HANDBOOK OF
SOIL SCIENCES
PROPERTIES AND PROCESSES

SECOND EDITION

Handbook of Soil Sciences

Handbook of Soil Sciences: Properties and Processes, Second Edition

Handbook of Soil Sciences: Resource Management and Environmental Impacts, Second Edition

HANDBOOK OF SOIL SCIENCES

PROPERTIES AND PROCESSES

SECOND EDITION

Edited by

Pan Ming Huang
Yuncong Li
Malcolm E. Sumner

To

Pan Ming Huang (*1934–2009*)

Editor in Chief

Dr. Pan Ming Huang died on September 13, 2009, after a short illness. He was a towering leader in the field of soil science, and soil chemistry in particular, having served the profession for over 44 years formerly as professor of soil science at the University of Saskatchewan, Saskatoon, Saskatchewan, Canada. He was a dedicated teacher and author of over 300 journal articles published in the most prestigious journals. In the midst of the revision of the *Handbook of Soil Sciences*, having developed and approved the outline, set the standards for the contributions, and solicited the cooperation of all the associate editors, he unexpectedly fell ill. One of his strengths was his ability to involve, organize, and inspire colleagues to cooperate in pursuing scientific endeavors as was evidenced by his initial inputs into this work. Unfortunately, he passed away before being able to see any of the fruits of his labors, the completed manuscripts. Dr. Huang is survived by his wife, Yun Yin, who was gracious in the handover of all the materials required for us to complete the handbook in a timely manner. We dedicate this book to Dr. Huang as a lasting memorial to his many contributions to soil science.

Contents

Part II   Soil Chemistry

Part III   Soil Mineralogy

Part IV   Soil Biology and Biochemistry: Soil Biology in Its Second Golden Age

Part V   Pedology

Preface

Handbook of Soil Sciences is the second edition of a comprehensive reference work on the discipline of soil science as practiced today. The new edition has been completely revised and rewritten to reflect the current state of knowledge. It contains definitive descriptions of each major area in the discipline, including its fundamental principles, appropriate methods to measure each property, many examples of the variations in properties in different soils throughout the world, and guidelines for the interpretation of the data for various applications (agricultural, engineering, and environmental).

This handbook assembles the core of knowledge from all fields encompassed within the discipline of soil science. It is a resource rich in data, which will provide professional soil scientists, agronomists, engineers, ecologists, biologists, naturalists, and students with their point of first entry into a particular aspect of soil science. The contributions serve those professionals seeking factual reference information on a particular aspect. The handbook provides a thorough understanding of soil science principles and practices based on a rigorous, complete, and up-to-date treatment of the subject matter compiled by the leaders in each field. In general, the following critical elements are present in each part: description of concepts and theories, definitions, approaches, methodologies and procedures, data in tabular and figure forms, and extensive references.

The handbook is organized into two books comprising nine parts, covering the six traditional areas of soil science together with a new part dealing with interfacial interactions among the physical, chemical, and biological regimes within the soil, a part on the interdisciplinary aspects, and a final part on databases. The two books are organized as follows:

Handbook of Soil Sciences: Properties and Processes

Part I: Soil Physics
Part II: Soil Chemistry
Part III: Soil Mineralogy
Part IV: Soil Biology and Biochemistry
Part V: Pedology

Handbook of Soil Sciences: Resource Management and Environmental Interactions

Part I: Soil Physical, Chemical, and Biological Interfacial Interactions
Part II: Soil Fertility and Plant Nutrition
Part III: Interdisciplinary Aspects of Soil Science
Part IV: Soil Databases

The subdivision of each part into a series of chapters was made by the associate editors and, in some cases, may appear to be somewhat arbitrary. The chapters have been arranged in such a way as to produce a thread running through each part. A complete table of contents, which is provided at the beginning of each book, gives a general outline of the scope of the covered subject material. In addition, a comprehensive subject index and an assemblage of units in common usage in soil science are provided at the end.

When the revision of this work was started, Dr. Pan Ming Huang, who was the associate editor for the soil chemistry part in the first edition, took on the role of editor in chief. Tragically and unexpectedly, he passed away in September 2009 in the midst of overseeing this expansive work. At short notice, the publisher approached us to complete this massive task to which we agreed in November 2009. Posthumously, we wish to recognize the vision, hard work, and dedication of our friend and colleague, Dr. Huang, in developing the revised outline of the book and soliciting the participation of all the associate editors and, in turn, the authors.

The chapters of this handbook have been written by many authors, all experts in their own fields, and peer-reviewed by independent reviewers. The nine parts have been carefully edited and integrated by the associate editors, all distinguished soil scientists in their own fields. This handbook is a tribute to the dedication of the authors, associate editors, and the publisher and its editorial associates. We wish to thank all the authors for their valuable contributions, the many nameless reviewers for their useful and helpful comments and criticisms, and the associate editors for all their hard work that they willingly contributed.

We also wish to recognize John Sulzycki who was highly effective in twisting our arms to pick up the threads and continue with the project after Dr. Huang's death, Randy Brehm who continued with the management of this project, and Jill Jurgensen, Vinithan Sethumadhavan, and Suganthi Thirunavukarasu who dealt with the minutiae of the editorial process. In addition, several research associates, graduate and undergraduate students, and visiting scientists, especially Helena Ren, Gaelan Jones, and David Li at the Tropical Research and Education Center, University of Florida, helped us in the editorial process for which we are duly grateful. Finally, we would like to thank our wives, Priscilla and Zhitong, for their patience, sacrifice, understanding, support, and encouragement without which this project would not have been possible.

Editors

Dr. Pan Ming Huang (deceased) was the professor emeritus of soil science at the University of Saskatchewan, Saskatoon, Saskatchewan, Canada, and served for 44 years in that institution. He received his BSc (1957) from the National Chung Hsing University, Taiwan; his MSc (1962) from the University of Manitoba; and his PhD (1966) from the University of Wisconsin, Madison, Wisconsin. His research work significantly advanced the frontiers of knowledge on the formation, chemistry, nature, and surface reactivity of mineral colloids, organic matter, and organomineral complexes in soils and sediments and their role in the dynamics, transformations, and fate of nutrients, toxic metals, and xenobiotics in terrestrial and aquatic environments. His research findings, embodied in well over 300 refereed scientific publications, are fundamental to the development of sound strategies for managing land and water resources in the Earth's critical zone. In addition to developing and teaching courses in soil physical chemistry and mineralogy, soil analytical chemistry, and ecological toxicology, he trained and inspired MSc and PhD students and postdoctoral fellows, and received visiting scientists from all over the world. He served on numerous national and international scientific and academic committees. He also served as a member of editorial boards, including the *Soil Science Society of America Journal*. He served the International Union of Pure and Applied Chemistry and the International Union of Soil Sciences in a number of capacities. He received the Distinguished Researcher Award from the University of Saskatchewan, the Soil Science Research Award from the Soil Science Society of America, the Distinguished Alumnus Award and the Chair Professorship Award from the National Chung Hsing University, and the Y. Q. Tang Chair Professorship Award from Zhejiang University. He was a fellow of the Canadian Society of Soil Science, the Soil Science Society of America, the American Society of Agronomy, the American Association for the Advancement of Science, and the World Innovation Foundation.

Dr. Yuncong Li is the University of Florida Research Foundation (UFRF) professor of soil science in the Department of Soil and Water Science at the Tropical Research and Education Center, Institute of Food and Agricultural Sciences (IFAS), University of Florida in Homestead, Florida. He received his BS (1982) in soil science and agricultural chemistry from the Shandong Agricultural University, China; his MS (1990) in agronomy from the University of Georgia; and his PhD (1993) in environmental science from the University of Maryland. He is also an affiliated professor at the University of Florida's Center for Tropical Agriculture, Hydrologic Sciences Academic Cluster, School of Natural Resources and Environment, and Water Institute, and a courtesy professor at the Shandong Agricultural University. He has received many awards and distinctions including a fellow of both the American Society of Agronomy and the Soil Science Society of America, Food and Agriculture Organization (FAO) Fellow, Wilson Popenoe Award (InterAmerican Society for Tropical Horticulture), Outstanding Paper Award (Florida State Horticultural Society), Jim App Award (University of Florida), Junior Faculty Research Award (Sigma Xi, the international honor society of science and engineering), Senior Faculty Award (Gamma Sigma Delta, the honor society of agriculture), Research Innovation Award (University of Florida), Wachovia Extension Professional Award (Extension Association of Florida), and Research Foundation Professor Award (University of Florida). His research and extension program focuses on water and soil quality monitoring, assessment and remediation, management practices to improve nutrient use efficiency, and nutrient cycling in soils/sediments. He has authored or coauthored over 160 research papers, 70 extension articles, and 15 book chapters. He recently edited a book, *Water Quality Concepts, Sampling, and Analyses* (Taylor & Francis). He serves as an associate editor for *Critical Reviews in Environmental Science and Technology* and *Communications in Soil Science and Plant Analysis*. Additionally, he has chaired or cochaired 20 graduate students, served as a committee member for 24 other graduate students, and supervised 15 postdoctoral fellows and many international visiting scientists.

Dr. Malcolm E. Sumner is the Regents' professor emeritus of environmental soil science in the Department of Crop and Soil Sciences at the University of Georgia, Athens, Georgia. He received his BSc Agriculture (1954) in chemistry and soil science and his MSc Agriculture (cum laude) (1957) in soil physics from the University of Natal, South Africa, and his DPhil (1961) in soil chemistry from the University of Oxford, Oxford, United Kingdom. His alma mater, the University of Natal, awarded him an honorary doctor of science degree in recognition of his career

contributions to soil science. Before coming to the University of Georgia, he was professor and head of the Department of Soil Science and Agrometeorology at the University of Natal, South Africa. He has spent sabbatical leaves at the Agricultural University (Wageningen, the Netherlands), University of Missouri (Columbia), University of Wisconsin (Madison), University of Newcastle-upon-Tyne (United Kingdom), Vista University (South Africa), University of Adelaide (Australia), and the Commonwealth Scientific and Industrial Research Organization (CSIRO) (Australia). Over 90 graduate students and postdoctoral associates have studied under his guidance in South Africa and the United States where he taught courses in soil chemistry, soil physics, and soil fertility. In addition to being a fellow of both the American Society of Agronomy and the Soil Science Society of America, he holds the Agronomic Research Award and the Werner L. Nelson Award for Diagnosis of Yield Limiting Factors from the American Society of Agronomy and the Soil Science Research Award, International Soil Science Award, and the Soil Science Distinguished Service Award from the Soil Science Society of America. He is the holder of many other awards and distinctions including the Sir Frederick McMasters

Visiting Fellowship (Australia) and the D. W. Brooks Faculty Awards for Excellence in Research and International Agriculture (University of Georgia). He has presented addresses at more than 100 universities and research institutions throughout the world on his research findings. He has also served as associate editor for the *Soil Science Society of America Journal* and is a member of editorial boards of many other journals. His published works cover a wide range of topics, including subsoil acidity, the agricultural uses of gypsum, diagnosis of yield-limiting factors, beneficial use of anthropogenic wastes, and transport of nutrients in soils. A widely respected author, Dr. Sumner's works include *Soil Acidity* (Springer-Verlag, 1991), *Soil Crusting: Chemical and Physical Processes* (Lewis Publishers, 1992), *Suelos de la Agroindustria Cafetalera de Guatemala* (University of Georgia, 1994), *Distribution, Properties and Management of Australian Sodic Soils* (CSIRO Publications, 1995), *Sodic Soils: Distribution, Properties, Management, and Environmental Consequences* (Oxford University Press, 1998), and the *Handbook of Soil Science* (CRC Press, 1999). He has authored or coauthored over 350 scientific papers, including 220 refereed journal articles, and has contributed chapters to over 30 books.

Associate Editors

Dr. Paolo Nannipieri (Soil Biology and Biochemistry) received his BSc and MSc biology degrees from Pisa University and his PhD (soil biochemistry) from the University of Pisa. He is a member of the Soil Science Society of America, the Italian Society of Soil Science, and the Italian Society of Agricultural Chemistry. He is currently professor of agricultural biochemistry and head of the Department of Plant, Soil and Environmental Sciences at University of Firenze, Firenze, Italy. He served as president of the Italian Society of Agricultural Chemistry and is currently serving as the president of the Commission 2.3 Soil Biology of the International Union of Soil Societies (IUSS). He is also the editor in chief of *Biology and Fertility of Soils*. He holds several distinctions for his global contributions to soil science.

Dr. Donald L. Sparks (Soil Chemistry) is the S. Hallock du Pont Chair in soil and environmental chemistry, Francis Alison professor, and director of the Delaware Institute for the Environment at the University of Delaware (UD). Dr. Sparks is internationally recognized for his research in the areas of kinetics of geochemical processes and mechanisms of metal/oxyanion/nutrient reactions at biogeochemical interfaces using in situ spectroscopic and microscopic techniques. He is the author or coauthor of 281 publications and 2 soil chemistry textbooks. He has given plenary and keynote presentations at scientific conferences throughout the world and been a lecturer at 80 universities and institutes and serves on the editorial boards of eight scholarly journals. He has received numerous awards and honors, including fellow of the American Society of Agronomy, the Soil Science Society of America, the American Association for the Advancement of Science, the Geochemical Society, and the European Association of Geochemistry; the M. L. and Chrystie M. Jackson Soil Science Award and the Soil Science Research Award from the Soil Science Society of America; the Environmental Quality Research Award from the American Society of Agronomy; the Francis Alison Award, the UD's highest faculty award; UD's Doctoral Student Advising and Mentoring Award; Distinguished Alumni Awards from the University of Kentucky and Virginia Tech; Honorary Membership in the Polish Soil Science Society; and the Geoffrey Marshall Mentoring Award. Dr. Sparks has also served as president of the Soil Science Society of America and the International Union of Soil Sciences.

Dr. Joseph W. Stucki (Soil Mineralogy) is professor of environmental soil physical chemistry in the Department of Natural Resources and Environmental Sciences at the University of Illinois, Urbana-Champaign, Illinois. He is a fellow of both the American Society of Agronomy and the Soil Science Society of America and holds the Marion L. and Chrystie M. Jackson Soil Science Award from the Soil Science Society of America. He served as associate editor of the *Soil Science Society of America Journal* and chaired the Soil Mineralogy Division of the Soil Science Society of America. He also received the Marion L. and Chrystie M. Jackson Mid-Career Clay Scientist Award from the Clay Minerals Society, where he served as president in 1997–1998.

Dr. Markus Tuller (Soil Physics) received his MS and PhD in civil engineering and water management from the University of Natural Resources and Applied Life Sciences (BOKU) in Vienna, Austria. He is currently an associate professor of soil and environmental physics in the Department of Soil, Water, and Environmental Science at the University of Arizona in Tucson, Arizona, with adjunct appointment in hydrology and water resources. He is the recipient of the Soil Science Society of America Soil Physics Early Career Award. His general area of research is focused on modeling and measurement of mass and energy transport and distribution in porous media. Specific research interests include hydraulic behavior of swelling soils and liner materials, interfacial pore-scale phenomena, liquid behavior in porous media in reduced gravity, and the application of x-ray computed tomography (CT) for porous media research.

Dr. Larry T. West (Pedology) received his BSc and MSc in soil science from the University of Arkansas and his PhD (soil science) from Texas A&M University. He served on the faculty at the University of Georgia for 20 years and is currently with the USDA-Natural Resources Conservation Service as National Leader for Soil Survey Research and Laboratory at the National Soil Survey Center, Lincoln, Nebraska. Dr. West has served as chair of Division S-5 of the Soil Science Society of America.

Contributors

Robert J. Ahrens
National Soil Survey Center
Natural Resources Conservation Service
United States Department of Agriculture
Lincoln, Nebraska

Richa Anand
Faculty of Land and Food Sciences
Department of Soil Science
The University of British Columbia
Vancouver, British Columbia, Canada

Olafur Arnalds
Agricultural University of Iceland
Reykjavik, Iceland

Richard W. Arnold
Soil Survey Division
Natural Resources Conservation Service
United States Department of Agriculture
Washington, District of Columbia

Susumu Asakawa
Soil Biology and Chemistry
Graduate School of Bioagricultural
    Sciences
Nagoya University
Nagoya, Japan

Judith Ascher
Department of Plant, Soil and
    Environmental Science
University of Florence
Florence, Italy

Jeffrey A. Baldock
Land and Water Division
Commonwealth Scientific and Industrial
    Research Organisation
Glen Osmond, South Australia, Australia

Raffaella Balestrini
Consiglio Nazionale delle Ricerche
Istituto per la Protezione delle Piante
and
Dipartimento di Biologia Vegetale
Università di Torino
Turin, Italy

Fredrich H. Beinroth
Department of Agronomy and Soils
University of Puerto Rico
San Juan, Puerto Rico

J.C. Bell
Department of Soil, Water, and Climate
University of Minnesota
St. Paul, Minnesota

Valeria Bianciotto
Consiglio Nazionale delle Ricerche
Istituto per la Protezione delle Piante
and
Dipartimento di Biologia Vegetale
Università di Torino
Turin, Italy

Paul R. Bloom
Department of Soil, Water, and Climate
College of Food, Agricultural and
    Natural Resource Sciences
University of Minnesota
St. Paul, Minnesota

J.G. Bockheim
Department of Soil Science
University of Wisconsin-Madison
Madison, Wisconsin

Janis L. Boettinger
Plants, Soils, and Climate Department
Utah State University
Logan, Utah

Paola Bonfante
Consiglio Nazionale delle Ricerche
Istituto per la Protezione delle Piante
and
Dipartimento di Biologia Vegetale
Università di Torino
Turin, Italy

Michael Bonkowski
Department of Terrestrial Ecology
Institute of Zoology
University of Cologne
Koln, Germany

J. Bouma
Environmental Sciences Group
Wageningen University and Research
    Centre
Wageningen, the Netherlands

Ian C. Bourg
Earth Sciences Division
Lawrence Berkeley National Laboratory
and
Geochemistry Department
University of California, Berkeley
Berkeley, California

Kris Broos
Flemish Institute for Technological
    Research
Boeretang, Belgium

Dominic A. Brose
Department of Environmental Science
    and Technology
University of Maryland
College Park, Maryland

Else K. Bünemann
Institute of Plant, Animal and
    Agroecosystem Sciences
Eidgenössische Technische Hochschule
    Zürich
Zürich, Switzerland

M.A. Callaham, Jr.
Southern Research Station
United States Department
    of Agriculture-Forest Service
Athens, Georgia

Florence Carré
Institute for Environment and Sustainability
European Commission Joint Research
    Centre
Ispra, Italy

Maria Teresa Ceccherini
Department of Plant, Soil and
    Environmental Science
University of Florence
Florence, Italy

Oliver A. Chadwick
Department of Geography
University of California, Santa Barbara
Santa Barbara, California

Yin Chen
Department of Biological Sciences
University of Warwick
Coventry, United Kingdom

G. Jock Churchman
School of Agriculture, Food and Wine
University of Adelaide
Glen Osmond, South Australia, Australia

Marianne Clarholm
Department of Forest Mycology and
    Pathology
Swedish University of Agricultural
    Sciences
Uppsala, Sweden

David C. Coleman
Department of Zoology
University of Georgia
Athens, Georgia

M.E. Collins
Department of Soil and Water Science
Institute of Food and Agricultural Sciences
University of Florida
Gainesville, Florida

D.A. Crossley, Jr.
Department of Entomology
University of Georgia
Athens, Georgia

R.G. Darmody
Department of Natural Resources and
    Environmental Sciences
University of Illinois
Urbana and Champaign, Illinois

Guo-Chun Ding
Federal Research Centre for Cultivated
    Plants
Julius Kühn-Institut
Institute for Epidemiology and Pathogen
    Diagnostics
Messeweg, Germany

James A. Doolittle
National Soil Survey Center
Natural Resources Conservation Service
United States Department of Agriculture
Newton Square, Pennsylvania

Steven G. Driese
Department of Geology
Baylor University
Waco, Texas

Hari Eswaran
Soil Survey Division
Natural Resources Conservation Service
United States Department of Agriculture
Washington, District of Columbia

Steven R. Evett
Soil and Water Management Research Unit
Conservation and Production Research
    Laboratory
Agricultural Research Service
United States Department of Agriculture
Bushland, Texas

Delvin S. Fanning
Department of Agronomy
University of Maryland
College Park, Maryland

Robert G. Ford
National Risk Management Research
    Laboratory
Office of Research and Development
United States Environmental Protection
    Agency
Cincinnati, Ohio

D.P. Franzmeier
Department of Agronomy
Purdue University
West Lafayette, Indiana

Emmanuel Frossard
Institute of Plant, Animal and
    Agroecosystem Sciences
Eidgenössische Technische Hochschule
    Zürich
Zürich, Switzerland

Jean-Claude Germon
Department of Soil and Environmental
    Microbiology
Institut National de la Recherche
    Agronomique
Dijon, France

Teamrat A. Ghezzehei
School of Natural Sciences
University of California, Merced
Merced, California

Sabine Goldberg
United States Salinity Laboratory
Agricultural Research Service
United States Department of Agriculture
Riverside, California

Robert C. Graham
Department of Environmental Science
University of California, Riverside
Riverside, California

E.G. Gregorich
Agriculture and Agri-Food Canada
Ottawa, Ontario, Canada

Bryan Griffiths
Teagasc
Environment Research Centre
Wexford, Ireland

Peter M. Groffman
Cary Institute of Ecosystem Studies
University of Georgia
Millbrook, New York

C.T. Hallmark
Department of Soil and Crop Sciences
Texas A&M University
College Station, Texas

A.G. Hardie
Department of Soil Science
Stellenbosch University
Stellenbosch, Western Cape, South Africa

Willie Harris
Department of Soil and Water Sciences
Institute of Food and Agricultural Sciences
University of Florida
Gainesville, Florida

James Harsh
Department of Crop and Soil Sciences
Washington State University
Pullman, Washington

Paul F. Hendrix
Department of Crop and Soil Sciences
Odum School of Ecology
University of Georgia
Athens, Georgia

Holger Heuer
Federal Research Centre for Cultivated
    Plants
Julius Kühn-Institut
Institute for Epidemiology and Pathogen
    Diagnostics
Messeweg, Germany

Daniel R. Hirmas
Department of Geography
University of Kansas
Lawrence, Kansas

Rainer Horn
Institut für Pflanzenernährung und
    Bodenkunde
Christian-Albrechts-Universität zu Kiel
Kiel, Germany

Robert Horton
Department of Agronomy
Iowa State University
Ames, Iowa

Pan Ming Huang (Deceased)
Department of Soil Science
University of Saskatchewan
Saskatoon, Saskatchewan, Canada

Wayne H. Hudnall
Department of Plant and Soil Science
Texas Tech University
Lubbock, Texas

Bruce R. James
Environmental Science and Technology
University of Maryland
College Park, Maryland

Jan Jansa
Institute of Plant, Animal and
    Agroecosystem Sciences
Eidgenössische Technische Hochschule
    Zürich
Zürich, Switzerland

H.H. Janzen
Agriculture and Agri-Food Canada
Lethbridge, Alberta, Canada

Jiri Jirout
Faculty of Science
Institute of Soil Biology
Biology Centre
University of South Bohemia
České Budějovice, Czech Republic

Scott B. Jones
Department of Plants, Soils, and Climate
Utah State University
Logan, Utah

Nestor Kämpf
Departamento de Solos
Universidade Federal do Rio Grande do
    Sul
Porto Alegre, Rio Grande do Sul, Brazil

Michael A. Kertesz
SUNFix Centre for Nitrogen Fixation
Food and Natural Resources
The University of Sydney
Sydney, New South Wales, Australia

Makoto Kimura
Laboratory of Soil Biology and
    Chemistry
Graduate School of Bioagricultural
    Sciences
Nagoya University
Chikusa, Nagoya, Japan

Hideomi Kodama
Centre for Land and Biological Resources
    Research
Agriculture and Agri-Food Canada
Ottawa, Ontario, Canada

Randall K. Kolka
Northern Research Station
United States Department of Agriculture
    Forest Service
Grand Rapids, Minnesota

Sylvia Koszinski
Institute of Soil Landscape Research
Leibniz Centre for Agricultural
    Landscape Research (ZALF)
Müncheberg, Germany

Deepak Kumaresan
Department of Biological Sciences
University of Warwick
Coventry, United Kingdom

Hannan E. LaGarry
Department of Physical and Life Sciences
Chadron State College
Chadron, Nebraska

Loretta Landi
Department of Soil Science and Plant
    Nutrition
Institute of Biological Chemistry
University of Florence
Florence, Italy

Inmaculada Lebron
Environment Centre Wales
Centre for Ecology & Hydrology
Bangor, United Kingdom

Feike J. Leij
Department of Civil Engineering and
    Construction Management
California State University, Long Beach
Long Beach, California

Yuncong Li
Department of Soil and Water Science
Tropical Research and Education
    Center
Institute of Food and Agriculture Sciences
University of Florida
Homestead, Florida

Zamir Libohova
National Soil Survey Center
National Resources Conservation Service
United States Department of Agriculture
Lincoln, Nebraska

Henry Lin
Department of Crop and Soil Sciences
Pennsylvania State University
University Park, Pennsylvania

David Long
Department of Geological Sciences
Michigan State University
East Lansing, Michigan

David J. Lowe
Department of Earth and Ocean Sciences
University of Waikato
Hamilton, New Zealand

Robert A. MacMillan
World Soil Information
International Soil Reference and
    Information Centre
Wageningen, the Netherlands

Terence L. Marsh
Department of Microbiology and
    Molecular Genetics and Center for
    Microbial Ecology
Michigan State University
East Lansing, Michigan

Alex B. McBratney
Department of Agriculture, Food and
    Natural Resources
The University of Sydney
Sydney, New South Wales, Australia

Paul A. McDaniel
Soil Science Division
University of Idaho
Moscow, Idaho

D.G. McGahan
Department of Agribusiness, Agronomy,
    Horticulture and Range Management
Tarleton State University
Stephenville, Texas

Robert McSorley
Department of Entomology and
    Nematology
University of Florida
Gainesville, Florida

Erika Michéli
Department of Soil Science and
    Agricultural Chemistry
Szent István University
Godollo, Hungary

Budiman Minasny
Department of Agriculture, Food and
    Natural Resources
The University of Sydney
Sydney, New South Wales, Australia

Per Møldrup
Department of Biotechnology, Chemistry
    and Environmental Engineering
Aalborg University
Aalborg, Denmark

H. Curtis Monger
Department of Plant and Environmental
    Sciences
New Mexico State University
Las Cruces, New Mexico

J. Colin Murrell
Department of Biological Sciences
University of Warwick
Coventry, United Kingdom

Paolo Nannipieri
Department of Soil Science and Plant
    Nutrition
University of Florence
Florence, Italy

Lee C. Nordt
Department of Geology
Baylor University
Waco, Texas

Jeanette M. Norton
Department of Plants, Soils and Climate
Utah State University
Logan, Utah

Astrid Oberson
Institute of Plant, Animal and
    Agroecosystem Sciences
Eidgenössische Technische Hochschule
    Zürich
Zürich, Switzerland

Tyson Ochsner
Department of Plant and Soil Sciences
Oklahoma State University
Stillwater, Oklahoma

Dani Or
Department of Environmental Sciences
Eidgenössische Technische Hochschule
    Zürich
Zürich, Switzerland

Phillip Owens
Department of Agronomy
Purdue University
West Lafayette, Indiana

E.A. Paul
Natural Resource Ecology Laboratory
Colorado State University
Fort Collins, Colorado

Stephan Peth
Institut für Pflanzenernährung und
    Bodenkunde
Christian-Albrechts-Universität zu Kiel
Kiel, Germany

Laurent Philippot
Department of Soil and Environmental
    Microbiology
Institut National de la Recherche
    Agronomique
Dijon, France

Giacomo Pietramellara
Department of Plant, Soil and
    Environmental Science
University of Florence
Florence, Italy

Chien-Lu Ping
School of Natural Resources and
    Agricultural Sciences
University of Alaska
Palmer, Alaska

James I. Prosser
Institute of Biological and
    Environmental Sciences
University of Aberdeen
Aberdeen, United Kingdom

John H. Prueger
Soil, Water, and Air Resources Research
    Unit
National Laboratory for Agriculture and
    the Environment
Agricultural Research Service
United States Department of Agriculture
Ames, Iowa

Martin C. Rabenhorst
Department of Environmental Science
    and Technology
University of Maryland
College Park, Maryland

David E. Radcliffe
Department of Crop and Soil Sciences
University of Georgia
Athens, Georgia

Paul F. Reich
Soil Survey Division
Natural Resources Conservation Service
United States Department of Agriculture
Washington, District of Columbia

Giancarlo Renella
Department of Soil Science and Plant
    Nutrition
University of Florence
Florence, Italy

Daniel deB. Richter, Jr.
Nicholas School of the Environment
Duke University
Durham, North Carolina

David A. Robinson
Environment Centre Wales
Centre for Ecology & Hydrology
Bangor, United Kingdom

Dennis E. Rolston
Department of Land, Air and Water
    Resources
University of California
Davis, California

Randall Schaetzl
Department of Geography
and
Department of Geology
Michigan State University
East Lansing, Michigan

Andreas C. Scheinost
Helmholtz-zentrum Dresden-Rossendorf
Institute of Radiochemistry
Dresden, Germany

Joshua P. Schimel
Department of Ecology, Evolution and
    Marine Biology
University of California, Davis
Santa Barbara, California

Michael Schloter
Department of Terrestrial Ecogenetics
German Research Center for
    Environmental Health
Institute of Soil Ecology
Neuherberg, Germany

Philip J. Schoeneberger
National Soil Survey Center
Natural Resource Conservation Service
United States Department of Agriculture
Lincoln, Nebraska

Darrell G. Schulze
Department of Agronomy
Purdue University
West Lafayette, Indiana

Paul Schwab
Department of Agronomy
Purdue University
West Lafayette, Indiana

Antonella Sciortino
Department of Civil Engineering and
    Construction Management
California State University, Long Beach
Long Beach, California

John C. Seaman
Savannah River Ecology Laboratory
University of Georgia
Athens, Georgia

Joey N. Shaw
Department of Agronomy and Soils
Auburn University
Auburn, Alabama

Jirka Šimůnek
Department of Environmental Sciences
University of California, Riverside
Riverside, California

Joseph M. Skopp
School of Natural Resources
University of Nebraska
Lincoln, Nebraska

Ulf Skyllberg
Department of Forest Ecology and
    Management
Swedish University of Agricultural
    Sciences
Umeå, Sweden

Kornelia Smalla
Federal Research Centre for Cultivated
    Plants
Julius Kühn-Institut
Institute for Epidemiology and Pathogen
    Diagnostics
Messeweg, Germany

Christopher W. Smith
Soil Survey Division
Natural Resources Conservation Service
United States Department of Agriculture
Washington, District of Columbia

M.P.W. Sonneveld
Department of Environmental Sciences
Land Dynamics Group
Wageningen University and Research
    Centre
Wageningen, the Netherlands

Randal J. Southard
Department of Land, Air and Water
    Resources
University of California, Davis
Davis, California

Otto C. Spaargaren
Taxonomy of Soils
World Reference and Information Centre
Wageningen, the Netherlands

Donald L. Sparks
Department of Plant and Soil Sciences
and
Delaware Environmental Institute
University of Delaware
Newark, Delaware

Garrison Sposito
Division of Ecosystem Sciences
University of California, Berkeley
Berkeley, California

Sharath Srinivasiah
Department of Plant and Soil Sciences
University of Delaware
Newark, Delaware

Mark G. St. John
Landcare Research
Lincoln, New Zealand

Mark H. Stolt
Department of Natural Resources Science
University of Rhode Island
Kingston, Rhode Island

J.J. Stoorvogel
Laboratory of Soil Science and Geology
Wageningen University and Research
    Centre
Wageningen, the Netherlands

Joseph W. Stucki
Department Natural Resources and
    Environmental Sciences
University of Illinois
Urbana, Illinois

Donald L. Suarez
United States Salinity Laboratory
Agricultural Research Service
United States Department of Agriculture
    Forest Service
Riverside, California

David Swanson
Natural Resources Conservation Service
United States Department of Agriculture
Palmer, Alaska

C. Tarnocai
Agriculture and Agri-Food Canada
Ottawa, Ontario, Canada

R. Greg Thorn
Department of Biology
University of Western Ontario
London, Ontario, Canada

Judy A. Tolk
Soil and Water Management Research
    Unit
Conservation and Production Research
    Laboratory
Agricultural Research Service
United States Department of Agriculture
Bushland, Texas

Arlene J. Tugel
Natural Resources Conservation Service
United States Department of Agriculture
Las Cruces, New Mexico

Markus Tuller
Department of Soil, Water and
    Environmental Science
The University of Arizona
Tucson, Arizona

Judith K. Turk
Department of Environmental Sciences
University of California, Riverside
Riverside, California

Goro Uehara
Department of Tropical Plant and Soil
    Sciences
University of Hawaii at Manoa
Honolulu, Hawaii

P.C.J. van Vliet
Blgg AgroXpertus
Wageningen, the Netherlands

Vicente Vasquez
Department of Plant and Soil Sciences
University of Kentucky
Lexington, Kentucky

Thomas Voice
Department of Civil and Environmental
    Engineering
Michigan State University
East Lansing, Michigan

Ole Wendroth
Department of Plant and Soil Sciences
University of Kentucky
Lexington, Kentucky

Larry T. West
National Soil Survey Center
Natural Resources Conservation Service
United States Department of Agriculture
Lincoln, Nebraska

Larry P. Wilding
Department of Soil and Crop Sciences
Texas A&M University
College Station, Texas

Kurt E. Williamson
Integrated Science Center
The College of William and Mary
Williamsburg, Virginia

K. Eric Wommack
Department of Plant and Soil Sciences
University of Delaware
Newark, Delaware

Jon M. Wraith
College of Life Sciences and Agriculture
University of New Hampshire
Durham, New Hampshire

Douglas A. Wysocki
National Soil Survey Center
Natural Resources Conservation Service
United States Department of Agriculture
Lincoln, Nebraska

Introduction

Malcolm E. Sumner
University of Georgia

Yuncong Li
University of Florida

Larry P. Wilding (Retired)
Texas A&M University

What Is Soil?

Soil, the extremely thin but precious skin covering our planet, supports all terrestrial life forms and contributes nutrients to aquatic and marine environments. This covering of the unweathered and partially weathered geological formations at the Earth's surface is a unique, fragile veneer. No longer rock nor geological sediment, soil has been altered during the process of soil formation by geological, topographical, climatic, physical, chemical, and biological factors to form a living entity, which inextricably links inorganic or mineral particles with organic matter and biota bathed in a milieu of liquid water and gases. Water, the major constituent of all living entities, is the solvent and conveyor of nutrients, which, together with the solid phase, become the fertile substrate from which all planetary life springs. This nurturing and life-supporting pedosphere zone is a biologically active, porous, and structured medium that effectively integrates and dissipates mass fluxes and energy. In its pristine state, soil is a self-regulating, slowly evolving biogeochemical system that weathers with time, simulating sponge-like behavior and regulating and buffering nutrient supply and water quality and quantity for the growth of macro- and microflora and fauna. Furthermore, it determines the partitioning of water into surface and subterranean groundwater reservoirs.

In addition to its primary role in promoting, supporting, and sustaining all forms of life, soil also acts as a living filter for anthropogenic waste. Soil can recycle water through biochemically mediated processes that cleanse, purify, detoxify, and counteract most toxins and pathogens in polluted waters that would irreparably contaminate and degrade our environment. Despite being the repository of human and animal cadavers, including those from epidemics of pestilence and plague, all but a few vectors have been rendered harmless, and it is seldom, if ever, involved in the transmission of diseases. On the contrary, soil has been the source from which great antidotes to disease and infection, the antibiotics, have been developed.

In ancient times, soil and water degradation, resulting in unsustainable crop production often due to the accumulation of salts, led to the downfall of civilizations. Directly or indirectly, the soil resource impacts, undergirds, and transcends all of society's urban, industrial, and agrarian interests. Current local to global policies and issues on conservation and sustainability, land use, energy, environmental quality, taxation, and food, feed, and fiber production are derived largely from the quality and extent of the available soil resource.

Despite the clearly pivotal role that soil plays in supporting life on Earth, a precise definition of soil is elusive. At this stage, one can say that soil is an evolving, living organic/inorganic covering of the Earth's surface, which is in dynamic equilibrium with the atmosphere above, the biosphere within, and the geology below. Soil acts as an anchor for roots, a purveyor of water and nutrients, a residence for a vast and still largely unidentified community of microorganisms and animals, a sanitizer of the environment, and a source of raw materials for construction and manufacturing. Soil is an essential component of terrestrial ecosystems used by nations as long-term capital to grow and develop. Because this foundation is essential to the life forms within and on it, one requires a fundamental understanding of this elastic, porous, three-phase system (solid, liquid, gas), and its components, processes, and reactions in order to be able to effectively manage and exploit this vital resource. In the words of Roy Simonson "Soil resources are the earthen looms that shape the lives of the people. The more completely they are understood, the better can be the fabric of life woven on these earthen looms" (Wilding et al., 1984).

Every soil consists of one to several layers, called horizons, a few to hundreds of centimeters thick that reflect the physical, chemical, and biological processes that have taken place during its formation. Horizons are composed of natural aggregates called peds, which are made up of associations of mineral and organic particles. Peds and particles are often separated from each other by pores that vary widely in size and shape. In addition, individual peds and particles may be coated by materials such as clay, organic matter, sesquioxides, or precipitated salts. Although the internal structure of peds is not readily visible to the naked eye, their spatial arrangement involving particles and pores (soil architecture) greatly influences soil behavior because

the organization is frequently systematic and related to macropore distribution. At the microscale, soil architecture governs soil water/solute movement and retention, soil structure/porosity, soil strength/failure, mineral synthesis/weathering, movement of toxic and nontoxic wastes, soil/root environments, root growth/proliferation, nutrient transfers, soil erosion, and oxidation/reduction reactions.

What Are the Soil Sciences?

Soil sciences are a spectrum of Earth sciences that address the importance of soils as very slowly renewable natural resources. They involve the study of soil formation, classification, and mapping; the physical, chemical, biological, and mineralogical properties of soil from microscopic to macroscopic scales of resolution; as well as the processes and behavior of soil systems and their use and management. Soils are integrative links between the atmosphere, biosphere, lithosphere, and hydrosphere. Soil sciences provide tools to integrate the components of Earth science systems, expanding knowledge of the causes and effects of spatial variability, and to take a more holistic approach to the dynamic processes affecting ecosystems.

Soil sciences focus primarily on near-surface processes that govern the quality and distribution of soil resources relative to landform evolution (geomorphology), geochemical environment (geochemistry), and organismal habitat (ecology/biology). Pedogenic processes are interactively conditioned by lithology, climate, and relief through geologic time. Soils are welded together into landscapes like chains; processes that perturb and impact higher topographic surfaces directly affect processes on adjacent surfaces below. Pedologists and other soil scientists study the energy flows and mass fluxes, which are the dynamic driving forces of pedogenesis through and over the three-dimensional soil system. They also quantify renewal and transfer vectors for biomass production, rainfall, and dusts, which counter constituent losses through drainage water, lateral interflow, and downslope migration or erosion products.

Soil sciences have their parentage in geology, chemistry, physics, and biology, but for the past 100 years have evolved as an independent, interdisciplinary body of knowledge with strong underpinnings to agriculture and, more recently, the environment. Because of an unparalleled success that soil sciences have enjoyed in helping to bring ample food, fiber, feed, and fuel to the world, the development of the basic soil sciences has evolved primarily as a by-product of research in agriculture, engineering, and the environment. There is growing evidence that the complexity of these problems requires a much broader approach to the science of the soil than can be stimulated by applied research alone. Soil sciences are taking their place alongside basic research efforts in the biosciences, geosciences, and atmospheric sciences to provide the reservoir of fundamental knowledge needed to develop lasting solutions to the challenges of balanced use and stewardship of the Earth. Despite the knowledge accumulated over the past century, Leonardo da Vinci's statement that "We know more about the movement of the celestial bodies than about the soil underfoot" is as true today as it was then. There is need for much more investigation of our most precious resource to stave off the fate that befell our ancestors in the ancient empires of Babylon, Egypt, China, Europe, and the Americas where soil degradation through nutrient loss, erosion, sodification, and salinization together with siltation of rivers and storage facilities laid waste their land and water resources. The ancients were constrained by their limited knowledge of processes driving the system, but we will be judged by generations to come based on our superior understanding of the soil system and the quality and stewardship of our land resources.

Modern parallels of soil and land degradation continue in both developing and industrialized countries. In addition to the problems facing the ancients that are still prevalent, chemical toxicities, irreversible land use conversion, environmental pollution, and desertification are the modern plagues. Sustainability of today's culture is threatened by loss of biodiversity, population growth, and land degradation, leading to enhanced greenhouse gas emissions culminating in global climate change. Land degradation is a complex technical, socioeconomic, and political issue without simple answers. Although soil scientists can provide technical solutions, without implementation as public policy, they are likely to be futile. Currently, public attention is focusing on soil, air, and water quality in efforts to maintain a clean environment. Soil sciences have a major role to play in this arena.

What Is the Purpose of This Handbook?

Handbook of Soil Sciences is a revision of the first comprehensive reference on the disciplines of soil science as practiced today. It contains definitive descriptions of each major area in the disciplines, including fundamental principles, appropriate methods to measure each property, many examples of the variations in properties in different soils throughout the world, and guidelines for the interpretation of the data for various applications (agricultural, engineering, environmental, biological, regulatory, educational, hydrological, biogeochemical). This handbook assembles the current core of knowledge from all fields encompassed within the soil sciences and provides a resource-rich database that will give professional soil scientists, agronomists, engineers, ecologists, biologists, naturalists, and students their point of first entry into a particular aspect of the soil sciences. The contributions also serve those professionals seeking specific, factual reference information. The handbook provides a thorough understanding of soil science principles and practices based on a rigorous, complete, and cutting-edge treatment of the subject compiled by leading scientists in each field. It is designed as a desk reference book.

This handbook has been extended to two books, *Handbook of Soil Sciences: Properties and Processes* comprises five parts dealing with soil physics, soil chemistry, soil mineralogy, soil biology and biochemistry, and pedology. *Handbook of Soil Sciences: Resource Management and Environmental Impacts* discusses soil physical, chemical, and biological interfacial

interactions; soil fertility and plant nutrition; interdisciplinary aspects of soil science; and soil databases.

In the *Handbook of Soil Sciences: Properties and Processes*, the part on soil physics opens with a description of the basic physical properties of soils followed by a treatment of their dynamic properties in relation to tillage and disturbance by machinery. Soil water is discussed, including content/potential relationships, its movement and transfer to the atmosphere both directly and through the plant, and the co-transport of solutes in moving water. The nature of soil structure and its bearing on soil behavior are explored followed by an examination of gas movement under unsaturated conditions. The role of macropores as the superhighway for bypass transport of water and solutes is considered. Thermal regime and heat transfer are discussed. Finally, the heterogeneity of soils is described culminating in an evaluation of spatial variability.

The part on soil chemistry begins with a discussion of the nature and dynamics of soil organic matter. The importance of the soil solution, for transport of nutrients and catalyst for reactions in soil, is then highlighted followed by an evaluation and description of the kinetics of reactions in soils. The reactions taking place when oxygen becomes limiting are then explored followed by a detailed account of soil colloidal phenomena important in predicting the behavior of the finest soil fraction. Ion exchange, sorption, and precipitation reactions are quantitatively treated as a framework to evaluate the behavior of labile constituents in soils. Catalytic reactions promoted by soil and its constituents are examined followed by an account of the effects of acidity and alkalinity on soil reactions and behavior.

Mineralogy of soils is covered in the next part starting with a discussion of the formation and occurrence of minerals in soils followed by three parts in which structure, occurrence, identification, and properties of phyllosilicates, oxide minerals, and poorly crystalline aluminosilicates are discussed. The part on soil biology and biochemistry begins with a discussion of viruses, bacteria, fungi, cyanobacteria and algae, followed by the topic of soil fauna comprising protozoa, nematodes, micro- and macro-arthropods, enchytriads, and vertebrates. The nature, function, and life cycles of each are described. Then the processes mediated by these organisms, including nutrient transformations, are discussed. A major portion of the discussion is devoted to nitrogen transformations because of their importance in soils.

Following a discussion of the geomorphology of soil landscapes and the pedogenic processes and models involved in soil formation, descriptions of the systems used to classify soils are presented in the part on pedology as a framework in which to discuss the soil orders in soil taxonomy and other classification systems. The part continues with a discussion of land evaluation followed by new chapters on hydropedology and subaqueous soils. Thereafter, digital soil mapping and its applications are presented, followed by a discussion of soil change in the anthropocene in which the bridge between pedology, land use, and soil management is built.

In the *Handbook of Soil Sciences: Resource Management and Environmental Impacts*, the new part on soil physical, chemical, and biological interfacial interactions opens with a discussion of the methodology to study metal(oid)s and nutrients at critical zone interfaces followed by insights into the interactions between clay and organic materials in soils. Various aspects of the involvement of nanoparticles in chemical, biological, and hydrological processes are discussed, followed by the effects of mineral and organic colloids on enzymatic activity in soils and its impact on biogeochemical processes. Thereafter, the effects of biophysical, biogeochemical, and biological processes on rhizosphere processes and reactions are discussed. Subsequent chapters deal with mineralogical, chemical, physical, and (micro) biological interactions involved in the turnover and stabilization of organic matter and the transformations of metal(oid)s. The part concludes with a discussion of the physicochemical and biological interfacial processes governing the fate and transport of nutrients and pollutants in soils.

Bioavailability of macro- and micro-nutrients as well as their interactions are discussed in the part on soil fertility and plant nutrition prior to evaluating methods for estimating their potential availability to crops and methods of application as fertilizers. The causes of soil acidity and strategies for its amelioration are also outlined. Finally, the efficiencies of nutrient and water use are discussed. Thereafter, various interdisciplinary aspects of soil science such as salinity, sodicity, water repellency, wetland biogeochemistry, acid sulfate soils, soils and environmental quality, soil erosion by wind and water, land application of wastes, and soil quality and conservation tillage are discussed. Finally, the soil databases available to assess worldwide soil resources are presented and discussed.

Reference

Wilding, L.P., N.E. Smeck, and G.F. Hall (eds.). 1984. Pedogenesis and soil taxonomy. II. The soil orders: Developments in soil science 11B, p. v. Elsevier, Amsterdam, the Netherlands.

I

Soil Physics

Markus Tuller
The University of Arizona

S OIL PHYSICS is concerned with the application of physical principles to characterize the soil system and mass and energy transport processes within the critical zone, which extends from the bedrock to the top of the plant canopy. The exploitation of soil physical principles is not only feasible for natural systems, but also relevant for many industrial and engineering applications such as oil recovery, filtration, ceramics, powder technology, or nanoporous materials. Because soil physical processes are intimately intertwined with biological and chemical processes simultaneously occurring within the same soil volume, the soil physics part is also important for other parts of this handbook.

The second edition of the soil physics part was notably expanded, containing a new chapter about the soil thermal regime and many updates regarding recent theoretical and experimental advances.

Chapters 1 through 3 emphasize the soil solid phase. Included are descriptions and state-of-the-art measurement methods for physical properties of primary soil particles, and the arrangement and organization of primary particles into an overall structure of solids and pore space. At first, soil is considered as static to facilitate definitions of particle density, shape, size, surface area, bulk density, and porosity. Later, the mechanical properties of soils and dynamics induced by tillage operations and temporal variations due to natural and anthropogenic actions are thoroughly examined.

Soil water is the primary focus of Chapters 4 and 5, a major part of Chapters 6 and 7, and relevant for other chapters as well. Amount and energy state are the two most fundamental characteristics of soil water that govern many hydrological, biological, and geochemical processes. Air content and gas exchange with the atmosphere, soil temperature, movement of dissolved chemicals, or diffusion of nutrients to plant roots are intimately related to water content. In addition, changes in water content and energy state affect many mechanical and engineering properties of soils such as strength, compactibility, penetrability, or trafficability. Water content is appropriate to describe storage of water, and knowledge about the energy state (i.e., water potential) is needed to describe direction and magnitude of fluxes. Standard and emerging methods for measurement of water content and potential as well as new numerical modeling tools

for steady and transient water flow are provided. Moreover, soil water movement is described in the context of the overall water balance and its key role in the hydrologic cycle.

Transport processes are principally for water, energy, solutes, and gas. Generally, flow laws are related to an appropriate gradient such as for soil water potential, temperature, or concentrations of solutes. The individual descriptions are related to the appropriate measurable quantities to overall mass conversion.

Chapter 6 contains a detailed description of the water and energy balances at the soil–plant–atmospheric interface, which represents a common convergence point for many of the world's problems in food production, water resources, and environmental pollution. Solute transport, the topic of Chapter 7, is of fundamental importance for crop production (i.e., soil quality and fertility) and environmental pollution and sustainability.

Separate chapters are devoted to soil gas transport and the soil thermal regime. While gas movement (Chapter 8) is historically linked to soil aeration relevant for plant growth, in-depth knowledge about gas transport mechanisms is also important for contaminant remediation, the movement of volatile organic chemicals from waste sites, the design of optimal plant growth media for space exploration, and most importantly, the emission of greenhouse gases and associated public concerns with global warming.

The soil thermal regime and heat transport are described in Chapter 9, which is new to the second edition of the soil physics part. Temperature governs numerous biological, chemical, and physical processes occurring in the soil. Plant germination and growth are closely related to soil temperature. Furthermore, soil temperature affects hydraulic transport properties and important environmental processes such as gas exchange between soil and atmosphere, the global carbon cycle, or transformation and transport of contaminants.

Spatial variability, Chapter 10, is treated separately in recognition of the heterogeneity of all soil properties, in fact, of all measurable natural quantities. In soil science, the development of the quantitative aspects of variation is necessary for both the description of the processes and for forecasting future events. This has been a rich area for coupling soil physical processes with other areas of soil science and remote sensing.

1

Physical Properties of Primary Particles

Joseph M. Skopp
University of Nebraska

This chapter discusses the following physical properties of primary particles: particle density, particle shape, particle size distribution (PSD), and surface area. The definitions and ideas behind these properties have built into it the concept or assumption of discrete primary particles as the primary soil constituent. If organic matter or amorphous materials and cementing agents are abundant, then the importance of primary particles is reduced. In either case, the arrangement or packing of soil particles and these "glues" that control most physical properties of the soil rather than a complete description of the primary particles will be presented in this chapter.

1.1 Particle Density (ρ_p)

1.1.1 Definition

The particle density represents one of the fundamental soil physical properties. One can define particle density as the mass of soil particles divided by the volume occupied by the solids (i.e., excluding voids and water). Typical values for mineral soils range from 2.5 to 2.8 Mg m^{-3} with some soils within the broader range of 1.0–3.0 Mg m^{-3}. The standard value of 2.65 Mg m^{-3} applies only to mineral soils dominated by quartz. Particle density provides few insights into soil physical processes. Consequently, one frequently overlooks the errors associated with particle density. However, its value enters into a variety of calculations of more interesting soil properties (e.g., porosity and PSD).

1.1.2 Typical Values–Mineral, Whole Soil

Each individual soil mineral has a characteristic particle density. Values for different minerals can be found in Klein and Hurlbut (1985). Quartz, a predominant soil mineral, has a value of about 2.65 Mg m^{-3}, which is why this value is frequently given as representing all soils. In contrast, gypsum, biotite, and hematite have values of 2.32, 2.80–3.20, and 4.80–5.30 Mg m^{-3}, respectively. The particle density of a soil is a weighted average over the distribution of soil minerals present.

A determination of particle density may be for individual minerals, size fractions, or the soil as a whole. Organic matter removal may take place in order to reduce variation. With standard procedures and removal of organic matter, a propagated error of less than ±0.01 Mg m^{-3} is possible. (Propagated error is the combination of all instrument errors when making a determination of a soil physical property.) Generally, determinate errors (i.e., biases) play a greater role in the analysis of particle density than indeterminate errors (i.e., random errors).

Perhaps, the most interesting question in the determination of soil particle density is the role of organic matter (with a typical density of 1.0 Mg m^{-3}). Most standard methods remove organic matter. Thus, the particle density reflects only the mineral phase. This value is the best value for use in particle-size analysis. However, it may not be the best value for calculating porosity. Including organic matter means that changes in soil management may change this particle density.

1.1.3 Measurement Methods—Water Pycnometer, Air Pycnometer, Other

Three methods of determining particle density will be examined. Other methods are discussed by Flint and Flint (2002). The most common determination of a soil's particle density uses a pycnometer or some variation. A pycnometer is any device that can be made to retain a reproducible or measurable volume. The soil sample is introduced into the pycnometer and the displaced volume of a fluid is determined. Water is typically the displaced fluid but air can also be used.

The water pycnometer method generally requires the removal of organic matter prior to use. This reduces problems with trapped air and decreases variability. The water pycnometer method requires good temperature control (Blake and Hartge, 1986).

An alternative air pycnometer procedure uses a gas as the displacing fluid and the ideal gas law to calculate the volume of solids. The principles of an air pycnometer are straightforward but care must be taken to prevent temperature changes. The air pycnometer does not require the removal of organic matter, which is particularly valuable when the total porosity (or total void space) or air-filled porosity is required without knowing soil bulk density or water content.

A less common method of determining particle density uses a vibrating tube (Elder, 1979), which is filled with a solution or suspension. The resonant frequency of the vibrating tube provides a very precise means of determining the density of the suspension.

1.2 Particle Shape

1.2.1 Definitions–Shape Factors, Fractals

Particle shape influences specific surface as well as particle-size analysis and has a strong influence on the packing of particles and soil strength, as well as transport properties. Unfortunately, particle shape is a difficult property to measure and few determinations exist in the literature. Recent articles by Jones and Friedman (2000), Santamarina and Cho (2004), and Cho et al. (2006) show the range of soil properties that are influenced by particle shape.

A variety of terms exist for defining particle shape. Some of the definitions require reference to a figure of an ideal particle (Figure 1.1). Let L, B, and T represent the length, breadth, and thickness, respectively, of a particle. Then, Heywood (1947) defines the following:

$$\text{Flatness ratio} = B/T,$$

$$\text{Elongation ratio} = L/B.$$

Other dimensionless terms are as follows:

$$\text{Sphericity} = \frac{\text{surface area of equivalent sphere}}{\text{actual surface area}},$$
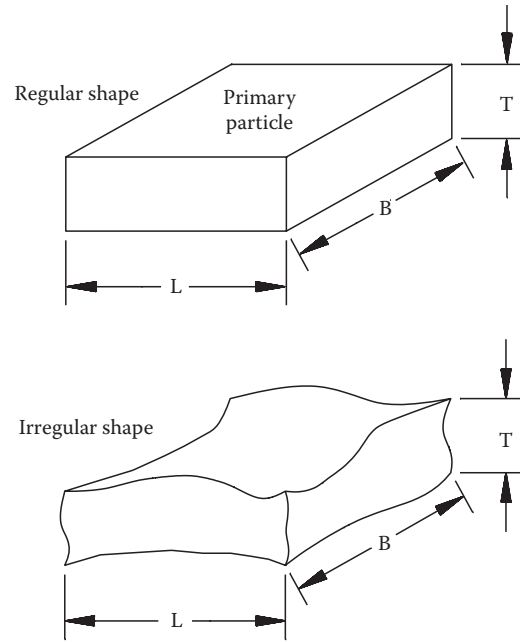
FIGURE 1.1 Ideal individual soil particle. (After Heywood, H. 1947. Symposium on particle size analysis. Trans. Inst. Chem. Eng. 22:214.)

$$\text{Circularity} = \frac{\text{circumference of circle with same area}}{\text{actual perimeter}},$$

$$\text{Rugosity} = \frac{\text{actual perimeter}}{\text{circumference of circumscribing circle}}.$$

These last two terms work best for a 2D image or the projection of the particle outline onto a flat surface.

A microscopic picture of soil describes either the solids or the voids. A simple picture of the solids might be a sphere or cube while that of the voids might be a cylindrical tube or a slit between two flat surfaces. These pictures or physical models of the soil make it possible to deduce relations describing surface area, packing, water retention, or water movement.

It is possible to apply these ideas to models that are not simple geometrical figures. Here, a "characteristic length" replaces the edge length of a cube and a dimensionless shape factor (C_k) describes the deviation of the geometry from simple geometries (e.g., Table 1.1). These two factors are used to describe the area and volume as

$$A = C_2 L^2 \quad \text{and} \quad V = C_3 L^3, \tag{1.1}$$
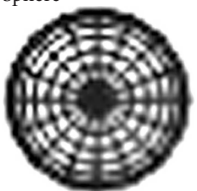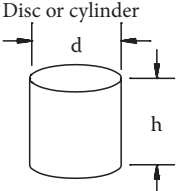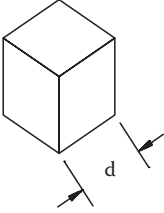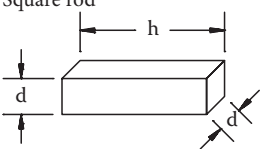
where
  L is the characteristic length
  C_2 or C_3 is the shape factor for area or volume, respectively

For a cube, if L equals d (the length of an edge), find $C_2 = 6$ and $C_3 = 1$. Examples of C_2 and C_3 for other shapes are in Table 1.1.

TABLE 1.1   Shape Factors for 3D Physical Models with L = d

Geometry	Area	C_2	Volume	C_3
Sphere	πd^2	Π	$\pi d^3/6$	$\pi/6$
Disc or cylinder	$(\pi dh + \pi d^2/2)$	$(\pi h/d + \pi/2)$	$\pi d^2h/4$	$\pi h/4d$
Cube	$6d^2$	6	d^3	1
Square rod	$2d^2 + 4dh$	$2 + (4h/d)$	d^2h	h/d

Two particles differing in size may or may not have the same shape. Typically, weathering or size reduction changes not only the total dimension but also the shape factor.

1.2.2 Measurement Methods

At least three methods exist to examine particle shape or shape factors. The first is direct observation under a microscope using an image analysis system (Davis and Dexter, 1972). Commercial image analysis systems exist that automatically provide output of shape factors or similar properties.

The second method is an indirect technique obtained from the variation of viscosity with the concentration of suspended particles. Increasing the solid concentration causes the suspension viscosity to deviate from pure liquid viscosity. These deviations are dependent on particle geometry as well as on concentration of the suspension. The viscosity of the suspension (η) and the viscosity of pure solvent (η_s) usually behave as follows:

$$\eta/\eta_s = 1 + Kf, \qquad (1.2)$$

where
  f is the volume fraction of suspended material
  K is an empirical constant, which is a shape parameter

This relation for spherical particles first appears in Einstein (1906). K = 2.5 is a constant for spheres but changes with the shape of the particle. Kahn (1959) applied this technique to examine the shape of clay particles. Similar techniques (Egashira and Matsumoto, 1981) provide estimates of a/b (the ratio of major to minor axis) for montmorillonite (200–300), kaolinite (15–25), and mica (10–20).

The third method of particle-shape analysis uses the scattering characteristics of light passing through a dilute soil suspension and relies on measurements of the angles of scattered light. Instruments commercially available for particle-size analysis can be used. This is an active area of research since particle shape has an influence on the accuracy of particle-size analysis based on light scattering (Naito et al., 1998).

1.3 Particle Size Distribution

1.3.1 Definitions

PSD is the most fundamental physical property of a soil and defines soil texture. The particle sizes present and their relative abundance sharply influence most physical properties.

Soil particle size (or effective diameter) provides the basis for a classification system. A range of diameters may be given a

special designation (e.g., 2.0–1.0 mm is "very coarse sand"). The ratio between the upper and lower bound is approximately 2 and particles within a given size range are called "soil separates." The size boundaries vary with country or discipline. Comparisons of the names given to a size range are given elsewhere (Gee and Bauder, 1986; Sheldrick and Wang, 1993).

The phrase "equivalent diameters" emphasizes the role of the measurement technique in determining particle size. If identical particles are measured by different techniques, they may appear to have different diameters. It is conceivable that two soils with identical PSDs (as determined by a single method) will show differences in other physical properties due to distinct particle shapes.

Defining the diameter of an irregularly shaped particle is not trivial. A single parameter, the diameter, characterizes a smooth sphere. The symmetry of the sphere and its smoothness mean that no other information is needed to describe it. Distorting the sphere (i.e., into a jelly bean shape) results in at least three possible diameters. Some particle-size analysis methods may orient the particle into a preferred direction (e.g., settling of a particle in a liquid). Other methods (e.g., image analysis) may observe several possible orientations.

1.3.2 Typical Distributions

Typical data for a variety of soils are presented in Table 1.2. Interpretation of particle-size analysis data requires either the drawing of graphs or the calculation of summary coefficients, which are discussed in Section 1.3.7.

Graphs of a PSD typically select the dependent variable as either the cumulative fraction up to a size limit or incremental fraction of soil in a size interval. The incremental fraction (F_i) is usually the mass of particles within a size interval (X_{i-1} to X_i) divided by the total mass of solids, with the index i specifying the size interval. The cumulative fraction (G_i) is the sum of all fractions for particle sizes less than the X_i value.

A typical graphical expression of PSD uses the logarithm of particle diameter (Figures 1.2 and 1.3). Soils frequently show a lognormal distribution of particle sizes, so that a graph of fraction versus the logarithm of particle diameters appears to be normally distributed. The shape and position of the lines provides qualitative information about the soil texture.

Alternative methods of representing the fraction include volume fraction and number fraction. Typically, conversion between distinct types of fractions may be necessary to compare distributions obtained with different methods. In these cases, the assumption is made of spherical particles and a standard particle density. It is important to understand how the method being used represents the PSD.

1.3.3 Dispersion and Fractionation

Particle size analysis (sometimes called mechanical analysis) consists of isolating various particle-size increments and then measuring the abundance of each size interval. Most methods accomplish this in two steps. First, the soil is dispersed or separated into individual primary particles. Second, the dispersed sample is fractionated, or the amounts of each size interval are measured.

There are three objectives of dispersion: (1) removal of cementing agents, (2) rehydration of clays, and (3) physical separation of individual soil particles. It is sufficient to recognize that organic matter and amorphous minerals are the primary bonding agents. When either of these are present in large amounts (e.g., histosols, andisols, or oxisols), then dispersion may be difficult or meaningless. Generally, soil dispersion is carried out using a combination of chemical and mechanical means (Gee and Bauder, 1986; Loveland and Whalley, 1991; Sheldrick and Wang, 1993; Gee and Or, 2002).

It is important to recognize that the fraction of soil that is a single size cannot be determined. What is detected is the fraction of soil within a particular particle-size interval or the cumulative fraction of all particles less than a given size. The use

TABLE 1.2 PSD of Soil Samples

| Soil Series | Sand | | | | | Silt | | | | Clay |
	Very Coarse 2–1 mm	Coarse 1–0.5 mm	Medium 0.5–0.25 mm	Fine 0.25–0.10 mm	Very Fine 0.1–0.05 mm	Coarse 50–20 µm	Medium 20–10 µm	Fine 10–5 µm	5–2 µm	2–0 µm
Anthony	18.05	13.71	17.68	12.93	8.92	7.41	2.69	2.20	1.37	15.04
Ava	0.53	0.56	1.25	0.82	0.67	12.80	21.69	15.71	9.63	30.63
Chalmers	0.74	0.62	1.67	1.38	2.52	19.42	20.18	11.18	7.44	34.89
Davidson	0.71	2.38	6.52	6.02	3.39	3.32	4.83	4.08	7.43	61.32
Fanno	8.45	4.87	2.40	9.96	9.06	5.92	4.27	4.05	5.56	46.46
Kalkaska	0.19	1.79	47.99	36.26	5.19	1.34	1.01	1.33	0.18	4.67
Mohave	15.25	11.30	12.40	8.02	5.42	30.36	5.00	1.34	0.43	10.45
Molokai	1.29	2.64	4.57	6.64	7.91	5.78	8.30	6.08	4.88	52.00
Nicholson	0.67	0.31	0.44	0.42	1.18	12.90	13.47	15.27	5.41	49.89
Wagram	7.48	20.70	32.06	21.81	5.84	2.00	1.37	1.59	3.31	3.84

Source: Data from D.M. Hendricks (1997, personal communication).
Mass percent of total sample for the indicated size class.
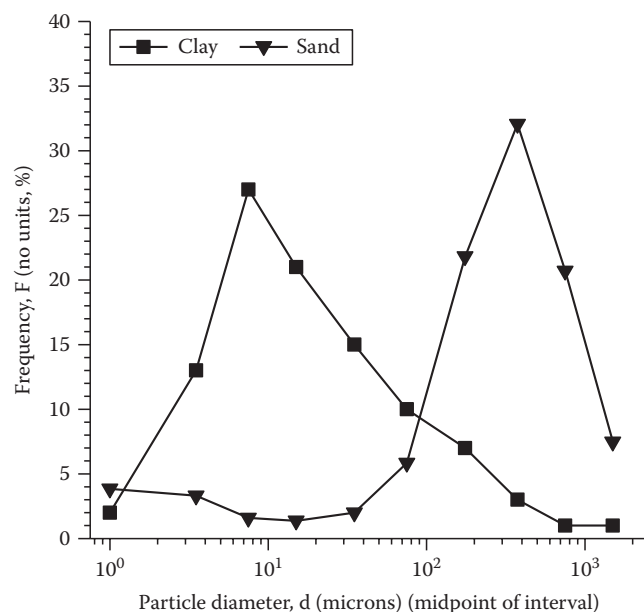
FIGURE 1.2   Frequency graph of PSD.
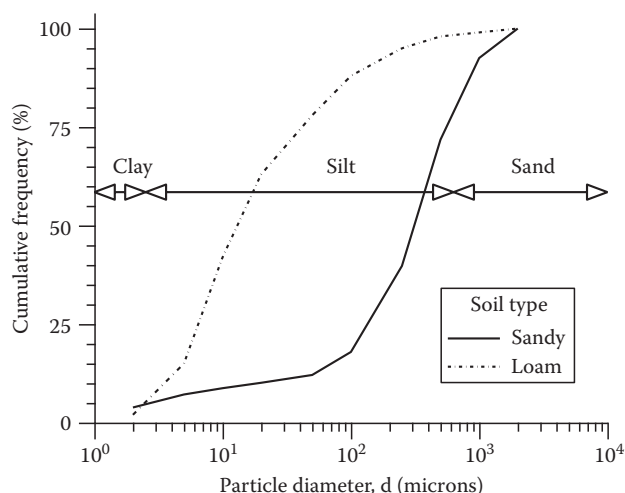
FIGURE 1.3   Cumulative frequency graph of PSD.

of sieves or sedimentation (e.g., pipette or SediGraph) typically determines the mass fraction (mass of particles in a size interval divided by total mass). Microscopic counting (e.g., image analysis systems or Coulter counter) results in a number fraction (number of particles in a size interval divided by total number of particles). Photometric techniques (e.g., FlowCAM®) typically determine the area fraction. Other methods result in volume or line fractions depending on the sensing procedure. A second complication of comparing PSDs from different methods is the manner in which the particle diameter is determined. In some cases, this is done directly (e.g., image analysis), rely on settling times (e.g., SediGraph), or diameter inferred from volume or other properties (e.g., Coulter counter). Thus, while all the methods are capable of observing PSDs, not all the methods provide results that are equal or directly comparable.

1.3.4 Sieving

The process of sieving is that of placing the particles on a pattern of holes. Small particles may fall through while the sieve retains the larger particles as the sieve is agitated. Either air or water may be the fluid to support the particles as they sort on the sieve. Dry sieving has a lower practical limit of about 50 μm, while wet sieving can separate smaller particle sizes. Changing the fluid can be expected to change the observed frequencies. Sieve holes may be square (using a wire cloth mesh) or round, although square holes are most common. The use of sieves with square openings will not result in measurements equivalent to those using sieves with round holes (Tanner and Bourget, 1952).

The use of words such as effective or nominal diameters with sieves is in recognition of the imperfect separation that may occur. Placement of a soil sample on a sieve does not result in instantaneous separation. Several factors influence the time to achieve a fixed level of separation. These factors include sample size, shaking intensity, particle shape, particle size, and hole geometry. Since samples vary in their sieving characteristics, it is best to run a trial sample. Errors on a single set of sieves typically are less than 1%, while comparisons between sieves show random errors of about 4%.

Many of the standard sieve sizes correspond to the class limits for USDA soil separates. Surprisingly, no standard sieve is available for the 50 μm cutoff between the sand and silt separates. Consequently, sieving cannot distinguish this class boundary using standard sieves.

1.3.5 Sedimentation

Below 50 μm, sieving is an inefficient and difficult procedure and for soil samples, sedimentation in water is an alternative procedure. A suspension of the dispersed sample (see Section 1.3.3) is allowed to settle in water, a measurement can be made of the density of particles suspended (mass of particles in a volume of liquid) at a specified depth within the sedimentation cylinder at known times. Variations in the method occur as to the determination of suspension density. In all cases, Stokes's law (Streeter and Wylie, 1975) is central to the derivation of an equation, which relates the time of settling to the size of particle sampled.

Two classic means of determining the density of a suspension exist: hydrometer and pipette methods. In the hydrometer method, the influence of suspension density on a floating object (the hydrometer) is observed. As density decreases (due to settling out of soil particles), the hydrometer sinks. A calibration scale converts the depth of the float (i.e., hydrometer) to the suspension density (expressed as g L^{-1}). The pipette method directly removes a sample from the solution. The concentration of solids is determined by drying a known volume of the pipetted sample. A proper accounting for dissolved salts requires a blank correction.

The hydrometer method uses higher concentrations of soil and may be less accurate than the pipette method. However, the hydrometer method allows repeated sampling at many points

on the distribution (since no sample is withdrawn). Problems in the pipette method due to convection currents near the tip of the pipette, effect of sample removal, and greater potential for operator error make the hydrometer method preferable in some circumstances (Gee and Bauder, 1979).

Stokes's law for the viscous drag on a sphere is combined with buoyant and gravitational forces to obtain a settling equation. The particle shape is also assumed to be a sphere in evaluating the other forces. Combining vertical forces and solving for v (the particle velocity) gives rise to the settling equation:

$$v = 2r^2 g (\rho_p - \rho_w)^a 9\eta, \qquad (1.3)$$

where
    r is the "equivalent radius"
    g is the gravitational acceleration
    η is the viscosity

The particle velocity is the distance traveled divided by the time or x/t. The settling equation allows a particle radius to be calculated at a particular x and t if the particle density and solution viscosity are known. This equation is basic to all gravity sedimentation procedures. Its use necessitates a number of assumptions not only in the derivation but also in the procedures, whereby the experiment is executed and the equation applied (Baver et al., 1972).

Another alternative to determine the suspension density is by gamma-ray attenuation (Karsten and Kotze, 1984). This method has the advantages of not disturbing the sedimentation column and allowing repeated nondestructive sampling over a large number of settling times. Sedimentation by pipette or hydrometer yields a mass frequency as a function of an effective settling diameter.

1.3.6 Coulter Counter

Individual particles can be counted in suspension by measuring the change in electrical resistivity (or conductivity) as particles pass through a small opening. Each particle displaces a volume of solution that is proportional to the signal. Thus, a single particle passing through the opening results in both a volume measurement (for size) and an increase by one in the count of the particles passing (Hurley, 1970). The commercial apparatus is called a Coulter counter and yields a number frequency as a function of a volume (or the equivalent radius of a sphere having the same volume). The range of diameters is typically between 0.4 and 1600 μm (Multisizer 4 COULTER COUNTER®, brochure http://www.beckmancoulter.com/coultercounter/showDoc.doc.do?filename=&fileid=11567).

1.3.7 Light Scattering

When light is passed through a suspension, several interactions can take place. Besides absorption, light can be scattered in a variety of directions. The behavior of the scattered light provides an alternative means to obtain a PSD. Commercial apparatus

may rely on the angle of scattered light or on dynamic measurements in the fluctuation of the scattered light to recover a PSD. The first method relies on the Mie theory and assumes spherical particles. The second method interprets the fluctuations in terms of a distribution of diffusion coefficients. These are interpreted as representing an equivalent spherical particle having the same diffusion characteristics. An example of the use of light scattering for soils can be found in Cooper et al. (1984).

1.3.8 Other Methods

A variety of variations exist to fractionation by settling under the influence of gravity in water. Settling can be speeded up through the use of a centrifuge (Svedberg and Nichols, 1923). Settling can occur in air, called air elutriation (Jensen and Hansen, 1961). The settling equation for air elutriation is the same as previously developed for water except that the velocity is that of the air floating particles up and the viscosity is that of air. Complete avoidance of settling may occur using microscopic methods to count and sort particles. Image analysis in conjunction with microscopic methods can determine particle shape or geometry parameters. Laser diffraction has also been used to obtain a PSD. More recently, acoustic methods have been developed (Guerin and Seaman, 2004). The acoustic methods also have the ability to determine electrical charge properties of soil particles. A variety of additional methods are discussed in the references by Davies (1984), Gee and Bauder (1986), Loveland and Whalley (1991), and Sheldrick and Wang (1993).

1.3.9 Interpretation of Results

A graph can clearly present the elements of a PSD. However, it is desirable to have a single parameter or a set of parameters to express the properties of the distribution. This is the origin of soil texture class that summarizes the particle-size properties in a single term. The phrase chosen depends on the relative abundance of sand, silt, and clay, irrespective of variations within these size ranges (e.g., Figure 1.4). Typical textural triangles and the textural class names are given in Gee and Bauder (1986) and Loveland and Whalley (1991). Note that while the names of three soil separates are similar to the names of textural classes, a textural class name does not limit the sizes that may be present. In other words, a clay texture class contains sand, silt, and clay separates while the clay separate contains only clay-sized particles.

The use of only the total amount of sand, silt, and clay to describe texture results in a bias that lies in the assumption that all particles within the range of the sand (or silt or clay) size are equivalent. For the clay fraction, this is a particularly misleading assumption and is partly responsible for the low correlations frequently observed when clay content is used in regression analysis. Two soils with the same percentage of clay may differ in the amounts of fine versus coarse clay. Additional differences in the shape and mineralogy of particles can also cause variations in soil behavior.

A more interesting approach determines the moments of the size distribution (Folk, 1966). This gives a quantitative measure
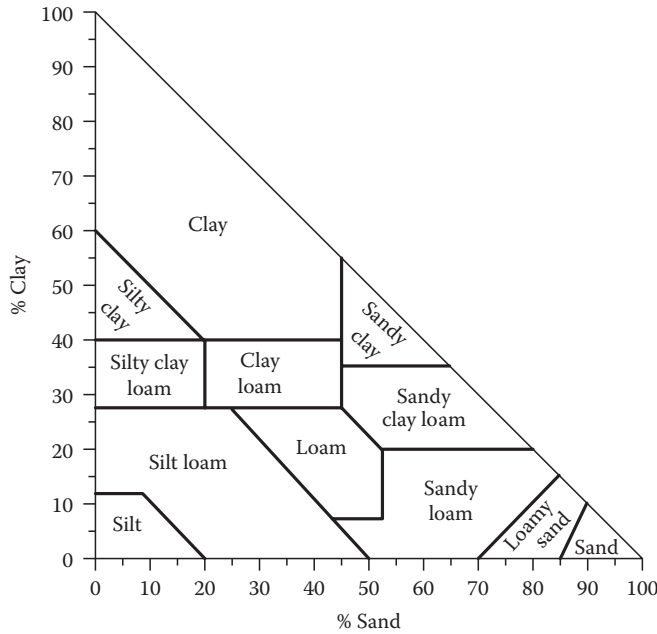
FIGURE 1.4 Alternative texture triangle using clay (%) and sand (%) to determine texture class name. For example, a soil with 50% of clay and 30% of sand sized particles would be assigned a textural class name of clay. (Reprinted from Elghamry, W., and M. Elashkar, 1962. Simplified textural classification triangles. Soil Sci. Soc. Am. Proc. 26:612–613. With permission of the Soil Science Society of America.)

of the mean (first ordinary moment), spread (second-order central moment or standard deviation), and asymmetry (third-order central moment or skewness) of a PSD. More importantly, it increases the power of correlative studies relating texture and any other physical, chemical, or biological factors of interest. As an example, the fourth moment of the size distribution has been related to soil compaction (Moolman, 1981).

If the PSD is known to follow a lognormal distribution, then the geometric mean diameter (d_g) and the geometric standard deviation (σ_g) summarize the PSD. These can be estimated from the mass fraction (f_i) of soil that is sand, silt, or clay. Then,

$$d_g = e^{\sum_1^3 f_i \ln(d_i)}, \tag{1.4}$$

$$\sigma_g = e^{\sqrt{\left\{\sum_1^3 f_i \ln^2(d_i)\right\} - (\ln d_g)^2}}, \tag{1.5}$$

where the d_i values are the average diameter of the size interval, $d_1 = 0.001$ mm, $d_2 = 0.026$ mm, and $d_3 = 1.025$ mm. See Shirazi and Boersma (1984) for more details. The equations above can be generalized if more than three soil separates have been measured and can be used even if the PSD does not follow a lognormal distribution.

Other parameters exist in the engineering literature and elsewhere to describe a PSD. One approach defines the grain diameter (D_n) at which n% passes through a sieve. While D_{40} represents that size for which 40% of the particles are smaller and 60% are larger.

Various combinations of these D_n characterize the distribution. One example is the uniformity coefficient (C_u):

$$C_u = \frac{D_{60}}{D_{10}}. \tag{1.6}$$

The uniformity coefficient provides information as to how narrow the distribution is with 1 being the minimum when only a single size is present. The second ratio is the coefficient of gradation (C_g):

$$C_g = \frac{(D_{30})^2}{(D_{60} D_{10})}. \tag{1.7}$$

Indirect descriptions of the PSD are possible through the use of an equation or model. These models contain parameters, which in turn characterize the distribution. The problem in using these models is first to determine the parameters; second, to determine the appropriateness of the model; and third, to interpret the parameters.

Two modes of particular interest are the power function and lognormal distribution. The use of a power function suggests an underlying fractal process. Where this applies, the exponent (δ) relates to the fractal dimension (n):

$$\delta = 3 - n, \tag{1.8}$$

and n is between 0 and 3. Tyler and Wheatcraft (1992) applied this technique to several materials and reported n values between 2 and 3. The lognormal model has also been useful (Shirazi and Boersma, 1984). Campbell (1985) shows how to estimate the water-holding properties of soil using lognormal model parameters.

1.4 Specific Surface Area

1.4.1 Definitions

The surface area of the individual particles is an important factor in nutrient or pesticide adsorption, water absorption, soil strength, and soil transport properties. The surface area of a soil has contributions from primary particles, amorphous mineral coatings, and organic matter. These individual contributions may overlap and cancel. Further, the surface area of some expanding minerals may change with the water content and chemical composition of the soil solution.

The surface area is an extensive quantity (i.e., depends on how much soil is present). A more satisfying alternative is the introduction of an intensive quantity, the specific surface, which is either the surface area per mass (S_m) or per volume (S_v). The specific surface per volume will change with soil compaction.

1.4.2 Typical Values—Separates, Whole Soil

Typical values for specific minerals as well as values for whole soils are presented in Table 1.3. Amorphous materials and soil organic matter (SOM) can greatly affect soil values. SOM has been reported to have a specific surface as high as 1000 m^2 g. The whole soil values include the effects of PSD as well as SOM and amorphous mineral levels for temperate zone soils.

TABLE 1.3   Ranges of Specific Surface of Clay Minerals, Soil Components, and Whole Soils Compiled from a Variety of Sources

Soils and Constituents	Specific Surface ($m^2 \, g^{-1}$)
Kaolinite	15–20
Illite	80–100
Bentonite	115–260
Montmorillonite	280–500
Organic matter	560–800
Calcite	0.047
Crystalline iron oxides	116–184
Amorphous iron oxides	305–412
Allophane and imogolite	700–1500
Sands	<10
Sandy loams and silt loams	5–20
Clay loams	15–40
Clays	>25

Sources: Nelson, R.A., and S.B. Hendricks. 1943. Specific surface of some caly minerals, soils, and soil colloids. Soil Sci. 56:285–296; Bower, C.A. and F.B. Gschwend. 1952. Ethylene glycol retention by soils as a measure of surface area and interlayer swelling. Soil Sci. Soc. Am. Proc. 16:342–345; Orchiston, H.D. 1955. Adsorption of water vapor: III. Homoionic montmorillonites at 25°C. Soil Sci. 79:71–78; Borggaard, O.K. 1982. The influence of iron oxides on the surface area of soil. J. Soil Sci. 33:443–449; Suarez, D.L., and J.D. Wood. 1984. Simultaneous determination of calcite surface area and content in soils. Soil Sci. Soc. Am. J. 48:1232–1235.

1.4.3 Methods

Direct measurements usually rely on the adsorption of either a gas (typically N_2 or Ar under high vacuum) or a liquid (historically ethylene glycol; but more recently, ethylene glycol monoethyl ether [EGME] or water) (Pennel, 2002). Either a multi- or monomolecular film is deposited or a gas adsorption isotherm is determined. A critical point in the use of all these procedures is the means by which a multi- or monomolecular layer is obtained. Control of the adsorbed phase occurs through the regulation of the gas phase. With gas adsorption, total pressure is controlled while with liquid adsorption, the use of a desiccator helps to fix both the total pressure and partial pressure of the adsorbed component. In the EGME method (Carter et al., 1986), a mixture of EGME and $CaCl_2$ regulates the vapor pressure of EGME. In water absorption, a separate saturated $LiNO_3$ solution regulates the relative humidity. The $LiNO_3$ and EGME methods are convenient procedures for soils. However, both methods are sensitive to variations in temperature. Values for specific surface vary with the method used.

The gas adsorption isotherm frequently takes the form of the BET equation (Brunauer et al, 1938). Here, pressure (P) is varied and the volume of adsorbed gas (V) is observed, all at constant temperature. The two are related as follows:

$$\frac{P}{V}(P_0 - P) = \frac{1}{V_m C} + \frac{(C-1)P}{V_m C P_0}, \tag{1.9}$$
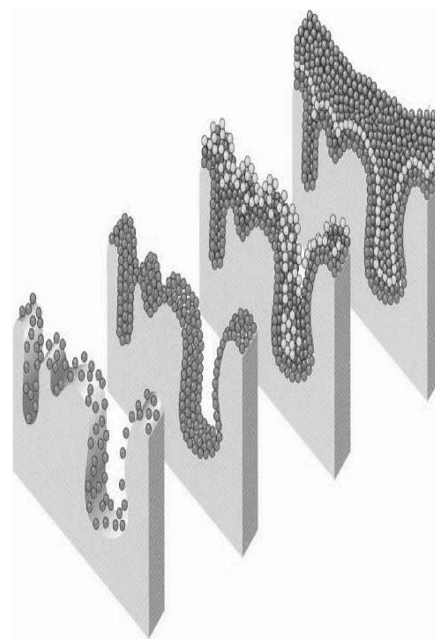
FIGURE 1.5   Surface coating with adsorbent molecules. Larger molecules will not fit in smaller pores.

where

V$_m$ is the volume of a monolayer

P$_0$ is the gas pressure at saturation

C = $\exp[(E_1 - E_2)/RT]$ where R is the gas constant, T is the absolute temperature, E_1 is the heat of adsorption of the first monolayer, and E_2 is the heat of subsequent layers or the heat of liquefaction of the gas

While this equation is frequently linearized, nonlinear regression should be used to estimate V$_m$ from which the specific surface is calculated. The use of either a gas or liquid to determine specific surface also raises questions about the size and nature (polar or nonpolar) of the adsorbing molecule (Figure 1.5). The smaller molecule can be expected to measure a greater surface area than a large one. Differences in surface area (when using distinct adsorbing molecules) are particularly noticeable when soils contain expanding clay minerals.

Another direct method to determine specific surface examines photomicrographs of primary particles. The classical approach determines the probability of a needle (randomly placed on the photo) either falling within a pore or intersecting the pore edge (Scheidegger, 1960). Current image analysis instrumentation allows this evaluation as part of many software packages. Unfortunately, most PSDs are not detailed enough in the smallest size range to accurately estimate specific surface.

1.4.4 Geometric Models and Regression

Geometric models are used to show how particle size influences specific surface and the general relation of particle geometry
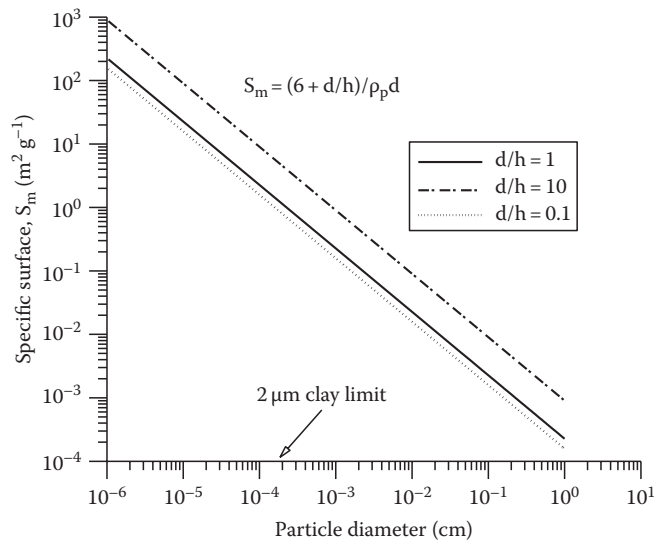
FIGURE 1.6 Effect of particle diameter on specific surface for materials of a single size.

to specific surface. Start with the specific surface, $S_m = A/\rho_p V$, where A is the surface area of solids and V is the volume of solids. Then, A and V are expressed in terms of the shape factors C_2 and C_3 and a characteristic dimension, L:

$$S_m = \frac{C_2}{\rho_p C_3 L}. \qquad (1.10)$$

This relation states that specific surface is the reciprocal of the characteristic length of the particles. The shape factors also influence surface area. Using Table 1.3, for spheres, $S_m = 6/\rho_p d$ while for a disc or flat plate, $S_m = [4 + 2(d/h)]/\rho_p d$. A large d/h ratio corresponds to a flat plate. A small d/h ratio corresponds to a prism or needle-like shape. A ratio of d/h = 1 is identical to the result for spheres (Figure 1.6). The specific surface of fine clays may be one or more orders of magnitude greater than for coarse clays.

Whole soils are best measured directly. In some cases, the amount of clay and knowledge of the type of clay can provide an approximate specific surface. Or and Wraith (1999) presented the following regression equations:

$$S_m = 5.65 \times \%\,clay - 18.9 \text{ for soils dominated}$$

$$\text{by montmorillonite} - \text{kaolinite clays}, \qquad (1.11)$$

$$S_m = 1.87 \times \%\,clay + 6.0 \text{ for soils dominated}$$

$$\text{by kaolinite} - \text{illite clays}. \qquad (1.12)$$

These and all regression equations should be used with care. The method used to measure or calculate surface area must be consistent with the use that will be made for the value.

References

Baver, L.D., W.H. Gardner, and W.R. Gardner. 1972. Soil physics. 4th ed. John Wiley & Sons, New York.

Blake, G.R., and K.H. Hartge. 1986. Particles density. *In* A. Klute (ed.) Methods of soil analysis. Part 1. SSSA, Madison, WI.

Borggaard, O.K. 1982. The influence of iron oxides on the surface area of soil. J. Soil Sci. 33:443–449.

Bower, C.A. and F.B. Gschwend. 1952. Ethylene glycol retention by soils as a measure of surface area and interlayer swelling. Soil Sci. Soc. Am. Proc. 16:342–345.

Brunauer, S., P.H. Emmett, and E. Teller. 1938. Adsorption of gases in multimolecular layers. J. Am. Chem. Soc. 60:309–319.

Campbell, G.S. 1985. Soil physics with basic. Elsevier, Amsterdam, the Netherlands.

Carter, D.L., M.M. Mortland, and W.D. Kemper. 1986. Specific surface, p. 414–424. *In* A. Klute et al. (eds.) Methods of soil analysis. Part 1. Physical and mineralogical methods. ASA, Madison, WI.

Cho, G., J.D. Dodds, and J.C. Santamarina. 2006. Particle shape effects on packing density, stiffness, and strength: Natural and crushed sands. J. Geotech. Geoenviron. Eng. 132:591–602.

Cooper, L.R., R.L. Haverland, D.M. Hendricks, and W.G. Knisel. 1984. Microtrac particle-size analyzer: An alternative particle-size determination method for sediment and soils. Soil Sci. 138:138–146.

Davies, R. 1984. Particle size measurement: Experimental techniques. *In* M.E. Fayel and L. Otten (eds.) Handbook of powder science and technology. Van Nostrand Reinhold Co., New York.

Davis, P.F., and A.R. Dexter. 1972. Two methods for quantitative description of soil particle shape. J. Soil Sci. 23:448–455.

Egashira, K., and J. Matsumoto. 1981. Evaluation of the axial ratio of soil clays from gray lowland soils based on viscosity measurements. Soil Sci. Plant Nutr. 27:273–279.

Einstein, A. 1906. A new determination of molecular dimensions. Ann. Phys. 19:289–306. Translated in: Investigations on the theory of Brownian movement. Dover Press, New York.

Elder, J.P. 1979. Density measurements by the mechanical oscillator, p. 12–25. *In* Methods in enzymology. Academic Press, New York.

Elghamry, W., and M. Elashkar, 1962. Simplified textural classification triangles. Soil Sci. Soc. Am. Proc. 26:612–613.

Flint, A.L. and L.E. Flint. 2002. Particle density, p. 229–240. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. SSSA Book Series No. 5. ASA, Madison, WI.

Folk, R.L. 1966. A review of grain-size parameters. Sedimentology 6:73–93.

Gee, G.W., and J.W. Bauder. 1979. Particle size analysis by hydrometer: A simplified method for routine textural analysis and a sensitivity test of measurement parameters. Soil Sci. Soc. Am. Proc. 43:1004–1007.

Gee, G.W., and J.W. Bauder. 1986. Particle size analysis, p. 383–411. *In* A. Klute (ed.) Methods of soil analysis. Part 1. 2nd ed. ASA, Madison, WI.

Gee, G.W., and D. Or. 2002. Particle-size analysis, p. 255–293. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4—Physical methods. SSSA Book Series No. 5. Soil Science. ASA, Madison, WI.

Guerin, M., and J.C. Seaman. 2004. Characterizing clay mineral suspensions using acoustic and electroacoustic spectroscopy: A review. Clays Clay Miner. 52:145–157.

Heywood, H. 1947. Symposium on particle size analysis. Trans. Inst. Chem. Eng. 22:214.

Hurley, J. 1970. Sizing particles with a Coulter counter. Biophys. J. 10:7679.

Jensen, E., and H.M. Hansen. 1961. An elutriator for particle-size fractionation in the sub-sieve range. Soil Sci. 92:94–99.

Jones, S.B., and S.P. Friedman. 2000. Particle shape effects on the effective permittivity of anisotropic or isotropic media consisting of aligned or randomly oriented ellipsoidal particles. Water Resour. Res. 36:2821–2833.

Kahn, A. 1959. Studies on the size and shape of clay particles in aqueous suspensions. Clays Clay Miner. 6:220–236.

Karsten, J.H.M., and W.A.G. Kotze. 1984. Soil particle size analysis with the gamma attenuation technique. Commun. Soil Sci. Plant Anal. 15:731–739.

Klein, C., and C.S. Hurlbut, Jr. 1985. Manual of mineralogy (after James D. Dana). 20th ed. Wiley & Sons, New York.

Loveland, P.J., and W.R. Whalley. 1991. Soil analysis physical methods, p. 271–328. *In* K.A. Smith and C.E. Mullins (eds.) Soil analysis: Physical methods. Marcel Dekker, Inc., New York.

Moolman, J.H. 1981. Soil textural properties influencing compactibility. Agrochemophysica 13:13–19.

Naito, M., O. Hayakawa, K. Nakahira, H. Mori, and J. Tsubaki. 1998. Effect of particle shape on the particle size distribution measured with commercial equipment. Powder Technol. 100:52–60.

Nelson, R.A., and S.B. Hendricks. 1943. Specific surface of some caly minerals, soils, and soil colloids. Soil Sci. 56:285–296.

Or, D., and J.M. Wraith. 1999. Temperature effects on soil bulk dielectric permittivity measured by time domain reflectometry: A physical model. Water Resour. Res. 35:371–383.

Orchiston, H.D. 1955. Adsorption of water vapor: III. Homoionic montmorillonites at 25°C. Soil Sci. 79:71–78.

Pennel, K.D. 2002. Specific surface area, p. 308–313. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. SSSA Book Series No. 5. ASA, Madison, WI.

Santamarina, J.C., and G.C. Cho. 2004. Soil behavior: The role of particle shape. *In* Advances in geotechnical engineering. 2: The Skempton Conference. March 2004. Thomas Telford Publishing, London, U.K.

Scheidegger, A.E. 1960. The physics of flow through porous media. Revised edition. University of Toronto Press, Toronto, Canada.

Sheldrick, B.H., and C. Wang. 1993. Particle size distribution, p. 123. *In* M.R. Carter (ed.) Soil sampling and methods of analysis. Canadian Society of Soil Science, Lewis Publishers, Boca Raton, FL.

Shirazi, M.A., and L. Boersma. 1984. A unifying quantitative analysis of soil texture. Soil Sci. Soc. Am. J. 48:142–147.

Streeter, V.L., and E.B. Wylie. 1975. Fluid mechanics, p. 752. 6th ed. McGraw-Hill Inc., New York.

Suarez, D.L., and J.D. Wood. 1984. Simultaneous determination of calcite surface area and content in soils. Soil Sci. Soc. Am. J. 48:1232–1235.

Svedberg, T., and J.B. Nichols. 1923. Determination of size and distribution of size of particles by centrifugal methods. J. Am. Chem. Soc. 45:2910.

Tanner, C.B., and S.J. Bourget. 1952. Particle-shape discrimination of round and square holed sieves. Soil Sci. Soc. Am. Proc. 16:88.

Tyler, S.W., and S.W. Wheatcraft. 1992. Fractal scaling of soil particle-size distributions: Analysis and limitations. Soil Sci. Soc. Am. J. 56:362–369.

2

Soil Structure

Teamrat A. Ghezzehei
University of California, Merced

2.1 Introduction

Soil structure refers to the organization and arrangement of soil particles and the resultant complex maze of pores. It is a composite soil quality that exerts significant control over most physical, chemical, and biological *processes* that occur in both natural and anthropogenically altered soils, including transport and storage of liquids, gases, and heat; penetration and proliferation of roots; microbial life; and decomposition and storage of soil organic matter. It is an inherently complex soil property because the constituent particles are heterogeneous in size, shape, and chemical nature as well as because the mechanisms that bond the particles are diverse. It is a hierarchical organization whereby fine primary particles bond together to form secondary particles and small secondary particles merge to form bigger peds or aggregates. It is also a dynamic soil property, continuously changing in response to various internal and external drivers, including moisture and temperature fluctuations, biological activity, and human intervention. As such it is a concept that spans multiple spatial and temporal scales: the arrangement of colloidal clay particles in a floccule, the arrangement of clods on the surface of a tilled layer, an array of earthworm tunnels, and the variability of soil strength from one point to another are all aspects of soil structure (Dexter, 1988). Due to these complexities, we still lack a truly objective description for it and a universally accepted and quantifiable measure of soil structure does not exist yet (Hillel, 1998).

The importance of soil structure has been acknowledged for most of the recorded agricultural history. For much of this time, it was perceived as the soil *tilth*—the structural state of seedbed resulting from tillage and affecting crop response (Hadas, 1997). For accounts of early writings about soil *tilth* and plowing, see the review by Warkentin (2008) and the citations therein. The Soil Science Society of America (1996) defines soil *tilth* as "the

physical conditions of soil as related to its ease of tillage, fitness as seedbed and its impedance to seedling emergence and root penetration." It refers to the favorable physical state of agricultural soil created by tillage, including balanced infiltrability and water retention, optimal aeration and exchange of gases, low resistance to root growth, resistance to erosion, and favorable conditions for microbial activity (Hadas, 1997). These conditions have in fact been the implicit goals of tillage over the ages. Interest in soil structure also arises from its importance to nutrient storage and cycling. A huge fraction of soil organic carbon is tied in soils as integral component of the structural units. Research in soil structure as reservoir of carbon and a large sink for atmospheric CO_2 has been intensified in recent years because of the importance of these processes to climate change issues (Six, 2004).

2.2 Development and Dynamics of Soil Structure

The concept of soil structure encompasses a broad spectrum of spatial arrangement of primary soil particles (Hillel, 1998; Brady and Weill, 2008). At one extreme, we have *single-grained* structure characterized by random arrangement of sand and silt particles with little cohesion. Sand blown dunes and fresh deposits of loess are examples of single-grained structure. At the opposite extreme, we have large cohesive *massive* structure of clay with no discernible internal features. Between these two extremes, soil particles form *aggregated* structure made of quasi-stable particles of various shapes and sizes. Older soil physics literature and *U.S. National Soil Survey Handbook* (USDA-NRCS, 2007) classify the single-grained and massive soils as *structureless*. For this reason, the concept of soil structure is often used to describe soil aggregates. Soil structure is inherently a dynamic feature that undergoes continuous change. There are two main

processes of soil structural change: (a) processes that lead to increased bonding or aggregation of noncohesive particles and (b) processes that cause fragmentation of aggregates or massive blocks of cohesive materials (such as polyhedral peds of soil-containing clay). These processes can and do occur concurrently, and it is not usually easy to separate aggregation from fragmentation. In the remainder of this section, we will discuss these two processes in depth.

2.2.1 Soil Aggregation

Prior to the 1930s, the concept of soil structure was synonymous with soil texture. A great deal of effort went into correlating soil physical properties with particle-size distributions. One attractive feature of soil texture is that it changes very little with time and as a result it is a reproducible index. However, most primary soil particles do not exist unattached. The vigorous mechanical and chemical forces applied to break bonds prior to textural analysis attest that the concept of "primary particles" does not have much meaning in aggregated soils. At around 1930s, there was a growing interest in agglomeration of soil particles mostly with extensive research focused on flocculation of clay particles to the extent that by mid-1940s soil aggregation became synonymous with flocculation. The following excerpt from a classic textbook *Soil Physics* by Baver (1940) describes the status of soil structure research at that time and what was about to happen afterward:

> Soil aggregation is often confused with flocculation. Numerous studies have tried to describe soil aggregate formation and stability based on observations of flocculation in dilute suspensions. These two are not synonymous. Flocculation is purely chemical process, and stability of the floccules is dependent on the presence of the flocculating agent. Stable-aggregate formation (granulation) requires that the primary particles be so firmly held together that they do not disperse in water. This implies that aggregation requires some form of cementation of the flocculated particles. In the words of Bradfield (1936) 'Granulation is cementation plus!'

For an extensive account of the historical development of the concept of soil structure, consult recent reviews by Warkentin (2008) and Six (2004).

The current conceptual model of soil structure describes soil aggregation as a hierarchical arrangement of particles. First introduced by Tisdall and Oades (1982), the model has been refined continuously by several researchers, including Oades (1984), Hadas (1987), Dexter (1988), and Six (2004). The hierarchies represent size and nature of particles of their respective constituents, nature of the bonds that form them, as well as temporal sequence of their creation. A schematic diagram depicting the hierarchical model is shown in Figure 2.1. The lowest order in the hierarchy involves bonding of primary clay particles into clay domains. Bonding between clay domains and/or with fine silt particles results in clusters. Aggregation of clusters creates
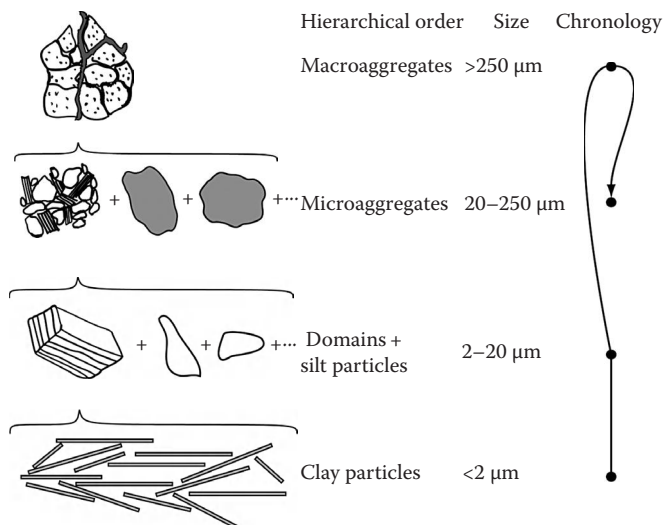
FIGURE 2.1 Hierarchical conceptual model of soil aggregation. (Modified from Dexter, A.R. 1988. Advances in characterization of soil structure. Soil Till. Res. 11:199–238; Six, J. 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil Till. Res. 79:7–31.)

microaggregates and bonding of microaggregates results in macroaggregates. In general, each particle of a given hierarchical order is made of 10^2–10^4 particles of the next lower hierarchical order. It is not necessary to have all the orders in every soil. The full spectrum of the hierarchical orders is likely to be present in soils that are rich in organic matter (Oades and Waters, 1991) for reasons that will become apparent in Section 2.2.1.6. Particles of lower hierarchical order are typically denser and stronger than higher order particles (Dexter, 1988). This is due to the fact the smaller order particles do not contain the larger (which are also weaker) pores that exist in higher order particles. An inverse correlation between tensile strength and aggregate size in a given soil was shown by Braunack et al. (1979) and Hadas (1987).

The dynamics of soil aggregation has two basic components: (a) *agglomerating processes* that bring soil particles to close proximity and (b) *binding agents* of organic and inorganic origins that render the inter particle linkages semipermanent. Development of the various hierarchical orders through the interplay between the agglomerating processes and binding agents is described later.

2.2.1.1 Clay Domains (<2 μm)

The basic building blocks of clays are aluminosilicate platelets that have a thickness of a few nanometers and lateral dimensions of several hindered nanometers (Rouse et al., 2004). Electrostatic bonding of these clay platelets (flocculation) forms clay domains, the lowest unit in the hierarchical order of soil aggregation. When clay platelets are brought close to one another, both attractive and repulsive forces can emerge. The repulsive forces arise because the ions that are attached to clay surfaces have identical (positive) charges. The attractive forces have several origins. Clay particles can be attracted to each other when the distance between them becomes so short that they can share the positive ions. This form of attraction is highly dependent on the valency

of the ions. Monovalent ions typically occupy large volumes when hydrated, resulting in very large separation between clay particles. Therefore, clay particles in contact with solution of monovalent ions exist in a *dispersed* state (opposite of flocculated state). In contrast, divalent cations have small hydrated volumes and can be readily shared by adjacent clay particles resulting in a stable and compact flocculation. When clay particles are in contact with acidic solution, the edges of the clay particles develop positive charges. In this condition, the edges of clay particles are attracted to faces of other clay particles, which are typically negatively charged. The face–edge attractions are weaker and more open than the face–face attractions mediated by divalent ions. In general, flocculation is a complex physical–chemical process that results in formation of stable clay domains through the binding of clay platelets by electrostatic attractions. The tendency for clay particles to flocculate is expressed in terms of critical coagulation concentration (CCC)—the minimum concentration of an electrolyte to initiate flocculation of dispersed clay suspension (Hsu and Tseng, 1996; Saejiew et al., 2004; Hsu et al., 2006; Liu et al., 2009). Advanced mathematical and experimental treatment of the formation and stability of clay domains can be found in Smalley (2006). The CCC of soils depends on the type of dominant clay minerals, organic matter content, pH, and cation exchange capacity (Frenkel et al., 1992; Heil and Sposito, 1993a, 1993b; Heil and Sposito, 1995).

2.2.1.2 Clusters (2–20 µm)

Clusters are formed through the bonding of clay domains and/or silt particles. This hierarchical order is so far the least understood and least studied. Clusters are predominantly formed when sand and silt particles are bonded together with colloidal particles, including clay domains, iron and aluminum oxides, and organic colloids (Baver, 1940). Soils that lack colloidal constituents typically do not form aggregated soil structure. Baver (1940) noted that there is a strong correlation between the concentration of 5 µm clay domains and aggregates larger than 50 µm. Baver also noted that the correlation between the concentration of 5 µm clay domains and aggregates larger than 100 µm is much weaker, suggesting that bonding by colloidal clay particles plays important role only at ~50 µm or below. When the mixtures contained organic matter, the correlations between colloidal clay and aggregation were weaker for both >50 and >100 µm aggregate sizes, indicating that colloidal clays play smaller role in the presence of colloidal organic matter. Colloidal organic matter such as mucilages exuded from soil microorganisms can bond clay clusters together and form very stable clusters (Turchenek and Oades, 1979). The bonds between the granular particles are formed upon partially irreversible desiccation of the colloidal particles that act as hardened glue. Sideri (1936) has shown the irreversibility of the bonds between desiccated clay and sand grains. In general, the strength of these bonds depends on the relative abundance of the colloidal particles in relation to silt particles, the size and chemical nature of the colloidal particles as well as wetting and drying history of the bonds.

2.2.1.3 Microaggregates (20–250 µm)

Microaggregates are formed when clusters and silt particles are bound by persistent binding agents (i.e., humified organic matter and polyvalent metal cation complexes), oxides, and highly disordered aluminosilicates (Six, 2004). In the original introduction of the hierarchical conceptual model of soil structure (Figure 2.1), Tisdall and Oades (1982) implied that the transition from lower to higher hierarchical orders is equivalent to chronological order (Six, 2004). Oades (1984) modified this notion and suggested that microaggregates are formed within macroaggregates through the process of fragmentation (see Section 2.2.3; Dexter, 1988; Six, 2004).

2.2.1.4 Macroaggregates (>250 µm)

Macroaggregates are formed from microaggregates that are bound together by weak binding agents such as the hyphae of roots and fungi as well as labile organic compounds derived from exudates of soil microorganisms and roots (Six, 2004). As a result, the strength, porosity, and internal hydraulic conductivity of macroaggregates are strongly influenced by soil management practices. In addition, there is considerable variability of physical properties within individual macroaggregates. Park and Smucker (2005a, 2005b) compared the porosity, hydraulic conductivity, and erosive strength of whole macroaggregates in the size range of 2–9.5 mm and concentric layers within individual macroaggregates. The aggregates used in the study were derived from two soils subjected to three different management practices, namely conventional tillage, no tillage, and native forest. Their findings show that porosity and hydraulic conductivity of whole macroaggregates increase with increasing size, whereas erosive strength decreases with increasing size. Within individual aggregates, the outer layers are more porous, more conductive, and less strong than the inner layers. In addition, they also showed that intensity of soil management decreases porosity, hydraulic conductivity, and erosive resistance.

2.2.1.5 Clods (>25 mm)

Clods are large lumps of soil that may or may not contain discernible particles of lower hierarchical order. In agricultural soils, clods may be formed due to compaction by agricultural machinery. Depending on the state of the soil wetness during compaction and the load of the machinery, the boundaries between macroaggregates may be totally lost or remain as microcracks. The ultimate goal of tillage in such soils is to reverse the creation of clods by liberating the aggregates "imprisoned" within clods (Dexter, 1988). Clods also represent large masses of soil created during desiccation of clayey soils. In this case, the clods need not contain particles of lower hierarchical order (Dexter, 1988). These clods (or peds) and their formation process are of great importance, particularly when dealing with subsoils. Subsoils typically do not have aggregates but show a similarly complex structure. This structure is the major concern when it comes to subsoil compaction since damage to subsoil structure is, in contrast to topsoil (aggregate) structure, considered almost irreversible.

2.2.1.6 Applications of the Hierarchical Model of Soil Structure

The hierarchical soil aggregation is an elegant and compact conceptual model for describing the soil aggregation process across multiple spatial and temporal scales. It has also been used creatively to elucidate some complex features of soil, including (a) the nature of organic matter lost by soil disturbance and (b) multiscale heterogeneity of soil pore space. These are described below.

According to the hierarchical model, microaggregates are formed from clusters that are bound together by relatively stable organic and inorganic colloidal particles. Whereas macroaggregates are formed when microaggregates are held together by weak organic bonds such as hyphae and labile organic matter. Therefore, mechanical soil disturbance (such as tillage) typically degrades the weak links between microaggregates and results in loss of the relative abundance of macroaggregates at the expense of increasing the abundance of "liberated" microaggregates (Oades and Waters, 1991). This state is close to the ideal soil *tilth* because (a) the fine pores within the microaggregates act as reservoirs of water for the seeds, (b) the voids between the microaggregates provide access to oxygen, and (c) the weak bonds between the aggregates exert minimal resistance to seedling establishment (Hadas, 1997). Because microaggregates are formed within macroaggregates (Oades, 1984), the creation of the ideal soil *tilth* is sustainable unless the labile organic matter released during cultivation is continuously replenished. Elliot (1986) used these concepts in describing the disproportionately high loss of labile organic matter during cultivation and his results were instrumental in establishing direct link between cultivation, loss of labile organic matter, and decreased aggregation (Six, 2004).

The hierarchical model of soil aggregation implies the existence of hierarchical pore categories. Elliott and Coleman (1988) proposed the following pore categories: (a) macropores, (b) intermacroaggregate pores, (c) intramacroaggregate or intermicroaggregate pores, and (d) intramicroaggregate pores. Such categorization can be useful in explaining spatial segregation of soil microorganisms. These pore categories house (a) microarthropods; (b) nematodes; (c) protozoa, small nematodes, and fungi; and (d) bacteria, respectively (Six, 2004). These concepts are very powerful because they can also be used to introduce spatial segregation of biochemical processes. For soil hydrologic processes, a slightly different pore categorization is employed and this includes (a) micropores, (b) capillary pores, and (c) macropores (Hillel, 1998). Micropores are generally pores that are less than 1 μm in diameter. Water in such pores is tightly held by adsorptive forces (Tuller et al., 1999) and its mobility is very restricted (Tuller and Or, 2001). Capillary pores range in width from several micrometers to a few millimeters (Hillel, 1998) and hold water that is subject to capillary laws and moves in accordance with Darcy and Buckingham–Darcy laws. The water held in these pores represents the bulk of plant available and mobile water. Macropores are several millimeters to a few centimeters in diameter and may be formed by burrowing organisms and/or

cracking during clay shrinkage. The sum of micropores and capillary pores can also be categorized as interaggregate and intraaggregate pores (Hillel, 1998; Or et al., 2000). These categorizations of soil pores are represented in flow and transport models using multiple overlapping continua. Dual-porosity and/or dual-permeability models, which separate the interaggregate and intraaggregate pores, are the most common forms of such models (Durner, 1994; Nimmo, 1997).

2.2.2 Factors Affecting Soil Aggregation

The formation and stabilization of soil aggregates are influenced by a number of biotic and abiotic factors that operate in concert. Oftentimes, it is difficult to distinguish the role of any particular factor from other factors.

2.2.2.1 Plants and Microorganisms

Soil microorganisms are largely supported by plants and their effect on soil aggregation cannot be viewed separately from the effect of plant roots. The interaction between plants and soil structure has been reviewed by Angers and Caron (1998), who recognized five classes of plants effects on soil structure.

2.2.2.1.1 Mechanical Effects of Root Penetration

When roots grow in soil, the volume occupied by the roots is accommodated by loss of an equal volume of pore space from the surrounding soil (Dexter, 1987b). Most of this displacement occurs at the root tips, which osmoregulate their internal stresses to counter the mechanical strength of the soil that they have to penetrate. The maximum stress that roots can exert ranges from 1 (Dexter, 1987a) to 2 MPa (Goss, 1991). The resulting densification can be modeled as an expansion of a cylindrical or spherical solid object within a homogenous continuum of deformable material (Farrell and Greacen, 1966; Greacen et al., 1968). These models, which were in good agreement with independent experimental observations, show that the density increases exponential from the undisturbed value at a distance away from the root to a maximum value on the root surface. Based on these observations, Dexter (1987a) proposed an empirical model that describes the soil porosity variation over short distances around plant roots:

$$\frac{\phi(r) - \phi_o}{\phi_i - \phi_o} = 1 - \exp\left(-k\frac{r - r_o}{r_o}\right), \qquad (2.1)$$

where
   ϕ is porosity
   r is radial distance measured outward from the center of a
      cylindrical root
   k is an empirical constant that ranges between 0.3 and 0.7
      based on two soils and root radii (Dexter, 1987b)

The subscripts o and i denote values at the root surface and far away from the root, respectively. Whereas these effects of

localized compaction reduce overall infiltration when the roots are living, the nearly vertical macropores left behind and stabilized by decaying roots greatly enhance infiltration (Barley, 1954). Plants with large-diameter, long, and straight roots (such as alfalfa) are known to promote macropore infiltration. The effect of roots on enhanced macropore flow increases with the number of years under plants such as alfalfa with minimal external disturbance (Meek et al., 1989, 1990).

2.2.2.1.2 *Moisture Dynamics Induced by Plants*

Wetting and drying cycles promote soil aggregation as well as fragmentation (see Section 2.2.2.4). Plant roots facilitate this effect by accentuating the wetting–drying cycles. Especially, in soils that exhibit shrink–swell behavior, the pattern and degree of cracking are strongly correlated with the distribution of the root system, spatial pattern of the plants (e.g., row crops vs. random distribution), and the biomass productivity of the vegetation (see Angers and Caron, 1998 and citations therein). Plant roots also promote small-scale moisture heterogeneities, which are efficient in creating small-scale stress imbalances. These stress imbalances can be very instrumental in the creation of microcracks and failure planes within macroaggregates and clods, thereby further promoting soil aggregation. Wetting and drying cycles induced by root hairs and fine roots were shown to cause orientated and compacted clay particles in microenvironments that are 50–200 μm in size (Dorioz et al., 1993).

2.2.2.1.3 *Root Exudates*

Roots promote the creation of stable soil aggregation in their immediate vicinity (rhizosphere) by releasing polymers and nutrient ions that (a) bind soil particles and microaggregates together or (b) support microbial life that generates complex and stable polymers and/or hyphae. Morel et al. (1991) has demonstrated that adding sticky mucilages exuded by maize roots to soil particles promotes instantaneous aggregation. However, these direct effects are typically transient not lasting much longer than 1 month. Consumption of the mucilages by soil microbes that release complex polysaccharides is considered one mechanism that prolongs the effect of root mucilages. In addition, the carbon and nutrients released by roots of many plant species support vesicular–arbuscular mycorrhiza (VAM), whose short hyphae are efficient in binding soil particles together and forming stable soil aggregates (Tisdall and Oades, 1979; Thomas et al., 1986).

2.2.2.1.4 *Carbon Inputs*

In addition to the short-term direct effects of plant roots discussed earlier, plant roots and litter promote long-term soil aggregation and stabilization through continuous supply of organic matter. It is estimated that around 20% of carbon assimilated by photosynthesis is deposited in the soil in the form of exudates and dead roots (Gregory, 2006). Consumption of this organic matter by soil fauna and microbes releases recalcitrant and complex organic molecules that promote formation of stable soil aggregates (Angers and Caron, 1998; Six, 2004).

2.2.2.1.5 *Entanglement by Roots*

Fine and coarse roots can increase stability of soil aggregates by physically entangling and enmeshing aggregates and particles of broad size range. Waldron et al. (1983) conducted shear tests on cylindrical soil columns of 1.2 m diameter—with fully developed root systems of alfalfa and yellow pine. Their experiments showed that shear resistance of the rooted columns was significantly higher than identical columns that were left bare. These large-scale effects are indeed exploited in soil stabilization on sloppy areas using herbaceous plants and trees. Similar effects also operate at much smaller scale. Tisdall and Oades (1982) showed that millimeter-size aggregates are formed and stabilized by the direct entanglement effects of fine roots and root hairs.

These effects of plant roots operate in concert, and it is usually very difficult to accurately distinguish the relative contribution of each effect toward soil aggregation. However, there is ample experimental evidence that demonstrates a positive correlation between vegetation and formation and stabilization of soil aggregates at multiple spatial and temporal scales (e.g., Angers and Caron, 1998; Six, 2004; Gregory, 2006).

2.2.2.2 Soil Fauna

Soil fauna plays an important role in the dynamics of soil aggregation by facilitating transport and breakdown of organic matter and forming burrows that enhance infiltration and fast moisture redistribution following rain or irrigation. The most influential soil faunas are perhaps earthworms. Earthworms are classified as (a) epigeic species that live in the litter layers of forest soils with minimal effect on soil aggregation, (b) anecic species that form and live in extensive burrows, and (c) endogenic species that feed on soil enriched with organic matter (Bouché, 1977; cited by Six, 2004). Networks of earthworm burrows, which are stabilized by the combined effects of compaction and mucous secretions (Edwards and Bohlen, 1996), form a system of fast-flow pathways that enhance infiltration and aeration. Earthworms also play a significant role in the formation and stabilization of soil aggregates by ingesting mixture of organic and inorganic soil materials and excreting the stable mixtures as cast. Numerous studies have shown that these casts are more stable than surrounding soil aggregates (e.g., Six, 2004 and references therein). The stability of earthworm casts depends on the degree of desiccation that they are subjected to, their age, and the nature of the organic matter incorporated in them by earthworms (Shipitalo and Protz, 1987, 1988; Shipitalo et al., 1988; Marinissen and Dexter, 1990). The amount of casting depends on feeding activity of the earthworms. Earthworms ingest larger volumes of soil when the organic content of the soil (food) is lower resulting in more casting. In general, it is estimated that the topsoil of temperate climates passes through the gut of earthworms every 4 years (Gardner, 1991).

2.2.2.3 Inorganic Binding Agents

Inorganic binding agents comprise polyvalent cations (calcium and magnesium) as well as colloidal clays and oxides. Calcium and magnesium operate at the lowest hierarchical order by promoting formation of stable clay domains. As such, their presence is very critical to the overall stability of soil structure at all the higher hierarchical orders. Calcium, in the form of lime, is a common form of soil amendment applied to reclaim dispersed sodic soils (Brady and Weill, 2008). In addition, calcium and magnesium promote microaggregation by forming clay–polyvalent cation–organic matter complexes (e.g., Baver, 1940; Clough and Skjemstad, 2000). Oxides of iron and aluminum can act as binding agents in three ways: (a) adsorption of organic compounds on oxide surfaces, (b) electrostatic attraction between positively charged oxides and negatively charged clay surfaces, and (c) formation of bridges between primary and secondary particles coated with oxides (El-Swaify and Emerson, 1975; Oades et al., 1989; Six et al., 2000). The first two processes act at the scale of microaggregates, whereas the latter process is responsible for stabilization of macroaggregates in oxide-rich soils (Six, 2004).

2.2.2.4 Wetting and Drying Cycles

Most soils experience continuous wetting and drying cycles as a consequence of a number of environmental and internal processes, including precipitation, irrigation, drainage, evaporation, and plant water uptake. These cycles induce soil aggregation and/or fragmentation through a number of processes, including swelling and shrinkage of clays, capillary stresses, as well as transport, deposition, and hardening of organic and inorganic binding agents.

2.2.2.4.1 Clay Swelling and Shrinkage

In soils consisting of substantial amounts of expansive clay fractions, one of the prominent effects of changing soil water content is the concurrent change of soil matrix volume. Nonuniform drying (hence, unbalanced shrinkage) creates internal stresses that eventually lead to cracking. In structureless heavy clay soils, formation of cracks that are several

centimeters wide and deep is a common phenomenon (e.g., Hillel, 1998; Brady and Weill, 2008). Similarly, in aggregated soils, unbalanced shrinkage creates weak planes and microcracks that lead to the formation of microaggregates within macroaggregates (Six, 2004). Swelling of aggregated soil (such as freshly tilled soil) results in joining of adjacent aggregates as they begin to invade one another space. Dexter and coworkers (unpublished data reported in Dexter, 1988) have shown that swelling significantly increases the tendency for spontaneous development of microaggregates in a thoroughly molded soil. In general, swelling and shrinkage can modify soil structure at multiple scales to varying degrees and/or directions (formation vs. destruction). A somewhat related process is freezing- and thawing-induced cracking of heavy clay soils. Farmers in cold regions typically plow clay-rich soils very coarsely. Subsequent freezing and thawing of the big clods create several microcracks and microaggregates (Scott, 2000).

2.2.2.4.2 Capillary Forces

Soils subjected to wetting and drying cycles are known to undergo structural changes by coalescence of adjacent aggregates. This process is primarily driven by strong tensile forces of capillary water (Keller, 1970; Silva, 1995; Or, 1996). Aggregate coalescence is responsible for post-tillage hardening of agricultural soils (Cockroft and Olsson, 2000). Micromorphological studies show that coalescence of aggregates occurs by plastic flattening and fusion of interaggregate contacts without any need for additional cementing agent (Day and Holmgren, 1952; Kwaad and Mucher, 1994). Plastic deformation occurs when the shear stress due to capillary forces exceeds the shear strength of the soil. Flattening of the contact faces was shown to greatly reduce the stresses, thereby restricting unlimited deformation. In an analogous experiment, where external triaxial stress was applied to mimic the effect of capillary forces, McMurdie and Day (1958) showed that soil aggregates behave in a manner similar to plastic grains. Figure 2.2 depicts aggregates that were fused together by capillary forces during multiple wetting and drying cycles. Ghezzehei and Or (2000) used modeled aggregate coalescence as a process driven by interplay between (a) energy liberation from contraction of capillary menisci and (b) energy

(a)   (b)   (c)
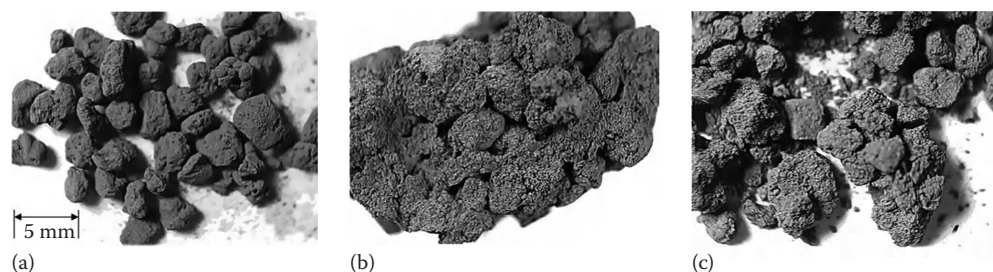
FIGURE 2.2   Coalescence of Millville silt loam soil aggregates (2–4 mm) under wetting–drying cycles. (a) Dry soil aggregates separated by mechanical sieving, (b) cluster of aggregates formed by coalescence due to wetting–drying cycles, and (c) partly crushed cluster of aggregates. Note that coalescence took place mainly at the contact points, and most of the aggregates maintained part of their original boundaries. (From Ghezzehei, T.A., and D. Or. 2000. Dynamics of soil aggregate coalescence governed by capillary and rheological processes. Water Resour. Res. 36:367–379. With permission.)

dissipation by viscous deformation of aggregate contacts. The model relies on the viscoelastic nature of wet soil aggregates (Ghezzehei and Or, 2001). An important conclusion derived from this model is the existence of a window of aggregate wetness where the capillary forces exceed the soil shear strength within which aggregate coalescence can occur. The window narrows with each incremental increase in the interaggregate contact area and completely disappears after some critical welding has occurred. Soil densification is traditionally viewed as deterioration of soil qualities that favor plant growth (Cockroft and Olsson, 2000). However, it is increasingly becoming apparent that loss of large interaggregate pores is accompanied by increased connectivity between aggregates leading to enhanced unsaturated hydraulic conductivity (e.g., Carminati et al., 2007a, 2007b, 2008; Berli et al., 2008).

2.2.2.4.3 *Alteration of Organic and Inorganic Colloids*

Slow wetting can promote stabilization of microaggregates by transporting dissolved organic carbon (produced during decomposition of organic matter and/or exuded by plant roots) to the interior of aggregates (Denef and Six, 2005). Drying facilitates localization of soil solutions in the fine crevices and wedges at interparticle contacts as pendular rings, where dissolved organic carbon and inorganic colloidal particles are deposited and form semipermanent bonds (e.g., Reid and Goss, 1982). The surfaces of microaggregates coated with dried organic matter are partially hydrophobic, which reduces their susceptibility to slaking (Caron et al., 1996; Czarnes et al., 2000a, 2000b). Park et al. (2007) compared the stability of aggregates subjected to five wetting–drying cycles with and without dissolved organic matter in the wetting solution. They observed that addition of $0.25\,\mathrm{mg\,g^{-1}}$ glucose soil during each wetting phase maintained the aggregate stability, while the aggregate stability deteriorated in the absence of additional organic-matter supply. Planted soils that were kept at constant water content were observed to have 10%–20% less aggregation than similar soils, which were allowed to periodically dry by plant uptake (Reid and Goss, 1982; Materechera et al., 1992, 1994). The foregoing observations indicate that wetting and drying cycles encourage aggregate stabilization by transporting and altering organic matter.

2.2.2.5 External Stresses

The primary objective of tillage of agricultural soils is to modify the structure so that the resulting morphological and hydrological properties favor optimal flow and storage of water and oxygen as well as minimal resistance to germination and root growth. Concurrently, tillage induces compaction and permanent destruction of inherent structure below the plow layer. A substantial portion of agricultural soil mechanics research has been devoted toward understanding and minimizing the unintended consequences of tillage. Soil structural dynamics as a result of external stresses is described in detail in Chapter 3.

2.2.3 Soil Fragmentation

Soil fragmentation is the flip side of aggregation and it begins with emergence of cracks within aggregates or massive structureless soil volume. Cracks originate in soil when the strain energy imposed by shrinking and swelling or tillage exceeds interparticle bonds (Hallett and Newson, 2001). Soil cracking is generally classified into brittle and ductile cracking. Brittle cracking is applicable for dry soils. Brittle cracking can be viewed as a two-stage process. First, strain energy that exceeds the interparticle bonds is needed to create failure planes. Once the planes are created, the growth of crack obeys a linear elastic law. In this case, the strain energy is stored in the reversible elastic crack.

For wet ductile soils, cracking is accompanied by substantial plastic deformation, which dissipates a significant portion of the strain energy. Unlike in brittle cracking, in ductile cracking the soil at the crack tip has the ability to flow and rearrange. This fluidity results in extra resistance to cracking, which is known as "toughness" in material science (Hallett and Newson, 2001). As a consequence, even if soil strength weakens upon wetting, the increase in toughness results in resistance to cracking. Therefore, both very wet and very dry soils have strong resistance to cracking. Somewhere in the middle, there is a wetness condition where the combined effect of strength and toughness is low and cracking of soils is the easiest. When the energy for cracking originates from shrinkage of clays, the energy also depends on wetness.

2.3 Characterization of Soil Structure

We started this chapter by noting the lack of precise and consistent definition of soil structure and the difficulties underlying this absence of precise definition are apparent from the discussions in Section 2.2. The implications of this vagueness are reflected in the methods employed for characterization of soil structure. In the remainder of this section, we describe four major types of soil structure characterization that are commonly employed in soil physics research. The first group of methods involves qualitative description of soil structure routinely used in soil survey. The second method is related to the morphological and hydrological soil properties that are influenced by soil structure. The third approach reflects the persistence of a given state of soil structure under disruptive forces. And finally, the fourth approach involves various imaging techniques.

2.3.1 Qualitative Description

Soil profile description, as part of soil survey and mapping, involves qualitative characterization of structure in terms of shape, size, and grade (distinctiveness). The terminology and designations used in *Soil Survey Manual* (Soil Survey Division Staff, 1993) are summarized in the following.
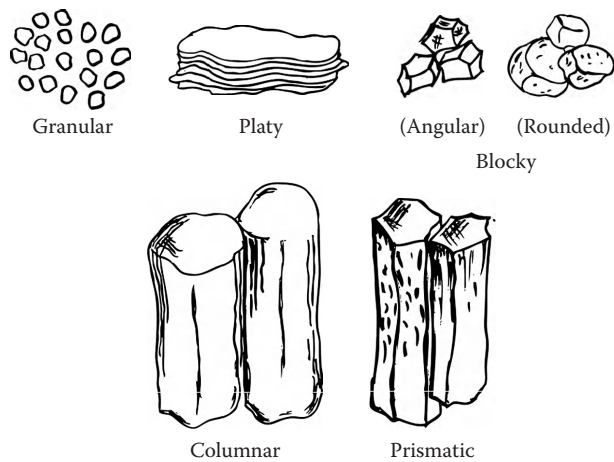
FIGURE 2.3 Shapes of soil structure based on the classification in *Soil Survey Manual*. (Adapted from Soil Survey Division Staff. 1993. Soil survey manual. Soil Conservation Service, USDA, Washington, DC.)

2.3.1.1 Shape

The following terms are used to describe the field observable shapes of individual structural units. Schematic illustrations of the various shapes (described below) are given in Figure 2.3.

Platy: Flat and horizontal structural units that commonly occur in clayey soils.

Prismatic: Elongated structural units with flat vertical faces that are matched with the faces of adjoining structural units. The vertices of the vertical faces are angular or subrounded and the tops are typically flat.

Columnar: These are similar to prismatic units, except that the tops are rounded and the vertical faces are also slightly rounded.

Blocky: Equidimensional and polyhedral structural units with flat or slightly rounded faces. The faces of these units are matched with the faces of adjoining units.

Granular: These are similar to blocky units, except that the units are more rounded and the faces are not matched with the faces of adjoining units.

2.3.1.2 Size

Five size classes are used to describe the size of soil structural units: very fine, fine, medium, coarse, and very coarse (Table 2.1). The size limits refer to the smallest dimension of plates, prisms, and columns.

2.3.1.3 Grade

Grade is used to describe the distinctness of individual structural units and the proportion of units that hold together when the soil is handled. Three grades are used:

Weak: It refers to barely observable structural units that disintegrate upon moderate handling.

Moderate: It refers to well-formed and readily recognizable structural units. Most of the units retain their integrity upon moderate handling/disturbance.

TABLE 2.1 Size Classes of Soil Structural Units

	Size Limits (mm)			
Size Class	Platy	Prismatic and Columnar	Blocky	Granular
Very fine	<1	<10	<5	<1
Fine	1–2	10–20	5–10	1–2
Medium	2–5	20–50	10–20	2–5
Coarse	5–10	50–100	20–50	5–10
Very coarse	>10	>100	>50	>10

Source: Soil Survey Division Staff. 1993. Soil survey manual. Soil Conservation Service, USDA, Washington, DC.

Strong: It refers to structural units that are distinctly observable and that separate cleanly when the soil is disturbed.

Qualitative description of soil structure in terms of shape, size, and grade is very valuable in understanding soil genesis as well as in classification and mapping soils. However, there is no objective means of translating these qualitative descriptions to quantifiable characteristics of soil structure and associated morphological and hydrologic properties.

2.3.2 Morphological and Hydrologic Properties

One of the main reasons we study soil structure is because it influences a number of morphological and hydrologic properties of soil that in turn determine storage and movement of various fluids, chemical cycling, and transport, as well as growth of plants and microbes. The most common morphological and hydrologic properties include porosity, bulk density, pore-size distribution, water retention characteristic, and saturated/unsaturated hydraulic conductivity. Determination of these properties requires extraction of *undisturbed* soil cores or better yet use of an in situ method. The notion of "undisturbed soil cores" and the great deal of care and effort invested in acquiring such samples are ultimately an indication of our recognition of soil structure—the organization and arrangement of soil particles—as an important component of soil physical properties.

Numerous attempts to infer the hydrologic properties using intrinsic soil properties (including particle-size distribution, organic-matter content, and soil solution chemistry) have had only limited success because they do not account for the structural component. The well-known method of Arya and Paris (1981) predicts water retention characteristic based on the knowledge of particle-size distribution, particle density, and bulk density. But this method works reasonably well only for soils that are not aggregated. For aggregated soils, empirical equations (also known as pedotransfer functions) are used to relate water retention characteristic and unsaturated hydraulic conductivity functions to soil constituents and descriptions collected during routine soil survey (Bouma and van Lanen, 1986; Schaap et al., 2001; McBratney et al., 2002). The method suffers from lack of physical basis for the prediction and its success greatly depends on the similarity of the database soils to the target soil.

Soil bulk density (ρ_b) is defined as the oven-dry mass of a unit volume of undisturbed soil. It is typically determined by measuring the volume and oven-dry mass of the same volume of soil. For most applications, a core of soil sample is extracted from the desired depth of the target soil using a metal cylinder. Density of soil macroaggregates and clods can be determined by measuring the mass of an individual clod and the volume it displaces when fully immersed in water (with a thin coat of wax to prevent water entry).

Porosity (φ) is a measure of the fraction of total soil volume occupied by pores. It is determined by measuring the total volume of a given soil sample and the volume of its pores. If the average density of the solids is known (which is typically around $\rho_p = 2650 \, kg \, m^{-3}$), then porosity can be readily determined from the bulk density as $\varphi = 1 - \rho_b/\rho_p$.

Water retention characteristic curve describes the amount of water that can be held in a soil by different degrees of capillary forces. Its determination involves concurrent measurement of water content and water potential (capillary pressure) at different degrees of wetness. While porosity and bulk density are measures of the total volume of pores, water retention curve is a measure of the pore-size distribution as well as wetting characteristics of pore surfaces. Water retention curve of structured soils captures the hierarchical nature of soil aggregation (and the resulting pore space). Soils with well-defined macroaggregates exhibit water retention curve with a discernible double-hump feature. Such distinction is represented by considering the pore system as a sum of two overlapping pore systems: one for the interaggregate pores and other for the intraggregate pores.

Hydraulic conductivity describes the water flux through soil to conduct water under unit total water potential gradient. It reflects not only pore-size distribution and wetting characteristics of pore surfaces, but also connectivity between pores. Direct determination of hydraulic conductivity involves measurement of flux density across a volume of soil subjected to known boundary water potentials. The value of hydraulic conductivity decreases by several orders of magnitude when a soil is drained. Therefore, determination of hydraulic conductivity of unsaturated soil is laborious and time consuming. Oftentimes, hydraulic conductivity of unsaturated soil is determined from water retention curve using models that represent soil pores using mathematically tractable representations. Of these, the most common model is that of van Genuchten (1980), which relies on a model of soil pores as bundle of capillary tubes (Mualem, 1976). Recent modifications of this indirect approach (Tuller and Or, 2001) recognize the angularity of soil pores and adsorption of thin films on clay surfaces and provide better predictive capability albeit at higher computational complexity. For more detailed discussion on soil–water relations, see Part I Chapters 4 and 5.

The above-described physical properties are dynamic in nature undergoing continuous change at multiple scales in time as illustrated in Figure 2.4. The data of Silva (1995) shown in Figure 2.4a were derived from columns of Millville silt loam soil subjected to wetting and drying cycles. Figure 2.4b depicts change in interaggregate porosity of Cotto clay soil under three tillage types (Rivadeneira, 1982). The interaggregate porosity, determined as the ratio of the total porosity for the three tillage treatments to the porosity for zero tillage, is plotted as a function of time after tillage. Change in total porosity is often accompanied by change in water retention curves. For the two illustrative examples shown in Figure 2.4, the corresponding evolution of water retention curve is shown in Figure 2.5. The striking feature of these data is that the comparable changes in porosity and water retention curve occurred within a few hours for the Millville silt loam soil while it took several weeks for the Cotto clay. The dynamics of the Millville silt loam soil was driven by capillary forces (Or et al., 2000) where as a complex interaction of various factors (including wetting and drying, plant roots, and microbial activity) may have contributed for the dynamics of the Cotto clay soil.

Because of the complexity of factors that are involved in the evolution of soil structure, the physical properties described
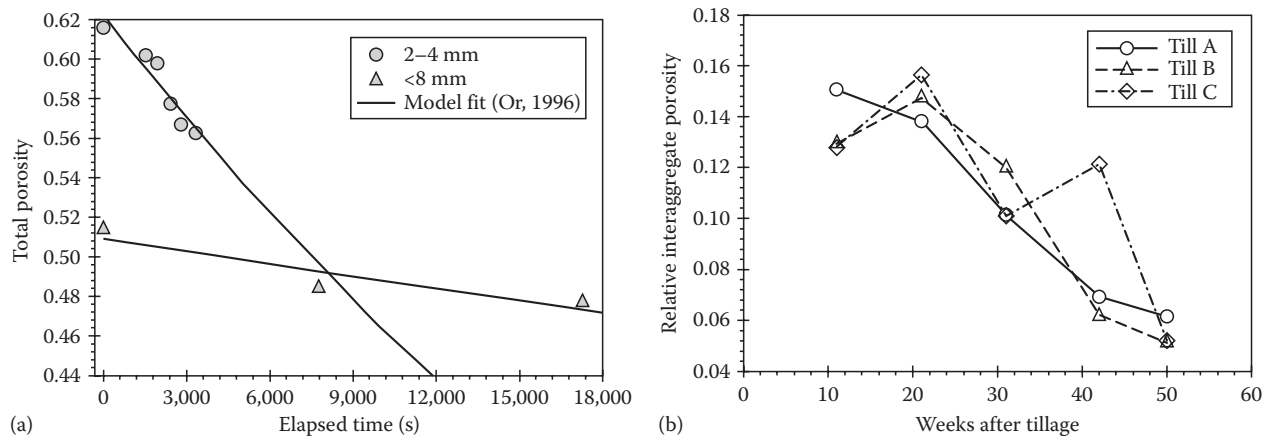
FIGURE 2.4 Rate of porosity loss, porosity measurements (a) versus elapsed time since initiation of wetting for two aggregate sizes (From Silva, H.R. 1995. Wetting-induced changes in near surface soil physical properties affecting surface irrigation. Ph.D. Dissertation. Utah State University, Logan, UT.) and relative interped porosity (b) for three different tillage practices versus weeks after tillage. (From Rivadeneira, J. 1982. Changes in physical properties in the plow layer of Coto soil as a function of tillage methods, and correlations with tanier development (in Spanish with English abstract). M.S. Thesis. University of Puerto Rico, Rio Piedras, Puerto Rico.)
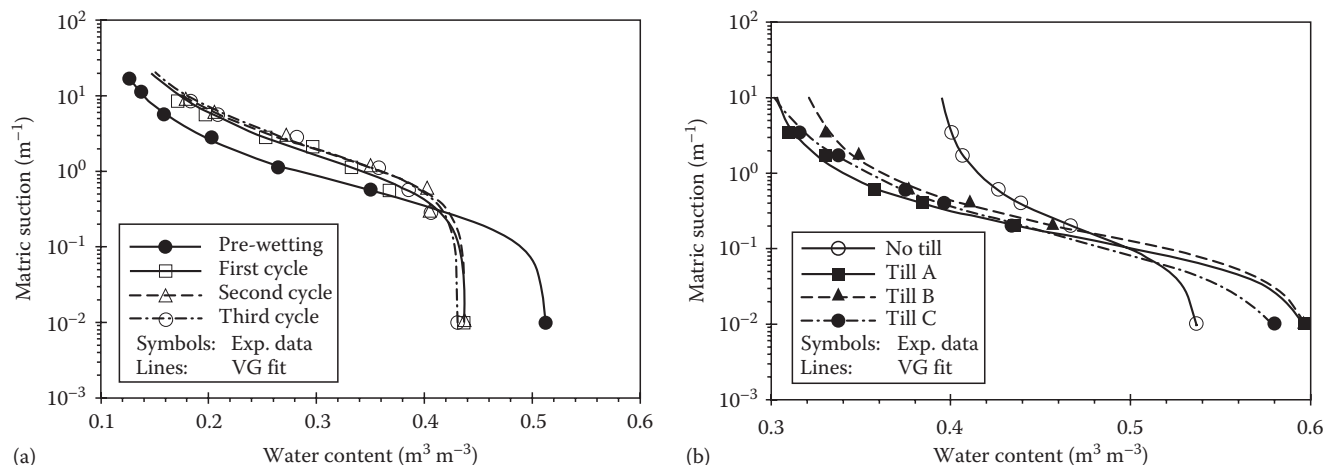
FIGURE 2.5 Comparison of evolution rates and shapes of soil water characteristic curves of two different soil types: (a) Millville silt loam soil before and after three cycles of wetting and drying in a few-hour timescale. (From Silva, H.R. 1995. Wetting-induced changes in near surface soil physical properties affecting surface irrigation. Ph.D. Dissertation. Utah State University, Logan, UT.) (b) Cotto clay soil under no till and three different types of tillage practices after 11 weeks of treatment. (From Rivadeneira, J. 1982. Changes in physical properties in the plow layer of Coto soil as a function of tillage methods, and correlations with tanier development (in Spanish with English abstract). M.S. Thesis. University of Puerto Rico, Rio Piedras, Puerto Rico.) Note that the tilled treatments in (b) correspond to the prewetting loose state in (a).

above are often treated as semipermanent characteristics of soil. Recently, there have been modest successes in predicting changes in soil structure and the associated physical properties. Ghezzehei and Or (2000) used energy-based analysis to develop a model that predicts the evolution of pore size during wetting and drying cycles. The model was subsequently used in a stochastic framework to predict evolution of pore-size distribution and water retention curves (Or et al., 2000). Figure 2.6 depicts evolution of the interaggregate pore space and the corresponding change in water retention after three cycles of wetting and drying. Here, only the interaggregate pore space is considered as dynamic as it directly depends on structure. The intraaggregate pore space is considered as constant as it depends mostly on the textural composition of the aggregates.

2.3.3 Stability of Soil Structure

Aggregate stability refers to the ability of aggregates to persist under internal and external disruptive forces. The disruptive force often impacts pores and/or aggregates of certain sizes only resulting in destruction of one or more hierarchical orders while concurrently forming another hierarchical order. The destruction of a certain hierarchical order is accompanied by destruction of all higher orders. Therefore, when discussing the stability of soil structure, it is very important to be specific at what order the destabilizing forces or processes operate (Dexter, 1988).

The reaction of soil aggregates to disruptive forces depends not only on the qualities of the soil itself but also on the nature of the disruptive forces. Therefore, the definition of aggregate
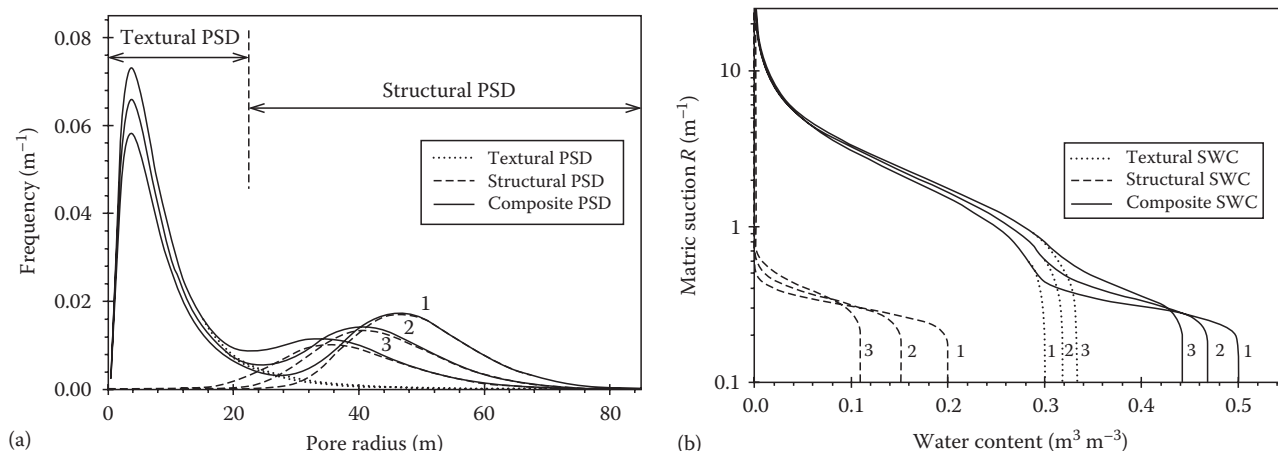
FIGURE 2.6 Evolution of (a) soil pore-size distributions and (b) soil water characteristic curves resulting from two cycles of wetting and drying (Figure 2.5a). Numbers represent the wetting–drying cycles (1 is initial; 2 is the end of first cycle; and 3 is the end of second cycle). Dashed lines represent the evolving structural (interaggregate) pore system.

stability and its characterization are linked with the disruptive forces of interest. In general, we can identify two main types of stability: (a) the ability of the soil to retain its structure under the action of water and (b) the ability of soil to retain its structure under the action of external mechanical stresses such as the impact of tillage implements and wheels (Dexter, 1988).

2.3.3.1 Water Stability

Water impacts soil aggregate stability in a number of ways. The initial action of water when aggregates are wetted is slaking, which refers to the disruption of soil aggregates by the action of entrapped air. During rapid wetting of dry aggregates, water enters each individual aggregate through its entire outer surface. This results in weakening of the outer layers of the aggregate, while concurrently the pressure of the air entrapped inside the aggregate rises. If and when the pressure of the entrapped air exceeds the strength of wetted outer layers of the aggregate, the air escapes by breaking the aggregate into smaller fragments as illustrated in Figure 2.7 (Emerson, 1954; Hillel, 1998). The violence of aggregate breakdown by slaking depends on the competition between air-pressure buildup and weakening of the soil strength upon wetting. For the entrapped air to gain pressure, it must be completely enclosed by saturated outer layers and must have sufficiently large starting volume so that subsequent wetting can result in the required gas compression. Emerson (1954) identified a number of factors that determine such pressure buildup. The primary factor is the rate of wetting, which in turn depends on the manner and source of wetting, the surface tension of the wetting liquid, and wettability of the aggregates. For example, aggregates with high-organic-matter content are less prone to slaking because the slight hydrophobicity of organic residues prevents rapid and even wetting of their outer surfaces. Another factor is the initial water content of the aggregates. Moist aggregates are typically less susceptible to slaking because they have smaller portion of their voids occupied by air as well as because preexisting continuous water films lead to diffuse wetting front as opposed to wet thin skin that forms upon wetting of very dry aggregates. Aggregates do not slake during rapid wetting unless the initial water potential is lower than $-1\,\text{MPa}$ (Grant and Dexter, 1986). Slaking produces aggregates that are larger than $250\,\mu\text{m}$ (Oades, 1986). If the air pressure is strong enough to break down the aggregates, slaking produces microcracks that are oriented parallel to the wetting front (McKenzie and Dexter, 1985).

Dry aggregate   Rapid wetting   Slaking
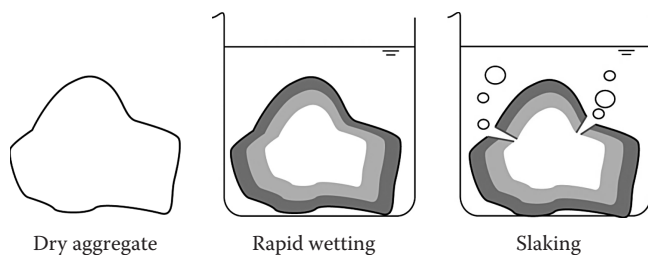
FIGURE 2.7 Slaking of an aggregate rapidly immersed in water. (Modified from Hillel, D. 1998. Environmental soil physics. Academic Press, San Diego, CA.)

Slow wetting can weaken aggregates by creating microcracks. Microcracks can be formed due to differential swelling of soil aggregates, which in turn can be caused by nonuniform wetting of individual aggregates and/or internal heterogeneities. The microcracks do not always join and result in breakdown of the aggregates. However, the creation of a network of microcracks results in considerable reduction in strength (Dexter et al., 1984) making the aggregate susceptible to disruption by other forces. Microcracks induced by the stresses generated by differential swelling on wetting are oriented mainly perpendicular to the wetting front (McKenzie and Dexter, 1985).

On uncovered ground surface, the kinetic energy of raindrop and flowing water can break down aggregates and transport the detached particles. Abrasion by particles suspended in flowing water provides additional scouring power that facilitates the disintegration of wet aggregates (Hillel, 1998).

Characterization of water stability of soil aggregates is conducted by simulating the processes that occur in nature. The most widely employed characterization method simulates the action of flowing water (Kemper and Rosenau, 1986). In this method, a sample of soil aggregates is placed on top of a nest of sieves. The sieves are immersed in water and shaken in a regular up-and-down motion at prescribed rate and vertical distance. The aggregates retained in each of the sieves after a fixed period of shaking are collected and dried in oven. Kemper and Rosenau (1986) suggest starting with 4 g of air-dry aggregates with 1–2 mm diameter and subject them to wet sieving at a rate of 35 cycles min^{-1} and vertical displacement of 1.3 cm for 3 min. Aggregates subjected to this treatment typically prewetted slowly to avoid slaking. Wet sieving provides a distribution of aggregate sizes. Van Bavel (1949) proposed the mean weight diameter (MWD) as a single-valued index of aggregate stability:

$$\text{MWD} = \int\limits_{0}^{x_{\max}} x f(x) dx, \tag{2.2}$$

where

  x is aggregate size
  $x_{\max}$ is the maximum aggregate size prior to wet sieving
  $f(x)$ is the empirical aggregate-size distribution

Before the availability of digital computers, the integration had to be done manually using a graphing paper. To simplify the time-consuming process, Youker and McGuinness (1956) proposed the use of summation in place of the definite integral:

$$\text{MWD} = \sum_{i=1}^{n} \overline{x}_i w_i, \tag{2.3}$$

where

  $i = 1, 2, \ldots, n$ is the index of sieves
  $\overline{x}_i$ is the mean diameter of the ith size class
  w_i is the corresponding weight fraction

MWD computed by summation is typically larger than the value computed by integration, especially if the number of the sieves is as small as the typical nest of five sieves (Kemper and Rosenau, 1986). This index of aggregate stability presupposes that the aggregate sizes are normally distributed. However, aggregate-size distributions reported by Van Bavel (1949) were skewed and Gardner (1956) proposed the use of a log-normal distribution based on additional data. Mazurak (1950) suggested the geometric mean diameter (GMD), which is better at representing a log-normal distribution than MWD:

$$\text{GMD} = \frac{\sum_{i=1}^{n} \log(\overline{x}_i) w_i}{\sum_{i=1}^{n} w_i}. \tag{2.4}$$

The GMD is getting wider use only recently (e.g., Abid and Lal, 2008) because of the extensive manual calculation it involved prior to the age of digital computers (Kemper and Rosenau, 1986).

Stability of soil aggregates under the disruptive energy of raindrops is determined by subjecting individual aggregates to drops in a standardized manner (Hillel, 1998). The fraction of the aggregate that remains intact after a prescribed duration or the number of drops required to completely destroy the aggregate are used as indices of the stability of the aggregate.

2.3.3.2 Mechanical Stability

Mechanical stability refers to resistance of aggregates to external forces such as tillage implements and wind. We can recognize at least three different types of external forces: tensile, compressive, and shear. Tensile forces applied to aggregates of a certain hierarchical order tend to cause failure surfaces between the particles and/or aggregates of the next lower hierarchical order while leaving those low-order aggregates intact (Dexter, 1988). This is generally because internal strength decreases with increasing hierarchical order (Braunack et al., 1979; Dexter, 1988).

Shearing of aggregated soil results in complex behavior that depends on the water content and density of the aggregates. If the density of the structured soil is below a characteristic critical value, the theory of critical state soil mechanics (Kurtay and Reece, 1970; Hettiaratchi and Ocallaghan, 1980) suggests that the stress propagates through the volume of the soil resulting in overall density reduction with disproportionate destruction of larger pores (Dexter, 1988). In contrast, if the structured soil is denser than the critical value, shearing is confined to a narrow band near the source of the stress, rolling of round particles or alignment of platy clay particles can occur. In the latter case, the gross changes in volume occur only in a small proportion of the soil.

Compression (uniaxial or multiaxial) typically results in densification that can extend throughout the volume of the soil. Dexter (1988) recognizes three types of response to uniaxial compression of aggregated soils. Hard (typically dry) aggregates respond to uniaxial compression by rearrangement and frictional attrition at their contacts that leads to overall densification. If the stress is large enough to cause brittle failure of individual hard aggregates, some aggregates are fractured and fragments released from the cracks fill larger voids. If the aggregates are soft, the aggregates deform at their contacts (points of maximum stress) resulting in flattening of contacts (Day and Holmgren, 1952; McMurdie and Day, 1958; Ghezzehei and Or, 2003a, 2003b). Similar flattening of contacts in response to internal capillary forces was proposed by Ghezzehei and Or (2000). The extent and rate of contact flattening (aggregate coalescence) can be related to the rheology of the soil forming the aggregates. Ghezzehei and Or (2000, 2001) showed that wet soils have viscoelastic behavior with the yield stress and apparent viscosity values that decrease with increasing water content.

2.3.4 Imaging Techniques

Imaging provides a direct and objective means of characterizing soil structure. Unlike the above three approaches, it can provide both qualitative and quantitative descriptions. The available imaging methods can be broadly classified into three groups: (a) thin-section micrography, (b) scanning electron microscopy (SEM), and (c) x-ray computed tomography (CT).

2.3.4.1 Thin-Section Micrography

The key developments that led to micrographic methods in soil research were thin-sectioning methods and petrographic microscopy in early twentieth century. The publication of the book *Micropedology* (Kubiena, 1938) introduced standardization of the methods and interpretation on micrographic studies of soil. The most important step of this method is the elaborate and time-consuming process of preparing thin sections for microscopy. For comprehensive description of procedures, consult the *Handbook for Soil Thin Section Description* by Bullock et al. (1985). Briefly, the method involves impregnation of soil pores with low-viscosity resins (such as unsaturated polyester, epoxy, and methyl methacrylate). To ensure complete saturation, this process is often carried out using dry sample under vacuum. After the resin is allowed to harden, the sample is affixed to microscope slide and sliced into thin sections, followed by polishing to remove traces of cutting. The final thickness of the thin section is typically 20–30 μm. Pigmented resins are used to enhance contrast. Thin-section preparation is inherently susceptible to several defects that can make subsequent image interpretation difficult. These include scratches and striations, entrapment of bubbles, contamination by translocated particles, and/or cutting abrasives.

Soil micrograph can be used to directly determine pore sizes and pore-size distributions, pore geometry, as well as aggregate sizes and shapes. Day and coworkers (Day and Holmgren, 1952; McMurdie and Day, 1958; McMurdie, 1963) used soil micrography to visualize deformation and welding of soil aggregates and accompanying the evolution of interaggregate porosity. However, a single two-dimensional (2D) cross section does not provide a complete picture of the three-dimensional (3D)

features of soil structure. Some common problems of 2D images include the following: (a) a given pore can appear to have different sizes depending on where the cutting plane passes through, (b) a wavy tubular pore (such as earthworm cast) can intersect the cutting plane multiple times and appear as a series of circular pores, and (c) adjacent particles (or aggregates) may appear disconnected.

2.3.4.2 Scanning Electron Microscopy

Compared to optical microscopy, SEM provides far more superior resolution albeit for much smaller field of view. Therefore, the application of SEM has been mostly used to study the microstructure of clay domains and their associations with cations and organic matter (Chen and Banin, 1975; Chenu, 1993; Skjemstad et al., 1996). Recently, a variation of SEM that utilizes flash freezing (Cry-SEM) has been successfully used to visualize the microstructure of rhizosphere (Gregory, 2006; Refshauge et al., 2006; McCully et al., 2009). These methods are instrumental in further understanding the role of roots and mycorrhizae in the development of soil structure (see Section 2.2.1).

2.3.4.3 X-Ray Computed Tomography

The application of 2D x-ray imaging in soil structure has been limited because of its inability to resolve the compounded effect in the third dimension. For example, Farrell and Greacen (1966) and Greacen et al. (1968) used medical x-ray facility to study soil density variations around a penetrometer inserted into a uniform soil. These observations were later used to develop conceptual and quantitative model of soil–root interactions by Dexter (1987a, 1987b).

In contrast, x-ray CT was adopted for soil research in early 1980s, soon after the technique was developed for medical applications. A thorough review of the application of x-ray CT in soil science and underlying principles is given by Taina et al. (2008). The method relies on the attenuation of x-ray as it passes through a subject medium (such as soil). The degree of attenuation is correlated with the bulk density as well as the actual and/or effective atomic weights of the materials encountered along the path. By integrating multiple projections of the same cross-sectional plane, tomographic reconstruction produces an effective density map (gray-scale image) of the cross section. Because these images are acquired noninvasively, it is possible to obtain several images of adjacent thin layers. This method has allowed visualization of undisturbed soil structure and the associated pore network in full 3D. X-ray CT have been used to quantitatively characterize porosity, pore geometry (diameter, perimeter, and area), tortuosity, and 3D hydraulic radius (e.g., Anderson et al., 1990; Perret et al., 1999; Gantzer and Anderson, 2002; Rachman et al., 2005). This method has also been successfully used in investigations of soil structural evolution due to disruptive forces such as tillage and erosion (Rogasik et al., 1999; Gantzer and Anderson, 2002; Lipiec and Hatano, 2003). Attempts to reduce the massive 3D x-ray CT description of soil structure to concise and mathematically tractable form have been performed using mathematical morphology (e.g., Pierret et al., 2002) and mass fractals (Rachman et al., 2005).

A key step in interpretation and quantification of x-ray CT imagery is segmentation—the process of classifying the images into two or more types of regions (e.g., voids, solids, water, and roots). The simplest approach is binarization of the gray-scale image (e.g., into voids and solids) by setting a global threshold in the histogram of the gray scales. This method is prone to miscategorization due to overlapping gray values common to the two populations (Taina et al., 2008). Alternative approaches include locally adaptive thresholding, region-growing methods (that assume voxels belonging to an object are connected and similar), probabilistic fuzzy clustering, and Bayesian methods (see Iassonov et al., 2009 and reference therein). Iassonov et al. (2009) used 14 segmentation algorithms that fall under global and locally adaptive approaches to derive soil porosity using common CT images. These algorithms yielded in vastly varying results and only a few gave promising results suggesting the need of reliable, consistent, computationally efficient, and automated algorithms applicable to a wide range of porous materials.

2.4 Concluding Remarks

Soil structure is a term widely used in all branches of soil science. The concept of soil structure is well established and has been continuously evolving. However, a precise definition of soil structure has not been adopted so far. Current definitions in various fields of soil research are summarized here.

For an agronomist, it refers to the quality of soil that relates to crop growth and productivity and may refer to water-holding capacity, ease of tillage, and suitability for root growth and microbial life. The agronomy-oriented literature on soil structure is replete with empirical correlations between various soil amendments (including lime, organic matter, and synthetic polymers) and the resultant changes in soil properties that depend on structure (such as porosity, water retention, and hydraulic conductivity).

For a conservationist, soil structure refers to resistance to water and wind erosion. Soil erodibility, a measure of the inherent susceptibility of any given soil to erosion, is closely tied to the stability of soil aggregates. In fact, the concept of aggregate stability was originally introduced to characterize soil erodibility (Yoder, 1936). Most aggregate stability characterization methods are designed to determine the ease with which particles can detach from soil aggregates by raindrop impact, flowing water, and/or wind. But it is still difficult, if not totally impossible, to predict erodibility based on the knowledge of aggregate stability because the methods of characterization cannot replicate all aspects of the erosive forces (Bissonnais, 1996).

For biogeochemists and global-change researchers, soil structure is viewed in association with nutrient storage and cycling (van Veen and Kuikman, 1990; Juma, 1993; Ladd et al., 1996). Soil structure is not only a product of organic carbon, but also a venue where carbon is stored and preserved from decomposition and released to atmosphere. Soil structure controls storage and cycling of organic carbon through its influence on the pore sizes that can be inhabited by soil organisms

as well as availability of water and oxygen. Large quantities of readily decomposable organic matter can be found in the vicinity of starving microbial populations but remains inaccessible to decomposers through chemical and physical protection (van Veen and Kuikman, 1990).

A common theme that has been developed within these diverse points of view is that the concept of soil structure spans multiple scales in time and space. The processes and factors involved in forming structure vary considerably with spatial and temporal scales. Similarly, the properties and functions influenced by soil structure vary depending on the scale of observation. The basis for this multiscale concept is the hierarchical model of soil aggregation, which has been undergoing continuous refinement and elaboration since its inception by Tisdall and Oades (1982). However, the interrelationships between the various factors involved in the creation and stabilization of soil structure on one hand and the numerous soil properties that are directly determined by soil structure on the other hand are very complex and rarely quantified. To make the concept of soil structure a useful descriptive as well as predictive tool, there is a need for quantification of the hierarchical conceptual model and the various factors that it entails.

References

Abid, M., and R. Lal. 2008. Tillage and drainage impact on soil quality—I. Aggregate stability, carbon and nitrogen pools. Soil Till. Res. 100:89–98.

Anderson, S.H., R.L. Peyton, and C.J. Gantzer. 1990. Evaluation of constructed and natural soil macropores using X-ray computed-tomography. Geoderma 46:13–29.

Angers, D.A., and J. Caron. 1998. Plant-induced changes in soil structure: Processes and feedbacks. Biogeochemistry 42:55–72.

Arya, L.M., and J.F. Paris. 1981. A physicoempirical model to predict the soil-moisture characteristic from particle-size distribution and bulk-density data. Soil Sci. Soc. Am. J. 45:1023–1030.

Barley, K.P. 1954. Effects of root growth and decay on the permeability of a synthetic sandy loam. Soil Sci. 78:205–210.

Baver, L. 1940. Soil physics. John Wiley & Sons, London, U.K.

Berli, M., A. Carminati, T.A. Ghezzehei, and D. Or. 2008. Evolution of unsaturated hydraulic conductivity of aggregated soils due to compressive forces. Water Resour. Res. 44:W00C09.

Bissonnais, Y. 1996. Aggregate stability and assessment of soil crustability and erodibility: I. Theory and methodology. Eur. J. Soil Sci. 47:425–437.

Bouché, M.B. 1977. Stratégies lombriciennes, p. 122–132. *In* U. Lohm and E. Persson (eds.) Soil organisms as components of ecosystems. Swedish Natural Science Research Council, Stockholm, Sweden.

Bouma, J., and H.A.J. van Lanen. 1986. Transfer functions and threshold values: From soil characteristics to land qualities quantified land evaluation procedures. Proc. Int. Workshop on Quantified Land Eval. Procedures, Washington, DC.

Bradfield, R. 1936. The value and limitations of calcium in soil structure. Am. Soil Surv. Assoc. Bull. 17:31–32.

Brady, N., and R. Weill. 2008. The nature and properties of soil. 14th ed. Pearson Prentice Hall, Upper Saddle River, NJ.

Braunack, M., J. Hewitt, and A. Dexter. 1979. Brittle fracture of soil aggregates and the compaction of aggregate beds. J. Soil Sci. 30:653–667.

Bullock, P., N. Fedoroff, A. Jongerius, G. Stoops, and T. Tursina. 1985. Handbook for soil thin section description. Waine Research, Wolverhampton, U.K.

Carminati, A., A. Kaestner, H. Fluhler, P. Lehmann, D. Or, E. Lehmann, and M. Stampanoni. 2007a. Hydraulic contacts controlling water flow across porous grains. Phys. Rev. E 76:026311.

Carminati, A., A. Kaestner, O. Ippisch, A. Koliji, P. Lehmann, R. Hassanein, P. Vontobel, E. Lehmann, L. Laloui, L. Vulliet, and H. Fluhler. 2007b. Water flow between soil aggregates. Trans. Porous Media 68:219–236.

Carminati, A., A. Kaestner, P. Lehmann, and H. Fluhler. 2008. Unsaturated water flow across soil aggregate contacts. Adv. Water Resour. 31:1221–1232.

Caron, J., C.R. Espinolda, and D.A. Angers. 1996. Soil structural stability during rapid wetting: Influence of land use on some aggregate properties. Soil Sci. Soc. Am. J. 60:901–908.

Chen, Y., and A. Banin. 1975. Scanning electron-microscope (SEM) observations of soil structure changes induced by sodium–calcium exchange in relation to hydraulic conductivity. Soil Sci. 120:428–436.

Chenu, C. 1993. Clay polysaccharide or sand polysaccharide associations as models for the interface between microorganisms and soil: Water related properties and microstructure. Geoderma 56:143–156.

Clough, A., and J.O. Skjemstad. 2000. Physical and chemical protection of soil organic carbon in three agricultural soils with different contents of calcium carbonate. Aust. J. Soil Res. 38:1005–1016.

Cockroft, B., and K.A. Olsson. 2000. Degradation of soil structure due to coalescence of aggregates in no-till, no-traffic beds in irrigated crops. Aust. J. Soil Res. 38:61–70.

Czarnes, S., A.R. Dexter, and F. Bartoli. 2000a. Wetting and drying cycles in the maize rhizosphere under controlled conditions. Mechanics of the root-adhering soil. Plant Soil 221:253–271.

Czarnes, S., P.D. Hallett, A.G. Bengough, and I.M. Young. 2000b. Root- and microbial-derived mucilages affect soil structure and water transport. Eur. J. Soil Sci. 51:435–443.

Day, P.R., and G.G. Holmgren. 1952. Microscopic changes in soil structure during compression. Soil Sci. Soc. Am. Proc. 16:73–77.

Dexter, A.R. 1987a. Compression of soil around roots. Plant Soil 97:401–406.

Dexter, A.R. 1987b. Mechanics of root-growth. Plant Soil 98:303–312.

Dexter, A.R. 1988. Advances in characterization of soil structure. Soil Till. Res. 11:199–238.

Dexter, A.R., B. Kroesbergen, and H. Kuipers. 1984. Some mechanical-properties of aggregates of top soils from the Ijsselmeer polders. 2. Remoulded soil aggregates and the effects of wetting and drying cycles. Neth. J. Agr. Sci. 32:215–227.

Dorioz, J.M., M. Robert, and C. Chenu. 1993. The role of roots, fungi and bacteria on clay particle organization: An experimental approach. Geoderma 56:179–194.

Durner, W. 1994. Hydraulic conductivity estimation for soil with heterogeneous pore structure. Water Resour. Res. 30:211–223.

Edwards, C.A., and P.J. Bohlen. 1996. Biology and ecology of earthworms. Chapman & Hall, London, U.K.

Elliott, E.T., and D.C. Coleman. 1988. Let the soil work for us. Ecol. Bull. Natl. Speleol. Soc. 39:23–32.

El-Swaify, S.A., and W.W. Emerson. 1975. Changes in physical-properties of soil clays due to precipitated aluminum and iron hydroxides. I. Swelling and aggregate stability after drying. Soil Sci. Soc. Am. J. 39:1056–1063.

Emerson, W.W. 1954. The determination of the stability of soil crumbs. J. Soil Sci. 5:233–250.

Farrell, D., and E. Greacen. 1966. Resistance to penetration of fine probes in compressible soil. Aust. J. Soil Res. 4:1–17.

Frenkel, H., M.V. Fey, and G.J. Levy. 1992. Organic and inorganic anion effects on reference and soil clay critical flocculation concentration. Soil Sci. Soc. Am. J. 56:1762–1766.

Gantzer, C.J., and S.H. Anderson. 2002. Computed tomographic measurement of macroporosity in chisel-disk and no-tillage seedbeds. Soil Till. Res. 64:101–111.

Gardner, W.R. 1956. Representation of soil aggregate-size distribution by a logarithmic-normal distribution. Soil Sci. Soc. Am. Proc. 20:151–153.

Gardner, W.R. 1991. Soil science as a basic science. Soil Sci. 151:2–6.

Ghezzehei, T.A., and D. Or. 2000. Dynamics of soil aggregate coalescence governed by capillary and rheological processes. Water Resour. Res. 36:367–379.

Ghezzehei, T.A., and D. Or. 2001. Rheological properties of wet soils and clays under steady and oscillatory stresses. Soil Sci. Soc. Am. J. 65:624–637.

Ghezzehei, T.A., and D. Or. 2003a. Stress-induced volume reduction of isolated pores in wet soil. Water Resour. Res. 39:TNN1.1–TNN1.7.

Ghezzehei, T.A., and D. Or. 2003b. Pore-space dynamics in a soil aggregate bed under a static external load. Soil Sci. Soc. Am. J. 67:12–19.

Goss, M. 1991. Consequences of the activity of roots on soil, p. 161–186. *In* D. Atkinson (ed.) Plant root growth: An ecological perspective. Blackwell Science Inc., Oxford, U.K.

Grant, C.D., and A.R. Dexter. 1986. Soil structure generation by wetting and drying cycles, p. 60–62. Proc. 13th Congress, Int. Soc. Soil Sci. Vol. 2. Hamburg, Germany.

Greacen, E.L., D.A. Farrell, and B. Cockcroft. 1968. Soil resistance to metal probes and plant roots. Trans. 9th Int. Soc. Soil Sci. Congress, Adelaide 1:769–779.

Gregory, P. 2006. Roots, rhizosphere and soil: The route to a better understanding of soil science? Eur. J. Soil Sci. 57:2–12.

Hadas, A. 1987. Long-term tillage practice effects on soil aggregation modes and strength. Soil Sci. Soc. Am. J. 51:191–197.

Hadas, A. 1997. Soil tilth—The desired soil structural state obtained through proper soil fragmentation and reorientation processes. Soil Till. Res. 43:7–40.

Hallett, P.D., and T.A. Newson. 2001. A simple fracture mechanics approach for assessing ductile crack growth in soil. Soil Sci. Soc. Am. J. 65:1083–1088.

Heil, D., and G. Sposito. 1993a. Organic-matter role in illitic soil colloids flocculation I. Counter ions and pH. Soil Sci. Soc. Am. J. 57:1241–1246.

Heil, D., and G. Sposito. 1993b. Organic-matter role in illitic soil colloids flocculation. 2. Surface-charge. Soil Sci. Soc. Am. J. 57:1246–1253.

Heil, D., and G. Sposito. 1995. Organic-matter role in illitic soil colloids flocculation. 3. Scanning force microscopy. Soil Sci. Soc. Am. J. 59:266–269.

Hettiaratchi, D.R.P., and J.R. Ocallaghan. 1980. Mechanical-behavior of agricultural soils. J. Agr. Eng. Res. 25:239–259.

Hillel, D. 1998. Environmental soil physics. Academic Press, San Diego, CA.

Hsu, J.P., and M.T. Tseng. 1996. Critical coagulation concentration of counterions. AIChE J. 42:3567–3570.

Hsu, J.P., H.Y. Yu, and S. Tseng. 2006. Critical coagulation concentration of a salt-free colloidal dispersion. J. Phys. Chem. 110:7600–7604.

Iassonov, P., T. Gebrenegus, and M. Tuller. 2009. Segmentation of X-ray computed tomography images of porous materials: A crucial step for characterization and quantitative analysis of pore structures. Water Resour. Res. 45:W09415.

Juma, N.G. 1993. Interrelationships between soil structure/texture, soil biota/soil organic matter and crop production. Geoderma 57:3–30.

Keller, J. 1970. Sprinkler intensity and soil tilth. Trans. ASAE 13:118–125.

Kemper, W.D., and R.C. Rosenau. 1986. Aggregate stability and size distribution, p. 425–444. *In* A. Klute (ed.) Methods of soil analysis, Part 1, physical and mineralogical methods. 2nd ed. SSSA, Madison, WI.

Kubiena. 1938. Micropedology. Collegiate Press, Ames, IA.

Kurtay, T., and A.R. Reece. 1970. Plasticity theory and critical state soil mechanics. J. Terra-Mech. Cohesive-Frict. Mater. 7:23–56.

Kwaad, F., and H.J. Mucher. 1994. Degradation of soil-structure by welding—A micromorphological study. Catena 23:253–268.

Ladd, J.N., R.C. Foster, and J.M. Oades. 1996. Soil structure and biological activity, p. 23–78. *In* G. Stotzky and J.M. Bollag (eds.) Soil biochemistry. Marcel Dekker Inc., New York.

Lipiec, J., and R. Hatano. 2003. Quantification of compaction effects on soil physical properties and crop growth. Geoderma 116:107–136.

Liu, L., L. Moreno, and I. Neretnieks. 2009. A novel approach to determine the critical coagulation concentration of a colloidal dispersion with plate-like particles. Langmuir 25:688–697.

Marinissen, J.C.Y., and A.R. Dexter. 1990. Mechanisms of stabilization of earthworm casts and artificial casts. Biol. Fertil. Soils 9:163–167.

Materechera, S.A., A.R. Dexter, and A.M. Alston. 1992. Formation of aggregates by plant roots in homogenised soils. Plant Soil 142:69–79.

Materechera, S.A., J.M. Kirby, A.M. Alston, and A.R. Dexter. 1994. Modification of soil aggregation by watering regime and roots growing through beds of large aggregates. Plant Soil 160:57–66.

Mazurak, A.P. 1950. Aggregation of clay separates from bentonite, kaolin, and a hydrous-mica. Soil Sci. Soc. Am. Proc. 15:18–24.

McBratney, A.B., B. Minasny, S.R. Cattle, and R.W. Vervoort. 2002. From pedotransfer functions to soil inference systems. Geoderma 109:41–73.

McCully, M.E., M.J. Canny, and C.X. Huang. 2009. Cryo-scanning electron microscopy (CSEM) in the advancement of functional plant biology. Morphological and anatomical applications. Funct. Plant Biol. 36:97–124.

McKenzie, B.M., and A.R. Dexter. 1985. Mellowing and anisotropy induced by wetting of moulded soil samples. Aust. J. Soil Res. 23:37–47.

McMurdie, J. 1963. Some characteristics of the soil deformation process. Soil Sci. Soc. Am. Proc. 27:251–254.

McMurdie, J.L., and P. Day. 1958. Compression of soil by isotropic stress. Soil Sci. Soc. Am. Proc. 22:18–22.

Meek, B.D., W.R. Detar, D. Rolph, E.R. Rechel, and L.M. Carter. 1990. Infiltration-rate as affected by an alfalfa and no-till cotton cropping system. Soil Sci. Soc. Am. J. 54:505–508.

Meek, B.D., E.A. Rechel, L.M. Carter, and W.R. Detar. 1989. Changes in infiltration under alfalfa as influenced by time and wheel traffic. Soil Sci. Soc. Am. J. 53:238–241.

Morel, J.L., L. Habib, and S. Plantureux. 1991. Influence of maize root mucilage on soil aggregate stability. Plant Soil 136:111–119.

Mualem, Y. 1976. New model for predicting hydraulic conductivity of unsaturated porous-media. Water Resour. Res. 12:513–522.

Nimmo, J.R. 1997. Modeling structural influences on soil water retention. Soil Sci. Soc. Am. J. 61:712–719.

Oades, J.M. 1984. Soil organic-matter and structural stability—Mechanisms and implications for management. Plant Soil 76:319–337.

Oades, J.M. 1986. Associations of colloidal materials in soils, p. 660–674. Proc. 13th Congr., Int. Soil Sci. Soc. Vol. 6. Hamburg, Germany.

Oades, J.M., and A.G. Waters. 1991. Aggregate hierarchy in soils. Aust. J. Soil Res. 29:815–828.

Oades, J.M., G.P. Gillman, and G. Uehara. 1989. Interactions of soil organic matter and variable-charge clays, p. 69–95. *In* D.C. Coleman et al. (eds.) Dynamics of soil organic matter in tropical ecosystems. Hawaii Press, Honolulu, HI.

Or, D. 1996. Wetting-induced soil structural changes—The theory of liquid phase sintering. Water Resour. Res. 32:3041–3049.

Or, D., F.J. Leij, V. Snyder, and T.A. Ghezzehei. 2000. Stochastic model for post-tillage soil pore space evolution. Water Resour. Res. 36:1641–1652.

Park, E.J., and A.J.M. Smucker. 2005a. Erosive strengths of concentric regions within soil macroaggregates. Soil Sci. Soc. Am. J. 69:1912–1921.

Park, E.J., and A.J.M. Smucker. 2005b. Saturated hydraulic conductivity and porosity within macroaggregates modified by tillage. Soil Sci. Soc. Am. J. 69:38–45.

Park, E., W.J. Sul, and A.J.M. Smucker. 2007. Glucose additions to aggregates subjected to drying/wetting cycles promote carbon sequestration and aggregate stability. Soil Biol. Biochem. 39:2758–2768.

Perret, J., S.O. Prasher, A. Kantzas, and C. Langford. 1999. Three-dimensional quantification of macropore networks in undisturbed soil cores. Soil Sci. Soc. Am. J. 63:1530–1543.

Pierret, A., Y. Capowiez, L. Belzunces, and C.J. Moran. 2002. 3D reconstruction and quantification of macropores using X-ray computed tomography and image analysis. Geoderma 106:247–271.

Rachman, A., S.H. Anderson, and C.J. Gantzer. 2005. Computed-tomographic measurement of soil macroporosity parameters as affected by stiff-stemmed grass hedges. Soil Sci. Soc. Am. J. 69:1609–1616.

Refshauge, S., M. Watt, M.E. McCully, and C.X. Huang. 2006. Frozen in time: A new method using cryo-scanning electron microscopy to visualize root-fungal interactions. New Phytol. 172:369–374.

Reid, J.B., and M.J. Goss. 1982. Interactions between soil drying due to plant water-use and decreases in aggregate stability caused by maize roots. J. Soil Sci. 33:47–53.

Rivadeneira, J. 1982. Changes in physical properties in the plow layer of Coto soil as a function of tillage methods, and correlations with tanier development (in Spanish with English abstract). M.S. Thesis. University of Puerto Rico, Rio Piedras, Puerto Rico.

Rogasik, H., J.W. Crawford, O. Wendroth, I.M. Young, M. Joschko, and K. Ritz. 1999. Discrimination of soil phases by dual energy X-ray tomography. Soil Sci. Soc. Am. J. 63:741–751.

Rouse, J.H., S.T. White, and G.S. Ferguson. 2004. A method for imaging single clay platelets by scanning electron microscopy. Scanning 26:131–134.

Saejiew, A., O. Grunberger, S. Arunin, F. Favre, D. Tessier, and P. Boivin. 2004. Critical coagulation concentration of paddy soil clays in sodium-ferrous iron electrolyte. Soil Sci. Soc. Am. J. 68:789–794.

Schaap, M.G., F.J. Leij, and M.T. van Genuchten. 2001. ROSETTA: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. 251:163–176.

Shipitalo, M.J., and R. Protz. 1987. Comparison of morphology and porosity of a soil under conventional and zero tillage. Can. J. Soil Sci. 67:445.

Shipitalo, M.J., and R. Protz. 1988. Factors influencing the dispersibility of clay in worm casts. Soil Sci. Soc. Am. J. 52:764–769.

Shipitalo, M.J., R. Protz, and A.D. Tomlin. 1988. Effect of diet on the feeding and casting activity of *Lumbricus terrestris* and *Lumbricus rubellus* in laboratory culture. Soil Biol. Biochem. 20:233–237.

Sideri, D.I. 1936. On the formation of structure in soil II: Synthesis of aggregates, on the bonds uniting clay with sand and clay with humus. Soil Sci. 42:461–481.

Silva, H.R. 1995. Wetting-induced changes in near surface soil physical properties affecting surface irrigation. Ph.D. Dissertation. Utah State University, Logan, UT.

Six, J. 2004. A history of research on the link between (micro) aggregates, soil biota, and soil organic matter dynamics. Soil Till. Res. 79:7–31.

Six, J., E.T. Elliott, and K. Paustian. 2000. Soil structure and soil organic matter: II. A normalized stability index and the effect of mineralogy. Soil Sci. Soc. Am. J. 64:1042–1049.

Skjemstad, J.O., P. Clarke, J.A. Taylor, J.M. Oades, and S.G. McClure. 1996. The chemistry and nature of protected carbon in soil. Aust. J. Soil Res. 34:251–271.

Smalley, M. 2006. Clay swelling and colloid stability. Taylor & Francis, Boca Raton, FL.

Soil Science Society of America. 1996. Glossary of soil science terms. SSSA, Madison, WI.

Soil Survey Division Staff. 1993. Soil survey manual. Soil Conservation Service, USDA, Washington, DC.

Taina, I.A., R.J. Heck, and T.R. Elliot. 2008. Application of X-ray computed tomography to soil science: A literature review. Can. J. Soil Sci. 88:1–20.

Thomas, R.S., S. Dakessian, R.N. Ames, M.S. Brown, and G.J. Bethlenfalvay. 1986. Aggregation of a silty clay loam soil by mycorrhizal onion roots. Soil Sci. Soc. Am. J. 50:1494–1499.

Tisdall, J.M., and J.M. Oades. 1979. Stabilization of soil aggregates by the root systems of ryegrass. Aust. J. Soil Res. 17:429–441.

Tisdall, J.M., and J.M. Oades. 1982. Organic-matter and water-stable aggregates in soils. J. Soil Sci. 33:141–163.

Tuller, M., and D. Or. 2001. Hydraulic conductivity of variably saturated porous media: Film and corner flow in angular pore space. Water Resour. Res. 37:1257–1276.

Tuller, M., D. Or, and L.M. Dudley. 1999. Adsorption and capillary condensation in porous media: Liquid retention and interfacial configurations in angular pores. Water Resour. Res. 35:1949–1964.

Turchenek, L.W., and J.M. Oades. 1979. Fractionation of organomineral complexes by sedimentation and density techniques. Geoderma 21:311–343.

USDA-NRCS. 2007. National soil survey handbook, title 430-VI. USDA-NRCS, Washington, DC.

Van Bavel, C.H.M. 1949. Mean weight-diameter of soil aggregates as statistical index of aggregation. Soil. Sci. Soc. Am. Proc. 14:20–23.

van Genuchten, M.T. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44:892–898.

van Veen, J.A., and P.J. Kuikman. 1990. Soil structural aspects of decomposition of organic matter by micro-organisms. Biogeochemistry 11:213–233.

Waldron, L.J., S. Dakessian, and J.A. Nemson. 1983. Shear resistance enhancement of 1.22-meter diameter soil cross-sections by pine and alfalfa roots. Soil Sci. Soc. Am. J. 47:9–14.

Warkentin, B.R. 2008. Soil structure: A history from tilth to habitat. Adv. Agron. 97:239–272.

Yoder, R. 1936. A direct method of aggregate analysis of soils and a study of the physical nature of erosion losses. J. Am. Soc. Agron. 28:337–435.

3

Mechanics of Unsaturated Soils for Agricultural Applications

Rainer Horn
Christian-Albrechts-Universität zu Kiel

Stephan Peth
Christian-Albrechts-Universität zu Kiel

3.1 Introduction

Soils undergo intensive changes in their physical, chemical, and biological properties during natural soil development and as a result of anthropogenic processes such as plowing, sealing, erosion by wind and water, amelioration, excavation, and reclamation of devastated land. In agriculture, soil compaction as well as soil erosion by wind and water are classified as the most harmful processes. Irrespective of land-use systems, soil deformation as the sum of soil compaction and shear processes leads to numerous environmental changes affecting soil quality for crop production, soil biodiversity, filtering and buffering functions of soils, soil-water household components, trace gas emissions, soil erosion and nutrient export, and related off-site effects (Figure 3.1).

In forestry, normal plant and soil management, tree harvesting, and clear-cutting also affect site-specific properties, including organic-matter loss and groundwater pollution and gas emissions, which have the potential to cause global changes. Furthermore, soil amelioration especially by deep tillage prior to replanting often causes irreversible changes in properties and functions. These interrelationships have been described by Jayawardane and Stewart (1994), Soane and van Ouwerkerk (1994), Horn (2000), and Horn et al. (2006) as well as by Birkas et al. (2004). Oldeman (1992) showed that about 33 million ha of arable land are already completely devastated by soil compaction in Europe alone while the total area of degraded land worldwide exceeds about 2 billion ha. Physical (soil erosion and deformation) and chemical processes are responsible for about 1.6 and 0.4 billion ha of degraded soils, respectively. Worldwide population growth will reduce the average area per person for food and fiber production from 0.27 ha today to <0.14 ha within 40 years, and even if the advances in technological developments continue to grow, more concern has to be made in order to prepare enough food for the population worldwide. Consequently, a more detailed analysis of soil and site properties is needed to manage soils in accordance with their potential properties. Numerous attempts to predict the strength of the soils were made; however, all are based on static loading experiments, which assume that equilibrium in settlement for stepwise increasing loads is achieved. This approach neglects the dynamic nature of soil loading processes during field traffic where loads are applied mostly short term and where multiple loading events may cumulatively add to the total compression. Thus, the increase in irreversible deformation of
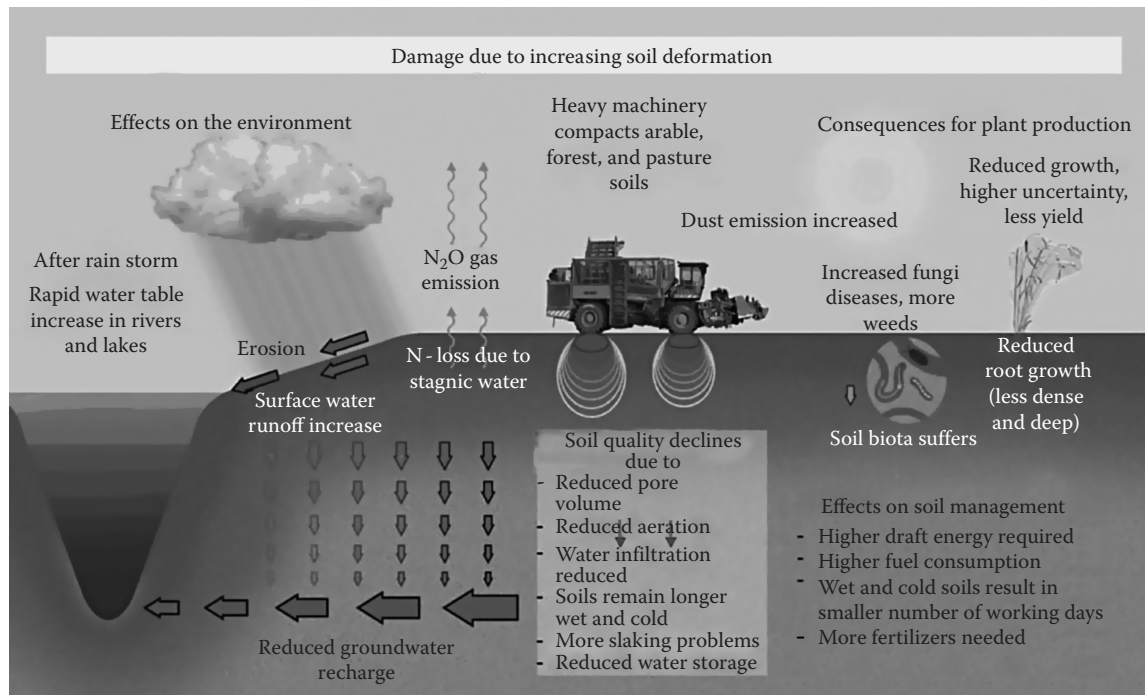
FIGURE 3.1 Summary of the effects of stress application on soil properties and functions. (Translated and slightly modified from van der Ploeg, R.R., W. Ehlers, and R. Horn. 2006. Schwerlast auf dem Acker. Spektrum der Wissenschaft August, 80–88. ISSN: 0170-2971.)

arable soil by agricultural machinery can not only be related to the increasing mass of machines (defined as static approach) but also enhanced by the increase of wheeling frequency as a dynamic component. The relevance of wheeling frequency can be estimated based on the calculations of Olfe (1995). Considering an average-sized wheat-production farm, the number of load repetitions in a time period of 5 years may add up to 50 events for 85% of the field and up to 100 events for permanent wheeling tracks.

An additional threat concerning soil degradation is caused by tillage erosion as the actual soil loss can exceed by far 20 Mg ha^{-1} and per tillage operation. This effect is the more pronounced, the drier the soil, the weaker the aggregates, the higher the dynamic energy input by agricultural machinery, and the greater the field size (Karlen, 2004; Reicosky et al., 2005; da Silva et al., 2006; Van Oost et al., 2006). The kinetic energy applied during seedbed preparation under dry conditions alone already results in an increased mass transport by wind. Additionally, the unproductive water loss due to evaporation from an increased accessible pore and particle surface and the enhanced organic-matter decomposition due to tillage both contribute to global change problems (Lal, 2005). Soil creep, mudflow after rainstorms, and intensified surface-water runoff are due to reduced/prevented vertical water infiltration in soils generating pronounced lateral fluxes of water, solids, and nutrients. The loss of shear strength, defined by the angle of internal friction and cohesion, of near-saturated soils subject to buoyancy forces that cause such lateral fluxes produces tremendous damage to the landscape and human beings. Related economic impacts are estimated to exceed billions of euros worldwide (van den Akker et al., 1998). In addition

to productivity losses of soils in farming regions, the loss in biodiversity and the effects on global changes by modifying physicochemical processes result in the emission of greenhouse gases and the mobilization of heavy metals in soils due to redox reactions, which in turn cause groundwater pollution threats.

If, on the other hand, the tilled soil dries out, transport by wind may occur, resulting in severe reductions in potential site productivity. Furthermore, the preparation of a seedbed leads to abrupt changes in the transport of gas, water, ions, and heat between the tilled and deeper soil layers (Boone and Veen, 1994; Ball et al., 2000; Lipiec and Hatano, 2003). This is especially true concerning preferential flow through structured soils (Hendrickx and Flury, 2001). When it comes to the quantification of soil deformation with respect to the induced changes of physical functions, again the dynamic nature of soil loading processes during field traffic must be considered, where loads are applied mostly short term and where multiple loading events may cumulatively add to the total compression, although the statically determined precompression stresses are not exceeded and should result in no further compression.

Such anthropogenic changes make clear that the discussion of soil process dynamics within the soil profile is most relevant. Some of the interactions between soil structure, water status, and aeration of structured soils in relation to root growth and compressibility of arable land have been described by Emerson et al. (1978). In current discussions on sustainability, soil compaction is repeatedly mentioned as one of the main threats to agriculture, which should be avoided (Soane and van Ouwerkerk, 1994; Pagliai and Jones, 2002; Horn et al., 2006; Toth et al., 2008).

The aim of this chapter is to analyze the relationships between stress, strain, and strength as well as the consequences of soil deformation for physical and physicochemical properties. It also includes a description of well-established and new methods for measuring mechanical properties of soils and introduces possibilities for modeling soil strength and stress–strain relationships from the micro- to the macroscale.

3.2 Stress–Strain Relationships in Soils

3.2.1 Stress Theory

3.2.1.1 Definitions

Before discussing the methods of field and laboratory stress measurement as well as factors influencing compaction, one needs to differentiate between several terms used to define compressive properties. These definitions have been taken from Fredlund and Rahardjo (1993), Hartge and Horn (1999), Parry (2004), and McCarthy (2007).

3.2.1.2 Mechanical Stresses

Stress is defined as force per area within a solid body. Stress can be induced by external or internal forces, which lead, if the body is nonrigid, to a change in the body's volume and/or shape expressed as deformations or strains. The mechanical behavior of a soil can therefore be characterized by its stress–strain relationships.

Strength is typically referred to as the maximum amount of stress a solid material can withstand before it fails; thus, exceeding soil strength results in soil failure or yield. Strength depends on internal parameters such as particle-size distribution, type of clay minerals, nature and amount of adsorbed cations, content and type of organic matter, aggregation induced by swelling and shrinkage, stabilization by roots and humic substances, bulk density, pore-size distribution and pore continuity of the bulk soil and single aggregates, water content, and/or matric potential (Horn, 1981).

3.2.1.2.1 Stress State

The number of stress state variables required to define the stress state depends primarily upon the number of phases involved. The effective stress σ' can be defined as a stress variable for saturated conditions (Terzaghi, 1936) and is the difference between the total stress (σ) and the neutral stress (u_w), which is equal to the pore water pressure:

$$\sigma' = \sigma - u_w, \tag{3.1}$$

where
σ' is transmitted by the solid phase
u_w is transmitted by the liquid phase, respectively

In unsaturated soil, stresses are transmitted by the solid, liquid, and gaseous phases. Thus, Equation 3.1 becomes (Bishop, 1959)

$$\sigma' = (\sigma - u_a) + \chi(u_a - u_w), \tag{3.2}$$

where
u_a and u_w are pore air and water pressures, respectively
χ is a factor that is often assumed to be directly related to the degree of saturation (S_r)

At saturation ($u_w = 0$ kPa), $\chi = 1$, while at $u_w = -10^6$ kPa, $\chi = 0$.

For sandy, less compressible, and nonaggregated soils, χ can be estimated as

$$\chi = 0.22 + 0.78 S_r. \tag{3.3}$$

For silty and clayey soils, the values of the parameters in Equations 3.1 and 3.2 depend on soil aggregation, pore arrangement and strength, and hydraulic properties. Thus, the material function of the components in structured soils is only valid as long as the internal soil strength is not exceeded by the externally applied stresses. It changes if, for example, aggregates are destroyed during soil deformation and the structural properties reduce to those that depend merely on texture.

The extent of soil deformation can be described by stress–strain relationships and by their relative proportions. In the absence of gravitational and other applied forces, effective stresses in three-phase soil systems can be expressed as a tensor containing three normal stresses and six shearing components; assuming symmetry of the stress tensor, the shear stresses are reduced to three independent components. Therefore, three normal stress (σ_x, σ_y, σ_z) and three shear stress (τ_{xy}, τ_{xz}, τ_{yz}) components must be determined to fully define the stress state at every location in a solid body (Nichols et al., 1987; Horn et al., 1992). For an unsaturated soil, the stress state can be described completely by a symmetric matrix, which can be written as two independent stress tensors (Fredlund and Rahardjo, 1993):

$$\begin{bmatrix} \sigma_x - u_a & \tau_{yx} & \tau_{zx} \\ \tau_{xy} & \sigma_y - u_a & \tau_{zy} \\ \tau_{xz} & \tau_{yz} & \sigma_z - u_a \end{bmatrix} \tag{3.4}$$

and

$$\begin{bmatrix} (u_a - u_w) & 0 & 0 \\ 0 & (u_a - u_w) & 0 \\ 0 & 0 & (u_a - u_w) \end{bmatrix}. \tag{3.5}$$

A graphical representation of the stress state variables and the action of water menisci at intergranular contacts in an unsaturated soil is provided in Figure 3.2.

For symmetric tensors, it is always possible to find a coordinate system in which the shear components become 0 and the tensor becomes diagonal. In this principal axis system and assuming that the pore air pressure u_a = atmospheric pressure, the stress tensor in Equation 3.4 simplifies to

$$\begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \tag{3.6}$$
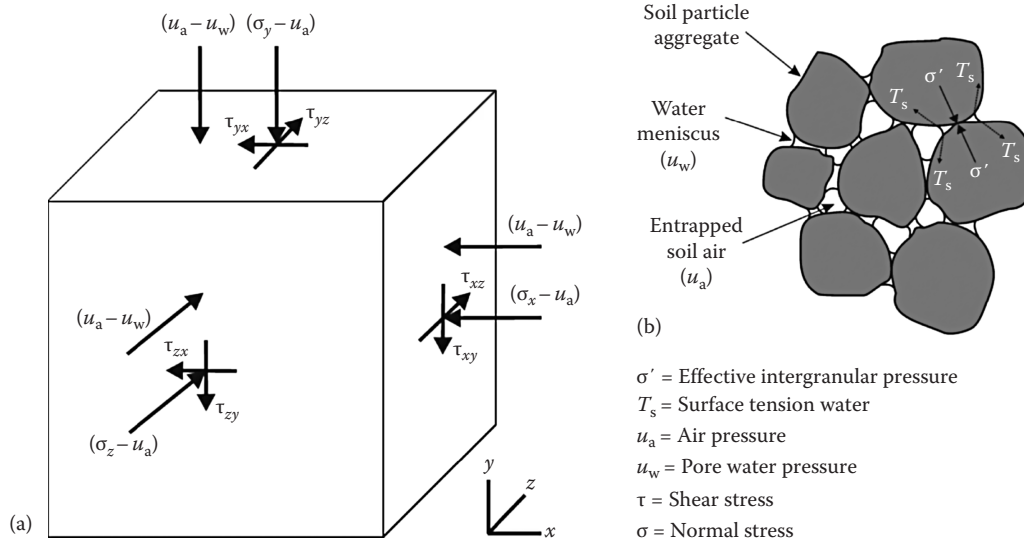
FIGURE 3.2 (a) Stress state variables for an unsaturated soil. (b) Schematic sketch of water menisci at interparticle/interaggregate contacts illustrating the effect of pore water pressure (u_w) on intergranular tension stresses. (Supplemented after Fredlund, D.G., and H. Rahardjo. 1993. Soil mechanics for unsaturated soils. John Wiley & Sons, New York.)

with principal stresses σ_1, σ_2, and σ_3 acting on the solid. For a single-valued characterization of the stress state, two invariants of the stress matrix are often used—the mean normal stress (MNS) and the octahedral shear stress (OCTSS; Koolen, 1994):

$$MNS = \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3), \qquad (3.7)$$

$$OCTSS = \frac{1}{3}\sqrt{(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2}. \qquad (3.8)$$

3.2.1.2.2 Stress Propagation

Each applied force per contact area is transmitted into the soil in three dimensions and can alter the physical, chemical, and biological properties of the soil (e.g., water infiltration, rootability) if the internal mechanical strength is exceeded. Stress propagation theories are rather old and have been often modified and adapted to in situ situations. The fundamental theory of Boussinesq (1885) is only valid for completely elastic material, while Fröhlich (1934) and Soehne (1958) included elastoplastic properties through the introduction of concentration factor values (υ_k). More comprehensive descriptions of these models are given by Koolen and Kuipers (1983), Bailey et al. (1986), Johnson and Bailey (1990), and Bailey et al. (1992). Horn et al. (1989) introduced precompression stress (P_v)-dependent values for the concentration factor, which are smaller in the recompression stress range, while they increase in the virgin compression stress range. The latter can be explained by the plastic deformation behavior, which causes a deeper stress transmission closer to the perpendicular line.

3.2.1.3 Hydraulic Stresses

The effective stress (Equation 3.1) defines the forces per given area, which can stabilize the soil particles against any kind of soil deformation. The hydraulic stress component (u_w) can either be negative (concave water menisci) or positive (convex water menisci). In case of convex menisci, it can result in weakening of the total soil system, especially when shear forces are applied. In case of pure compression under saturated conditions, however, effective and hence compressive stresses are reduced while the neutral stress (water pressure) bears part or even the entire externally applied load. In most cases, air pressure (u_a) in Equation 3.2 is ignored, assuming that gas pressure in soil during loading is in equilibrium with atmospheric conditions.

If, on the other hand, hydraulic stresses become negative (=suction), the contractive forces even increase the effective stress (Figure 3.2). In addition to capillary forces associated with water, salt effects (i.e., water potential = sum of matric and osmotic potential) and hydrophobic substances can also increase soil strength by altering wetting angle, cohesion, internal friction, as well as elastic and viscous properties of the soil (Barré and Hallett, 2009; Markgraf and Horn, in press).

Hydraulic stresses result in shrinkage in almost all soils. In an initially homogenized state, soils undergo proportional and thereafter residual shrinkage during the first drying phase. Depending on the history of the formerly applied hydraulic and mechanical stresses, the shrinkage curve pattern, however, shifts and it must be now differentiated between structural shrinkage (=structural rigidity), virgin, and residual shrinkage behavior (Groenevelt and Grant, 2001; Braudeau et al., 2004; Peng and Horn, 2005). Structural shrinkage defines the internal soil strength caused by deformation related to capillary forces. Structural shrinkage is more pronounced, the more often and intensive soil dries out. Horn (1994a), for example, described the effect of the drying frequency and intensity on shrinkage behavior and pointed out that soil strength increases the more often and longer contracting water menisci occur while only a few but

very intense drying processes are mostly not capable of rearranging particles to form a rigid structure with smaller entropy (Horn and Dexter, 1989).

The power of capillary forces and their strengthening potential for soil architectures can be demonstrated by a simple computation. Given the mass of the earth (~6×10^{24} kg) and the area of, for example, Germany (~$360,000 \, km^2$) of which 50% is arable soil (=$180,000 \, km^2$) and assuming a thickness of the plowed topsoil of 30 cm with a hypothetical specific surface area of $800 \, m^2 \, g^{-1}$ (for simplicity we assume the soil to consist of smectitic clay and 2%–4% of organic carbon), we would end up with a total surface area of ca. $6.5 \times 10^{19} \, m^2$. If we furthermore presume that water menisci at a potential of pF 4 would be effective over the total surface ($\chi \rightarrow 1$, assuming a very high negative air-entry pressure), then the sum of menisci forces acting within the topmost 30 cm of arable soils in Germany could carry the total mass of the earth.

3.2.1.4 Stress Coupling

Based on the findings by Baumgartl (2003), who described the similarity of mechanical stress–strain curves and shrinkage curves resulting from hydraulic stresses (Figure 3.3), Peng and Horn (2008) proved that the link between mechanical and hydraulic "pre"stresses results in nearly identical changes in shrinkage curves. Additionally, Baumgartl (2003) explained the differences between mechanically induced collapse and matric potential-dependent shrinkage patterns by the dimensionless χ factor, which, however, is difficult to determine experimentally. Accounting for the interactions between mechanical and hydraulic processes is very important to understand the deformation behavior of unsaturated soils. Nevertheless, their description is inherently complex involving effective stress theory, loading rate dependency, and nonlinear hysteretic coupled transport and stress–strain functions.
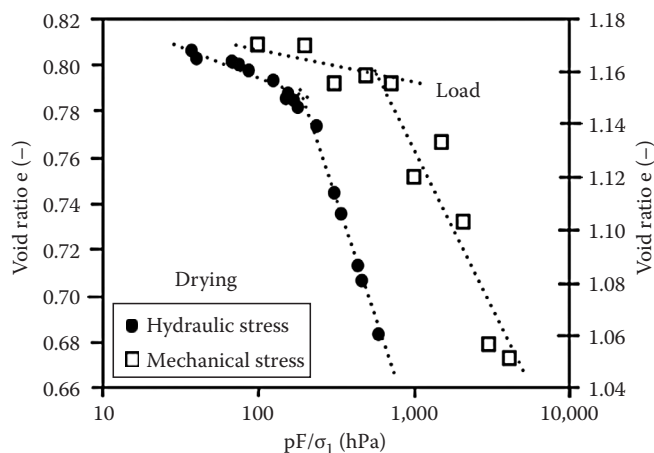
FIGURE 3.3 Stress–strain relationships showing the similarity between compaction (right ordinate) and shrinkage curves (left ordinate). (Modified from Baumgartl, T. 2003. Koppelung von mechanischen und hydraulischen Bodenzustandsfunktionen zur Bestimmung und Modellierung von Zugspannungen und Volumenänderungen in porösen Medien. Habilitation Thesis. Schriftenreihe Inst. Pflanzenernaehrung und Bodenkunde, Germany, No. 62, 133 pp. ISSN: 0933-680X.)
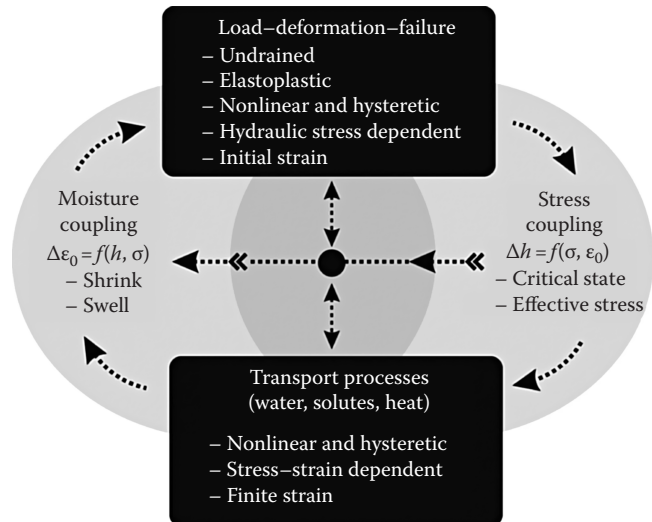
FIGURE 3.4 Conceptual framework for modeling coupled soil mechanical and hydraulic processes. The *stress-coupling* routine models pore water pressure changes (Δh) as a function of stress (σ) or strain (ε_0). The *moisture-coupling* routine models strain changes ($\Delta \varepsilon_0$) as a function of pore water pressure (h) and effective stress (σ'). (From Richards, B.G., and S. Peth. 2009. Modelling soil physical behaviour with particular reference to soil science. Soil Till. Res. 102:216–224.)

According to Richards and Peth (2009), mechanical and hydraulic processes can be coupled by treating the initially independent load-deformation and transport processes by an incremental approach as shown in Figure 3.4. This coupling is most important in modeling the interaction between water flow and deformation/failure as, for example, encountered in swelling soils, landslides on slopes during rainstorms, and consolidation of high water content soils. In a stress-coupling routine, changes in pore water pressure (h) are calculated as a function of changes in stress (σ) or strain (ε), and in a moisture-coupling routine, changes in stress or strain are calculated as a result of pore water pressure changes related to both water transport and load deformation. A detailed mathematical description of the finite element model (FEM) used for mechanical and hydraulic stress coupling and the underlying constitutive relations and material parameters is given in Richards and Peth (2009).

3.2.1.5 Soil Strength Changes due to Stress-Dependent Effects on Hydraulic Properties

Total stress application affects interparticle bonding as well as pore-size distribution, pore geometry, and the degree of saturation, which either weakens the soil during compression or makes it even stronger. While the former is caused by soil settlement, reducing pore volume and hence increasing the degree of saturation, the latter is related to the stress-dependent water redistribution in newly formed finer pores at the expense of bigger and already air-filled ones causing more negative matric potential values, thus increasing effective stresses (Equation 3.2). Fazekas and Horn (2005) describe the effect of static loading on the transition between stress-induced soil strengthening and weakening and found that this change occurs close to the precompression

stress value. The changes resulted initially in a strength increase and the pore water pressure became more negative, while at higher stresses the pore water pressure even reached positive values starting from −30 kPa initial matric potential. If external forces are applied as repeated short-time loading (e.g., due to wheeling or trampling) or if shear deformation-induced rearrangement of particles occurs, we also have to consider the "water menisci pumping effect due to short-term suction changes" because it enhances the weakening process due to a more pronounced and more complete swelling, finally resulting in a complete soil homogenization (Horn, 1976, 2003; Pietola et al., 2005; Krümmelbein et al., 2008). Janssen (2008) measured stress-induced changes in pore water pressure in paddy soil samples. Increasing frequency of loading and unloading changed water menisci forces from negative to positive pore water pressure values and caused complete soil softening. Similar effects were also reported for short-term cyclic loading of homogenized soil, where even when the soil was unsaturated at an initial matric potential of −6 kPa, positive pore water pressure was measured during the short-term loading phases (Peth and Horn, 2006b). Water-induced soil softening is also evident when shear forces are applied to the soil surface, for example, by wheeling due to slip effects (Horn, 2000; Horn et al., 2006). Horn et al. (1995) proved the effect of shear-induced changes in pore water pressure in comparison with the previously applied static stress (Figure 3.5). Even if the consolidated stress–strain state had been

reached due to a given static loading, the following so-called consolidated shear process again caused an additional compaction and an extra change in the pore water pressure values. The more rapid the shear process occurs and the smaller the hydraulic conductivity even at increased hydraulic gradients, the more positive become the pore water pressure values and the weaker the total soil (Hartge and Horn, 1999).

3.2.2 Strain Theory

Mechanical processes in soil are generally stress dependent. All stresses exceeding the internal soil strength result in failure, which can be defined either by an irreversible plastic volume decrease (=soil compaction) or by a rearrangement of soil particles at a constant bulk soil volume (=shearing). Analogous to stress, strain can be described by normal (ε_x, ε_y, ε_z) and shear strain (ε_{xy}, ε_{xz}, ε_{yz}) components, which are also described completely by a symmetric strain tensor:

$$
\begin{bmatrix}
\varepsilon_x & \varepsilon_{xy} & \varepsilon_{xz} \\
\varepsilon_{xy} & \varepsilon_y & \varepsilon_{yz} \\
\varepsilon_{xz} & \varepsilon_{yz} & \varepsilon_z
\end{bmatrix}. \tag{3.9}
$$

In the corresponding principal axis system, the strain tensor reduces to

$$
\begin{bmatrix}
\varepsilon_1 & 0 & 0 \\
0 & \varepsilon_2 & 0 \\
0 & 0 & \varepsilon_3
\end{bmatrix}. \tag{3.10}
$$

The volumetric strain (ε_{vol}) can be calculated as

$$
\varepsilon_{vol} = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 = \varepsilon_x + \varepsilon_y + \varepsilon_z. \tag{3.11}
$$

Note that the tensor trace is invariant under coordinate transformations. Furthermore, the strain components and their proportions depend on internal and external parameters and require the determination of all components in a 3D volume.

3.2.2.1 Volumetric Deformation

Compression refers to a process that describes the increase in soil mass per unit volume (increase in bulk density) under an externally applied load or under changes in internal pore water pressure. Considering the soil as a continuum, soil movement can be described as a translation field (*d*) whose local properties are usually characterized by three components of its spatial derivative:

FIGURE 3.5 Changes in shear stress, sample height, and pore water pressure during shearing. The samples were predried to −6 kPa and preconsolidated prior shearing. Upon shearing at the same applied stress, the sample further consolidated (height change) and pore water pressure increased reaching even positive values during shearing. After normal and shear stress were removed, pore water pressure decreased to a final value more negative than the initial pore water pressure. (From Horn, R., T. Baumgartl, R. Kayser, and S. Baasch. 1995. Effect of aggregate strength on changes in strength and stress distribution in structured bulk soils, p. 31–52. *In* K.H. Hartge and R. Stewart (eds.) Soil structure: Its development and function. CRC Press, Boca Raton, FL.)

$$
\nabla \cdot \vec{d} =
\begin{bmatrix}
\dfrac{\partial}{\partial x} d_x & \dfrac{\partial}{\partial x} d_y & \dfrac{\partial}{\partial x} d_z \\[2ex]
\dfrac{\partial}{\partial y} d_x & \dfrac{\partial}{\partial y} d_y & \dfrac{\partial}{\partial y} d_z \\[2ex]
\dfrac{\partial}{\partial z} d_x & \dfrac{\partial}{\partial z} d_y & \dfrac{\partial}{\partial z} d_z
\end{bmatrix}. \tag{3.12}
$$

Every change in stress state therefore results in soil deformation. This divergence (volumetric strain) can be expressed as the divergence of the deformation field:

$$\varepsilon_{vol} = \mathrm{div}(\vec{d}) = \nabla \cdot \vec{d} = \mathrm{tr}(\nabla \cdot \vec{d}) = \frac{\partial}{\partial x}d_x + \frac{\partial}{\partial y}d_y + \frac{\partial}{\partial z}d_z, \quad (3.13)$$

and results in volumetric compression or extension. The terms of the right-hand side of Equation 3.13 show the corresponding strains in *x*, *y*, and *z* directions and how they are related to respective deformation components.

3.2.2.2 Shear Deformation

Shearing a soil volume always leads to a pronounced volume constant rearrangement of particles often resulting in significant changes in pore functions by decreasing pore connectivity or continuity and changes in interparticle strength due to a loss of cohesion (Figure 3.6). The stress/strain at which the interparticle strength is lost and the soil begins to fail is referred to as yield stress/strain. The resistance of a soil toward shear deformation is described by shear moduli (*G*), which will be explained in more detail in Section 3.3 (*rheometrical parameters*).

3.2.3 Stress–Strain Processes

Generally, mechanical processes in soils are described by the stress/strain equation:

$$\begin{bmatrix} \sigma_x(\vec{x},t) & \tau_{xy}(\vec{x},t) & \tau_{xz}(\vec{x},t) \\ \tau_{xy}(\vec{x},t) & \sigma_y(\vec{x},t) & \tau_{yz}(\vec{x},t) \\ \tau_{xz}(\vec{x},t) & \tau_{yz}(\vec{x},t) & \sigma_z(\vec{x},t) \end{bmatrix}$$
$$= f \begin{bmatrix} \varepsilon_x(\vec{x},t) & \varepsilon_{xy}(\vec{x},t) & \varepsilon_{xz}(\vec{x},t) \\ \varepsilon_{xy}(\vec{x},t) & \varepsilon_y(\vec{x},t) & \varepsilon_{yz}(\vec{x},t) \\ \varepsilon_{xz}(\vec{x},t) & \varepsilon_{yz}(\vec{x},t) & \varepsilon_z(\vec{x},t) \end{bmatrix}, \quad (3.14)$$

(a)　　(b)
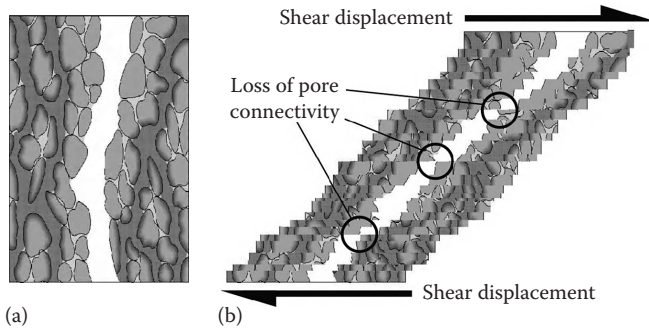
FIGURE 3.6 Schematic sketch of the effect of shear displacement on pore continuity (b) in comparison with an "unstressed" pore system (a).

where *f* defines the soil-specific material properties. In the case of a linear elastic isotropic material, the tensor notation reads as follows:

$$\begin{bmatrix} \sigma_x & \tau_{xy} & \tau_{xz} \\ \tau_{xy} & \sigma_y & \tau_{yz} \\ \tau_{xz} & \tau_{yz} & \sigma_z \end{bmatrix} = \frac{E}{1+v}\begin{bmatrix} \varepsilon_x - \varepsilon_m & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{xy} & \varepsilon_y - \varepsilon_m & \varepsilon_{yz} \\ \varepsilon_{xz} & \varepsilon_{yz} & \varepsilon_z - \varepsilon_m \end{bmatrix}$$
$$+ \frac{E}{3(1-2v)}\begin{bmatrix} \varepsilon_m & 0 & 0 \\ 0 & \varepsilon_m & 0 \\ 0 & 0 & \varepsilon_m \end{bmatrix} \quad (3.15)$$

where

　$E = \Delta\sigma/\Delta\varepsilon$ (Young's modulus or modulus of elasticity)
　v is the Poisson's ratio

The first tensor on the right-hand side of Equation 3.15 describes the volume constant shear deformation and the second tensor the volumetric strain with the mean volume strain ε_m defined as

$$\varepsilon_m = \frac{1}{3}(\varepsilon_x + \varepsilon_y + \varepsilon_z) = \frac{1}{3}\varepsilon_{vol}. \quad (3.16)$$

Further details on the notions of stress–strain relationships can be found in Fredlund and Rahardjo (1993) and Warrick (2002).

To describe nonlinear elastic–plastic soil behavior as a function of stress, soil–water suction, and stress history, Richards (1978) introduced, based on the analysis of experimental data, a constitutive model, where the mechanical material properties are defined as hyperbolic functions allowing to treat soil as nonlinear and irreversible hysteretic material. The stress-dependent mechanical material properties are defined based on standard soil mechanical material parameters (bulk modulus, shear modulus, and Mohr–Coulomb failure criteria), which can be readily derived from soil tests (e.g., consolidation test, shear test, triaxial test; Section 3.3.4). Bulk modulus *K* and shear modulus *G* characterize the stiffness and shear strength of the soil, respectively. The basic relationships for the two simple stress-dependent moduli are (Richards and Peth, 2009)

$$K = k_1\sigma_m^n + k_2h^m + k_0 \quad (3.17)$$

and

$$G = (g_1\sigma_m^p + g_2h^r)\left(1 - \left(\frac{\tau}{\tau_f}\right)^q\right) + g_0, \quad (3.18)$$

where

　σ_m is the previous maximum of the mean normal stress $((\sigma_1 + \sigma_2 + \sigma_3)/3)$
　h is the soil water suction
　τ is the shear stress
　τ_f is the yield stress according to the employed yield criteria

The shape of the material functions is further defined by empirical material constants k_{0-2}, g_{0-2}, n, m, p, r, and q, which are

derived directly from fitting soil test data or by back-analyzing soil test results with an FEM approach (for further details see Section 3.7.3 and Richards and Peth (2009)).

3.3 Soil Strength–Stability Indicators and Their Measurement

Soil strength and elasticity are mechanical key determinants for maintaining efficient functionality and ultimately quality of soils. Indicators describing and quantifying the mechanical resistance, deformation, and resilience of agricultural soils are important for several reasons. One is to assess if external mechanical loads applied to soils during field operations modify solid–void architectures and volume relationships through particle relocations and how far this may detrimentally affect pore functions. Another is to evaluate the penetrability of soil horizons and aggregates by roots and soil organisms while prospecting soils for water, nutrients, and food. A third reason is to estimate the risk for soil mass movements and soil erosion processes along slopes.

Soil scientists and civil engineers are likewise interested in measuring soil mechanical stability parameters, although with sometimes contrasting goals. In fact, many of the mechanical parameters that are commonly used in agricultural soil science originated in the field of civil engineering. On the other hand, civil engineers have adapted methods from agricultural soil science (e.g., Atterberg's limits) and nowadays realize the importance of understanding soils as bioactive media where interacting physical and biological processes in turn have a strong influence on soil strength. However, recognizing soils as deformable (elastoplastic) multiphase systems with an internal structure, we need to define boundary-dependent strength/stability limits in order to estimate and ultimately predict deformation-induced changes in physical, chemical, and biological soil functions.

Most commonly used and well-established parameters characterizing the mechanical strength/stability of soils on multiple scales ranging from interparticle to pedon levels will be described in the following, including methods for their determination. Where appropriate, we supplemented this section with new techniques and methodologies in the field of soil mechanical research.

3.3.1 Soil Consistency

The consistency of soils is a result of cohesive and adhesive forces acting between particles and between fluid-phase molecules and the solid surface, respectively. Together, they determine the interparticulate resistance to deformation, rupture, and flow. Soil wetness obviously has a strong influence on the kind of reaction against external forces (e.g., hard, friable, soft, plastic, sticky, and viscous). Atterberg (1911) has defined different classes of soil behavior based on soil wetness, which is a simple yet useful method to judge soil consistency and hence workability. The so-called *Atterberg's limits* quantify the soil's resistance to mechanical energy as a function of gravimetric water content (Θ_g):

3.3.1.1 Liquid Limit

The liquid limit is determined with a CASAGRANDE apparatus (Casagrande, 1932) where a homogenized moist sample is placed in a special bowl and a V-shaped trench is cut from the top to the bottom (Kézdi, 1974). The bowl with the sample is bounced up and down by a special crank onto a hard rubber block at a frequency of two strokes per second until the trench is closed at a length of <1 cm. The number of strokes when this condition is met is counted and the gravimetric water content determined on a subsample at the end of the test. The liquid limit is defined as the water content (Θ_g) where the test criteria are fulfilled at 25 strokes corresponding to a certain "standardized" amount of energy applied to the soil. The test is repeated for at least four water content values and the number of strokes plotted versus the water content on a semilogarithmic paper, finally producing a linear line. The liquid limit increases with clay and organic-matter contents, ionic strength, cation valency, and proportion of 2:1 type clay minerals in the soil.

3.3.1.2 Plastic Limit

The plastic limit is defined as the water content (Θ_g) when homogenized soil samples begin to crack and crumble when rolled to a diameter <3 mm. It has practical implication in civil engineering and agriculture since it determines the lower critical water content at which soils become plastic and hence difficult to till or excavate.

3.3.1.3 Plasticity Index

The difference between the water contents (Θ_g) at the liquid limit (LL) and plastic limit (PL) is the so-called plasticity index (PI = LL – PL), which is often used as an indicator for soil workability. With an increasing plasticity index, soils become more sensitive to plastic deformation. The higher the plasticity index the smaller is the water-content range at which soils can be tilled efficiently and trafficked without considerable soil deformation. The smaller is also the angle of internal friction over a wider range of water contents. Many attempts have been made to correlate PI to soil strength (Kézdi, 1974; Hartge and Horn, 1992; Kretschmer, 1997). In principle, this test only gives information on minimum strength values correlated with soil water content. It is furthermore limited to silty, loamy, and clayey soil samples while such correlation fails and is too insensitive for sandy material. Thus, especially coarse-textured soils cannot be classified by these data.

3.3.1.4 Rheometrical Parameters

Advances in rheological research as an independent science have brought sophisticated techniques for quantifying flow behavior and deformation properties of viscoelastic substances (mainly fluids)—the most important parameters are shear modulus (G), plastic viscosity (η_p), and yield stress (τ_y).

Although first rotational viscometers (1888) and later rheometers (~1951) have been developed, early modern microprocessor-based systems, which are available since the 1980s, added more

options for rheological testing including creep, relaxation, and oscillation tests. The methods are today well established and widely used in polymer, chemical, and material sciences as well as in the paper and food industries.

Despite its potential applicability, however, they are still not commonly employed in soil science research. One of the earliest contributions has been made though by Yasutomi and Sudo (1967) who used a low-frequency forced oscillation viscometer to study viscoelastic properties of soil. Later, Keren (1988) and Hesterberg and Page (1993) used a viscometer to investigate the influence of electrolyte composition and pH on rheological characteristics of clay mineral suspensions. Studies based on mineral suspensions certainly aid the understanding of the principle influence of surface chemistry on mechanical solid–interface interactions but have the drawback of ignoring the pertinent action of water as a lubricant during deformation processes. More recently, Ghezzehei and Or (2001) have used a rotational rheometer with a parallel-plate sensor system to determine these *rheometrical parameters* on natural soils and clay samples under more realistic field-moisture conditions. Higher water contents and high stress amplitudes resulted in lower viscosity of all investigated soils. Although this is not surprising, the authors could demonstrate that rheometry is a suitable and sensitive technique that enables us to quantify rheological properties of soils based on physically well-defined boundary conditions. Also Markgraf et al. (2006) investigated rheological characteristics of natural soil samples (Avdat loess, smectitic vertisol, and kaolinitic oxisol) using a parallel-plate rheometer. By conducting so-called amplitude sweep tests (oscillatory test with continuously increasing deformation), they were able to reveal differences in structural stabilities between the smectitic vertisol and the kaolinitic oxisol and the influence of electrolyte concentrations on rheological parameters of Avdat loess samples.

It would go beyond the scope of this chapter to describe the principles of rheometry in great detail. We therefore will only outline the basic test and analysis procedures in the following. Readers that are more interested in this technique will find useful staring points in Ghezzehei and Or (2001), Markgraf et al. (2006), and Mezger (2006).

The sample (usually a remolded paste at defined water content) is placed between two parallel plates (diameter 2.5 cm) with an adjustable gap size commonly between 2 and 4 mm (Figure 3.7). To set up the test, the upper plate is slightly pressed against the sample to ensure contact and time is given for stress dissipation. To avoid slip at the interface between the sample and the plates, both the upper and lower plates are furrowed. In an amplitude sweep test, the upper plate rotates (oscillatory) at a predefined frequency (e.g., 0.5 Hz) and variable amplitude beginning with very small shear deformations (corresponding to angular deflection) up to the maximum shear deformation where the angular deflection is equal to the gap spacing between the plates (e.g., 4 mm deflection at a gap spacing of 4 mm). Shear strain γ is expressed as percentage ranging from 0% (no deflection) to 100% (maximum deflection).

(a)

(b)
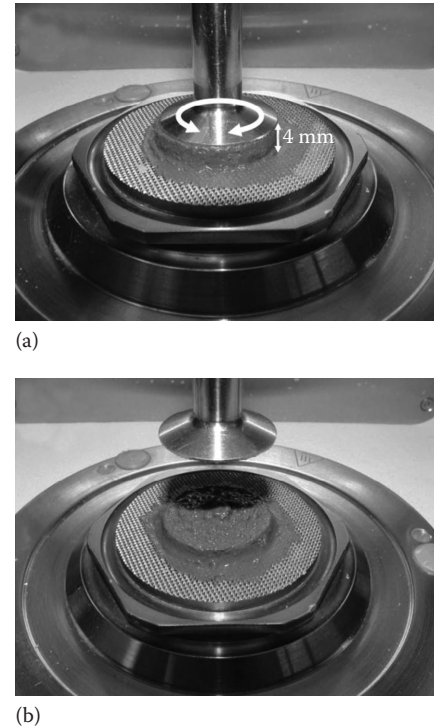
FIGURE 3.7 Rheometer with a parallel-plate measurement system. A remolded soil paste is placed between two furrowed plates by moving down the upper rotational shaft. (a) The upper plate rotates (arrows indicate the rotation in an oscillatory test) while the lower plate remains static. During oscillation, shear force (torque) and angular velocity are measured by sensors placed in the shaft of the upper plate. (b) Shows the sample after the test. A typical result of a rheometrical test is shown in Figure 3.8.

A typical rheological parameter calculated from an amplitude sweep test is the shear modulus $G = \tau/\gamma$, where τ is the shear stress and γ the shear strain. The shear modulus is distinguished into a storage modulus $G' = \tau_A/\gamma_A \cos \delta$ and a loss modulus $G'' = \tau_A/\gamma_A \sin \delta$ representing the elastic and viscous part of deformation, respectively (Figure 3.8). The subscript A denotes the amplitude of the deflection that corresponds to the shear stress and shear deformation pair. The phase shift angle δ describes the viscous reaction of the sample to the prescribed shear deformation or shear stress. At $\delta = 0°$ (tan $\delta = 0$), the sample deforms ideal elastic (system response is instantaneous, that is, stress is proportional to strain); at $\delta = 90°$ (tan $\delta = \infty$), the sample shows ideal viscous behavior (system response is shifted by 1/4th of a full oscillation). In other words, the higher the phase shift angle δ the higher is the viscous and the lower is the elastic part of deformation, respectively. Hence, at small tan δ values, the storage modulus G' exceeds the loss modulus G'' and vice versa for tan δ values >1. The magnitude of G' or G'' in turn represents the resistance of the sample against deformation irrespective of whether it is elastic or viscous. However, at tan $\delta = 1$, the yield point is reached indicating that the sample reacts predominantly viscous. For shear deformation exceeding the yield point, the sample becomes quickly more viscous (Figure 3.8). Tan δ is also referred to as loss factor since it indicates the fraction of mechanical energy introduced into
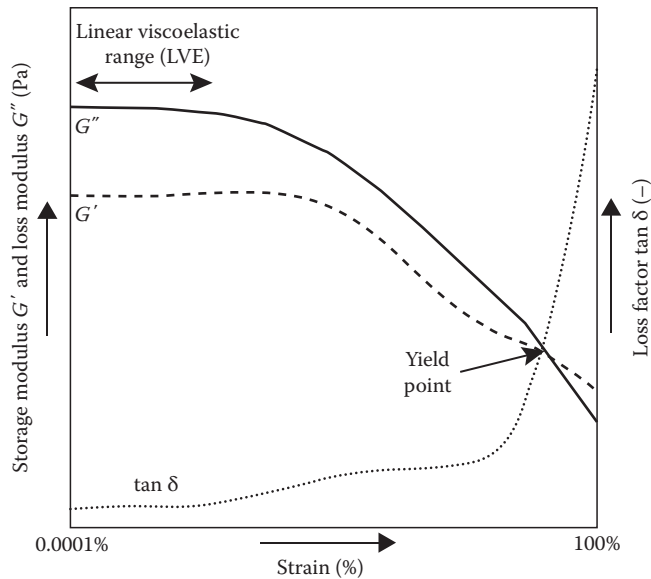
FIGURE 3.8 Schematic diagram showing a typical result of a rheo-metrical test. G' = storage modulus (Pa) and G'' = loss modulus (Pa) correspond to the elastic and viscous deformation, respectively. Tan δ indicates the viscoelasticity of the sample: for tan δ = 0, deformation is ideal elastic; for tan δ = ∞, deformation is ideal viscous; for tan δ = 1 (yield point), deformation is equally viscous and elastic. Note that both moduli (G' and G'') and strain are plotted on log scale.

the system that is lost in deformation. Note that possibilities for variations in boundary conditions during rheometer tests (stress or strain controlled, different stress–strain paths, stress relaxation, oscillation frequency) are numerous and choice of either methodology depends on the material under investigation. For soils, however, amplitude sweep tests have produced interpretable results and seem to be sensitive for a wide range of soil conditions (salt effects, water effects, and organic-matter effects) and allow expressed as yield stress a quantification of micromechanical soil strength (Markgraf et al., 2006; Markgraf and Horn, 2007).

3.3.2 Bulk and Proctor Density

Bulk density is defined as mass per unit volume ignoring the internal arrangement of solids and voids, which limits its use to interpret strength properties and functions of structured soils. For example, two soils at the same bulk density, one aggregated and the other not, can have significantly different strengths. Employing bulk density for deriving site-specific soil properties should therefore in general be done with care (Horn and Kutilek, 2009).

Nevertheless, as far as remolded or freshly deposited soil material is concerned, bulk density may be a suitable indicator for the degree of compaction of such soils. Especially, in civil engineering, often the state of maximum compaction of a given material is of interest and the water content where compactibility of a soil material is at optimum, that is, where least energy

input results in maximum effect (=maximum density). To evaluate the soil moisture condition where compaction of homogenized soil samples is most efficient, *Proctor tests* are commonly employed (Proctor, 1933). Here, a remolded soil at various water contents (Θ_g) is filled into a container and compacted by dropping a weight (~2.5 kg) on the soil from a height of ~30 cm and for a defined number of strokes (usually 25). A series of tests are performed at different water contents and the resulting bulk density measured after each test. Bulk densities and corresponding water contents (Θ_g) are plotted on a graph, which represents the so-called *Proctor curve*. The water content where the compaction reaches a maximum is called the optimum water content (Θ_{opt}) and the maximum density is referred to as *Proctor density* (ρ_p). The coarser the soil sample, the higher ρ_p and the smaller Θ_{opt}. For example, sandy soils have higher ρ_p values than silty or clayey soils, while the latter require a higher water content (Θ_{opt}) to reach the optimum bulk density. Further, the more heterogeneous the grain-size distribution, the higher are ρ_p and Θ_{opt}. Strongly aggregated soils behave like coarser soil materials, that is, also here the proctor density is higher and the corresponding optimal water content lower than in less-aggregated soils. If aggregates themselves are destroyed during the test, ρ_p gets smaller while Θ_{opt} increases compared with completely homogenized material.

3.3.3 Penetration Resistance

Soil resistance to any kind of deformation is determined by various types of penetrometers. The most simple is a thin metal rod (<0.5 cm diameter) with a defined tip shape. Frequently, the tip angle is 30° to simulate root properties or earthworm shapes. The penetrometer can be either pushed into the ground by the constant weight of a falling hammer or it can be driven by a motor at constant speed. The output readings can be either penetration depth per hammer stroke or depth stress depletions, which have to be overcome by the penetrating body in more sophisticated models. Penetration resistance is correlated with root growth, earthworm activity, and tillage effects. When penetration resistance exceeds 2 MPa, root growth is often reduced by half, while values >3 MPa often prevent root growth. Tillage may increase the critical stress value of a hardpan to >3.5 MPa depending on the nature of the pore system and the type of soil structure. Because the penetrometer needle is not as flexible as a root, which can choose planes of weakness for growth, penetrometer readings quantify resistances mainly in the vertical direction. In addition, the penetrometer readings only integrate impeding effects and cannot identify its causes. Despite a voluminous literature on the effect of increasing bulk density and water content on penetration resistance, extrapolation to land management situations is limited. Penetrometer readings can be used to create maps of derived properties (e.g., definition of sites with a given strength irrespective of its origin). Such data can be interpreted using statistical variograms, fractal analyses, or simply by stating that values are spatially different.

3.3.4 Precompression Stress and Compressibility Indices

The compression of a soil means a change not only in total pore volume but is generally accompanied by a change in pore-size distribution including related soil functions (e.g., permeability, storage of water, aeration, gas exchange, heat diffusivity). Hence, the determination of the soil's resistance to compaction and the magnitude of volume changes in soils exposed to external mechanical stresses are of paramount importance for preserving soil physical quality. A number of laboratory tests under precisely defined boundary conditions may be used to quantify soil strength and volume change behavior of soils (Table 3.1).

3.3.4.1 Uniaxial Compression Test

Uniaxial compression tests are used to obtain approximate values of shear strength and settlement behavior of fine-grained soils at unconfined conditions. A vertical normal stress (σ_1) is applied to the specimen (cylindrical or cubic sample), while the stresses on the planes mutually perpendicular to the σ_1 direction ($\sigma_2 = \sigma_3$) are zero. The vertical stress is increased until the specimen fails, which determines the strength of the soil at the given water content. The results of the test can also be used to define the tensile strength of single soil aggregates (Section 3.3.6).

3.3.4.2 Confined Compression Test

Soil stress–strain relationships of undisturbed structured and homogenized soils with respect to volume change behavior are quantified in confined compression tests (*oedometer* tests). In contrast to uniaxial compression tests, the stresses σ_2 and σ_3 are undefined (rigid wall of the soil cylinder) and the respective strains are defined and equal to zero. Both time- and load-dependent changes in soil deformation are measured (recorded as settlement). The slopes of the virgin compression and the recompression line in a void ratio *e* versus log σ plot are referred to as the compression index ($C_c = -\delta e/\delta \log \sigma$) and swelling index ($C_s = -\delta e/\delta \log \sigma$), respectively. The transition from the region of overconsolidation (recompression) to normal consolidation (virgin compression) is defined by the precompression stress, which separates the stress range where soil deformation is considered fully elastic (i.e., reversible) from the stress range where soil deformation is elastoplastic (i.e., partly irreversible). Precompression stress (σ_p) is most frequently determined based on the method by *Casagrande* (Casagrande, 1936); however,

a number of alternative approaches are found in the literature. For a review of different σ_p estimates refer to Grozic et al. (2003).

3.3.4.3 Triaxial Test

Triaxial tests mimic in situ stress distributions and are most realistic since horizontal and vertical stresses are defined and adjustable. However, they are also more difficult to conduct and more laborious than, for example, *oedometer* tests, where a high number of samples can be measured, and, hence, spatial heterogeneity of natural soils is better reflected. Triaxial tests on the other hand have the advantage that both volume and shape changes are determined simultaneously on the same sample and that failure occurs at the naturally weakest plane, which is important for deriving material parameters used in modeling approaches. In triaxial tests, undisturbed cylindrical soil samples are loaded with an increasing vertical principal stress (σ_1), while the horizontal principal stresses ($\sigma_2 = \sigma_3$) are kept constant (equal to the cell pressure). Shear stresses occur in any plane other than those of the principal stresses. Shear parameters (cohesion and angle of internal friction) can be determined from the slope of Coulomb's failure line embracing the Mohr circles at failure for various combinations of σ_1 and $\sigma_2 = \sigma_3$ (e.g., Kézdi (1974), Mitchell (1993), Wulfsohn et al. (1998)). Due to aggregate deterioration and prevented drainage of excess soil water, the *Mohr–Coulomb failure line* is bent toward smaller slope values at higher mean normal stresses. However, the numbers of contact points, strength per contact point, and pore geometry also affect triaxial test results.

There are three types of triaxial tests: (1) In the consolidated drained (CD) test, the soil sample is equilibrated with the mean normal stresses prior to an increase in the vertical stress (σ_1); the pore water drains off when the decrease in volume exceeds the air-filled pore space. Therefore, the applied stresses are assumed to be transmitted as effective stresses via the solid phase. However, Baumgartl (1991) found an additional change in the pore water pressure during extended CD triaxial tests depending on soil hydraulic properties. Thus, shear speed and low hydraulic conductivity, high tortuosity, and small hydraulic gradients further affect the drainage of excess soil water and the effective stresses (Horn et al., 1995). (2) In the consolidated undrained (CU) test, pore water is prevented from draining off the soil as vertical stress increases. Thus, high hydraulic gradients occur and the pore water reacts as a lubricant with a low surface tension. Thus, in the CU test (considering total stresses), shear

TABLE 3.1 Methods of Determining Soil Strength

Method	Dimension	Derived	Soil Condition
Uniaxial compression	Pressure (Pa)		Homogenized soil, structured bulk soil, and single aggregates
Confined compression test	Pressure (Pa)	Precompression stress (Pa)	Homogenized soil, structured bulk soil
Triaxial test	Pressure (Pa)	Cohesion (Pa), angle of internal friction (°)	Homogenized soil, structured bulk soil, and single aggregates
Direct shear test	Pressure (Pa)	Cohesion (Pa), angle of internal friction (°)	Homogenized soil, structured bulk soil, and single aggregates

parameters are much smaller and pore water pressure values are much greater than those in the CD test. (3) The highest neutral stresses and, therefore, the lowest shear stresses are measured in the unconsolidated undrained (UU) test, where neither the effective stresses nor the neutral stresses are equilibrated with the applied principal stresses at the beginning of the test. Thus, cohesion and angle of internal friction in terms of total stresses are strongly dependent on the compression and drainage conditions during deformation. In terms of strength of agricultural soils under traffic, texture, pore water pressure (in unsaturated conditions referred to as soil matric potential), nature of aggregates, and soil structure determine the preference for a particular test.

3.3.4.4 Cyclic Compression Test

The theoretical concept of precompression stress as an indicator for soil stability is based on the assumption that in the elastic range of deformation (recompression line), there is no permanent volume change. This is not strictly true since even at load magnitudes below σ_p, there is always a small but yet noticeable volume change. While volume changes for a low number of loading events below the precompression stress mostly only insignificantly affect soil porosity, the situation may change for a higher number of load repetitions. In fact, the formation of a plow layer and the development of a platy soil structure are the result of numerous loading events rather than the product of once exceeding the precompression stress value (Peth and Horn, 2006b). Volume change behavior for repeated loading conditions can be investigated by *cyclic compression tests*. Such tests allow estimating soil compaction due to short- and long-term multiple machine passages or animal trampling in grazing areas (Peth and Horn, 2006a; Krümmelbein et al., 2008). Samples are measured in a standard oedometer device with a constant load applied for a number of load cycles (e.g., 100 cycles). Soil deformation expressed as change in void ratio follows a logarithmic trend with number of load cycles (Figure 3.9). A simple model can be used to predict the compression of a soil subject to the same load for multiple loading repetitions (Peth, 2004):

$$e(N) = -C_n \times \log N + e_u^1. \tag{3.19}$$

Here, e and N are the void ratio and number of loading cycles, respectively. The axis intercept e_u^1 denotes the void ratio in the unloaded condition after the first loading event. Only the following loading events are considered to belong to the cyclic loading process. The slope C_n in Equation 3.19 is referred to as *cyclic compressibility* and indicates the sensitivity of a soil toward cyclic compression. The higher C_n the higher is the compression under repeated loading.

3.3.5 Shear Strength Parameters

The shear strength parameters, cohesion (*c*), and angle of internal friction (φ), may be determined by triaxial tests as described above or in a direct shear test. During direct shear tests, the type
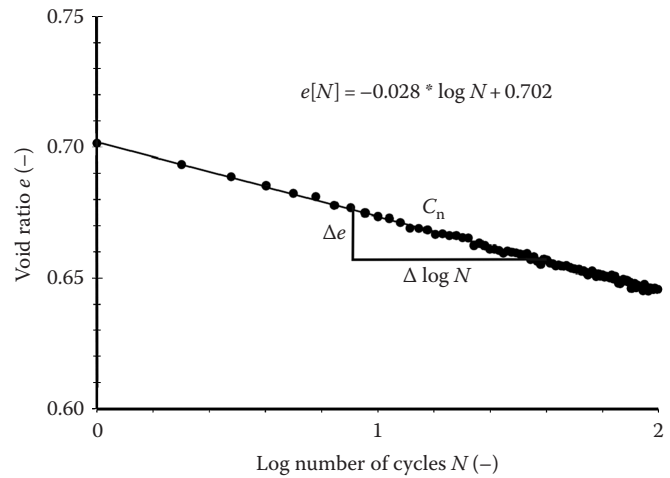
FIGURE 3.9 Compressibility as a function of repeated ("cyclic") loading. The slope C_n (cyclic compressibility index) denotes the sensitivity of the soil toward cyclic compression at the same magnitude of external load. (After Peth, S., and R. Horn. 2006b. The mechanical behavior of structured and homogenized soil under repeated loading. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 169:401–410.)

and direction of the shear plane, which is assumed to be affected only by normal and shear stresses, are fixed. Normal stress is applied to the specimen in the vertical and shear stress in the horizontal direction. To determine the Mohr–Coulomb failure line, at least four to five samples need to be tested each with a different normal stress. The maximum shear resistance (τ_{max}) is determined from a shear stress–displacement curve and plotted against the corresponding normal stress (σ_n). Plotting all pairs of τ_{max} versus σ_n gives the Mohr–Coulomb failure line in which the slope and intercept are the angle of internal friction (φ) and cohesion (*c*), respectively. As in triaxial tests, cohesion and angle of internal friction are influenced by shear speed and drainage conditions for a given soil.

3.3.6 Tensile Strength of Aggregates and Bulk Soil

Tensile strength of soils is a measure of the direct strength of interparticle bonds, which depend on chemical bonding, clay type and content, water potential (menisci forces), and organic matter. Also the internal geometry of particle arrangements and preexisting shrinkage cracks determine shear failure and crack propagation upon mechanical and hydraulic loads generating tensile forces. Tensile strength is an important indicator for the structural stability from the aggregate scale to the bulk soil level and a decisive factor for the preservation of physical quality of soils subject to mechanical stresses. The most widely used technique to measure tensile strength of soil aggregates is the *crushing test*. Soil aggregates are placed between two parallel plates in their most stable position (flat side downward) and the upper plate is loaded either with a weight (e.g., water dripping slowly into a bucket, Horn and Dexter, 1989, or by a load frame).

The force when failure occurs is recorded and tensile strength is calculated according to Dexter and Kroesbergen (1985):

$$Y = \frac{0.576 \times F}{d^2}, \tag{3.20}$$

where

   Y is the tensile strength
   F is the applied force
   d is the mean diameter of the aggregate given by

$$d = \frac{(d_x + d_y + d_z)}{3} \tag{3.21}$$

with indices *x*, *y*, and *z* denoting the aggregate diameter of the longest, intermediate, and shortest axes, respectively.

Other indirect techniques have been proposed to investigate crack growth and propagation in soils subject to tensile forces. One method is based on approaches adapted from fracture mechanics where crack formation and propagation are measured by so-called deep notch three-point bending tests (Hallett and Newson, 2001, 2005). Also direct tensile tests have been suggested to study crack propagation and tensile strength of soils, for example, using bone-shaped bulk soil samples that are torn apart by tension forces perpendicular to a predefined crack (Junge et al., 2000).

3.3.7 Other Methods to Determine Aggregate Stability

Aggregate stability may also be determined by wet sieving. Here, soil samples are prewetted to a given pore water pressure and then sieved through a set of sieves from 8 to 2 mm diameter. The difference between the aggregate-size distribution at the beginning and end of sieving under water for a given time is calculated as the mean weight diameter (MWD). This value is qualitatively related to aggregate strength and increases with increasing aggregate stability. The wettability of the solid interfaces has a pronounced effect on the test results in wet sieving and direct comparison with strength parameters derived from mechanical stress applications through polar loads is limited. Aggregate stability has also been measured by the application of ultrasonic dispersion techniques where aggregate stability can be expressed by the input of ultrasound energy resulting in the breakup of aggregates into smaller microaggregate units up to single particles (Raine and So, 1994). A comparison of different methods to determine aggregate stability is given by Baumgartl and Horn (1993).

3.4 Effect of Soil Structure on Soil Strength and Deformation Processes

As mechanical stresses are applied, soil deformation occurs first at the weakest point(s) in the soil matrix and further increases in stress resulting in the formation of failure zones. The strength of the failure zone is equal to the energy required to create a new unit of surface area or to initiate a crack (Skidmore and Powers, 1982) and is called the apparent surface energy (Hadas, 1987). Consequently, soil stability against shear or tensile stresses is related to strength distribution in failure zones. In principle, soil structure will be stable if the applied stress is smaller than the strength of the failure zone, that is, if the bond strength at the points of contact exceeds the shear or tensile stresses generated by external loads. If the resisting forces are smaller than the active forces, stresses are in disequilibrium and hence soil deformation will occur to generate more contact points until stress equilibrium is reached again. Reorientation of particles is accompanied by a change in soil structure and consequently functions. In extreme stress situations (especially high shear stresses) or when soil is subject to mechanical loads in unfavorable moisture conditions (near the liquid limit), soil structure may be almost completely returned to an immature state (homogeneous soil). In contrast to shear stresses that lead to volume constant deformation, normal stresses result in volume change, that is, compression, which is plastic as soon as the soil stability is exceeded. In situ stress conditions during field operations (traffic, soil–tool interactions) are generally characterized by a combination of both shear and compressive stresses.

3.4.1 Mechanical Properties of Homogenized Soils

If only mechanical properties of homogenized soils are compared, gravelly soils are less compressible than sandy or finer-textured soils (Horn, 1988), and any deviation of the particles from the spherical shape results in an increase in the shearing resistance and soil strength (Gudehus, 1981; Hartge and Horn, 1999). Ellies et al. (1995) found that the angle of internal friction for medium and coarse sands at comparable bulk density and water content but of different origins (quartz, basalt, volcanic ash, and shells) ranged from 25° to 50°. Soil strength also depends on the type and content of clay minerals and exchangeable cations. At the same bulk density and pore water pressure, compressibility of homogenized soil increases with clay content and decreases with soil organic-matter (SOM) content. Vice versa soils at the same clay content are more readily compressed at lower bulk density or higher moisture content. Cohesive forces between illite, smectite, and vermiculite are greater than for kaolinite, which in turn has a larger angle of internal friction because of its size and particle shape. Additionally, the higher the valency of the adsorbed cations and the higher the ionic strength of the soil solution, the greater is the shearing resistance under otherwise similar conditions (Yong and Warkentin, 1966; Mitchell, 1993).

Shearing resistance also increases depending on both the nature and content of SOM. In agricultural soils with the same physical and chemical properties, the mechanical parameters (precompression stress, angle of internal friction, and cohesion) are greater when the relative contents of carbohydrates, condensed lignin subunits, bound fatty acids, and aliphatic polymers are higher (Hempfling et al., 1990).

3.4.2 Mechanical Properties of Aggregated Soils

Mechanical strength of structured soils depends on aggregation and history of hydraulic stresses (degree of maximum predrying), as well as the shape and internal arrangement of particles, microaggregates, and voids. At comparable grain-size distribution, bulk density, and pore water pressure, soil strength increases with aggregation (i.e., coherent < prismatic < blocky < subangular blocky < crumbly). In the case of a platy structure, the strength depends on the direction of shear forces relative to the preferred orientation of the particles. In the direction of the elongated axes of aligned particles, shear strength is lower than perpendicular to it. This means that not only mechanical stresses are anisotropic but also strength properties that in turn are a function of aggregation. However, both the magnitude of previous hydraulic stresses (predrying intensity) and their dynamic changes determine soil strength. Therefore, mechanical properties also depend on the frequency of swell/shrink and wet/dry events and on the actual pore water pressure at the time of loading. Soils become stronger when redried and rewetted and when pore water pressure gradients over longer distances promote particle movement and rearrangement until entropy is minimized (Semmel et al., 1990). The effect of pore water pressure on strength itself is governed by the theory of effective stresses in unsaturated soils (e.g., *Bishop*'s effective stress equation (Equation 3.2)). It has been extended from the early considerations of *Terzaghi* (Terzaghi, 1936) for saturated soils; however, its application remains difficult since the evaluation of the parameter χ in Equation 3.2 is still complicated. It accounts for the tradeoff between pore water pressure (increases interparticle forces with curvature of menisci) and effective area of interparticle contact (when decreasing, the effect of pore water pressure is reduced). The interrelationship of pore water pressure and effective area of interparticle contact is structure dependent and hence also χ is related to structure and not as sometimes assumed purely a function of the degree of saturation. If the relative reduction in water-filled pores is smaller than the actual decrease in pore water pressure (more negative matric potential), soil strength increases while if the decrease in the effective surface of water menisci exceeds the decrease in matric potential (getting more negative), the effective stress of such drier soils gets absolutely smaller. In principle, soil strength is promoted by two different mechanisms: (1) the increase in strength results from an increase in the total number of contact points between single particles (i.e., an increase in effective stress) and (2) the increase in shear resistance per contact point (Hartge and Horn, 1984). Therefore, even if soil bulk densities are similar, strength properties may be quite different. Also freezing and thawing affect soil strength, because aggregates become either denser or they are destroyed by ice lens formation during freezing. The exfoliation process starts from the outer skin. Both effects are called soil curing, but result in completely different strength values and physical properties of the bulk soil (Horn, 1985b; Kay, 1990). Pedogenic effects on natural mechanical strength of three soils are illustrated in Figure 3.10. Luvisols derived from loess are characterized by clay
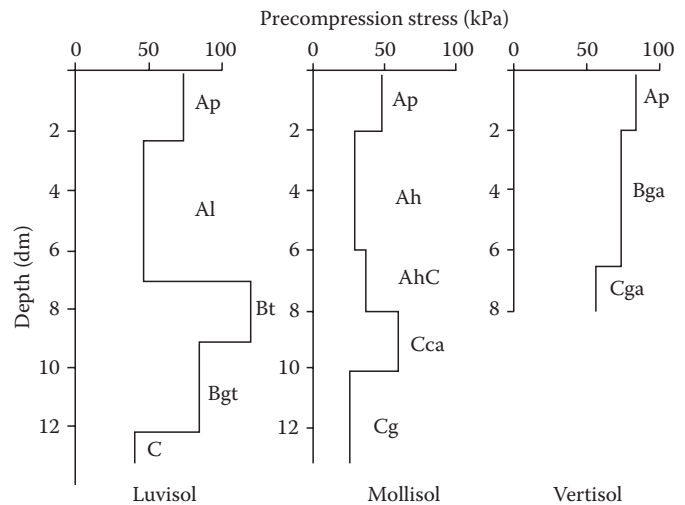
FIGURE 3.10 Precompression stress values for three soil profiles at soil matric potential of $\Psi_m = -6\,\text{kPa}$. (Modified from Horn, R. 1994b. Stress transmission and recompaction in tilled and segmently disturbed soils under trafficking, p. 53–87. *In* N.S. Jayawardane and E. Stewart (eds.) Subsoil management techniques. Lewis Publishers, Boca Raton, FL.)

illuviation, which results in strength decrease in the *Al* horizon and an increase in the *Bt* horizon due to aggregation. On the other hand, calcium precipitation in the Cca horizon of the mollisol also leads to a strength increase. In all three soils, the parent material (C horizon) was weakest although overburden pressure increases with soil depth suggesting the opposite should have been the case. This underlines the offset of soil strength parameters by pedogenic processes. On the other hand, anthropogenic processes such as annual plowing and tractor traffic may create very strong plow pan layers with precompression stress values similar to the contact pressure of a tractor tire or even higher (up to 300%) due to lug effects. The transition from the region of overconsolidation to the virgin compression line depends to a great extent on internal parameters and further on the hydraulic and mechanical stress history. In addition, strength decreases in the *Ap* horizon due to plowing and seedbed preparation can be observed until texture-dependent values are reached.

Pronounced strength decreases are also found in soil profiles that have been deep-loosened or deep-plowed (up to 60 cm) or loosened and/or mixed with other soil material. In the case of a partial deep loosening by a trenching slit plow (Reich et al., 1985; Blackwell et al., 1989), traffic must be restricted by the use of smaller machinery or controlled tracks perpendicular to the homogenized soil volume.

Soils with a well-developed vertical pore system are stronger than those with randomized or horizontal pore arrangements since elongated pores are more stable with their symmetry axis oriented in the direction of the major principle stress (σ_1) than perpendicular to it. Therefore, vertical (bio) pores are less prone to compression than oblique or horizontally oriented pores and, thus, untilled or minimum tilled soils are stronger than conventionally tilled soils (Horn, 1986).

Each soil deformation requires air-filled pore space and sufficiently high hydraulic conductivity during compression to drain off the excess pore water. The smaller the hydraulic conductivity, the hydraulic potential gradient, and the pore continuity, the more stable is the soil especially during short-term loading. In sandy soils, this effect is small and initial settlement equals the total strain. With increasing clay content, however, the proportion of initial to primary and secondary consolidation is reduced and time-dependent soil settlement becomes important. This results in an increase in precompression stresses for short-term loading due to an increase in pore water pressure that dissipates only slowly depending on the local hydraulic conductivity. Such drainage effects on the consolidation process are smaller in more strongly aggregated soils and with coarser soil textures.

The development of structure generally results in increasing soil strength since the formation of secondary large pores is always accompanied by the development of denser aggregates. Although the aggregate assemblages carry the same stresses over fewer contact points, the aggregates themselves are more resistant to shear stresses. Strength differences may also result from changes in the angle of internal friction and cohesion as indicated in Figure 3.11 for a single aggregate, an undisturbed aggregate, and a homogenized bulk soil, respectively. More intense aggregation increases soil strength at comparable hydraulic stresses.
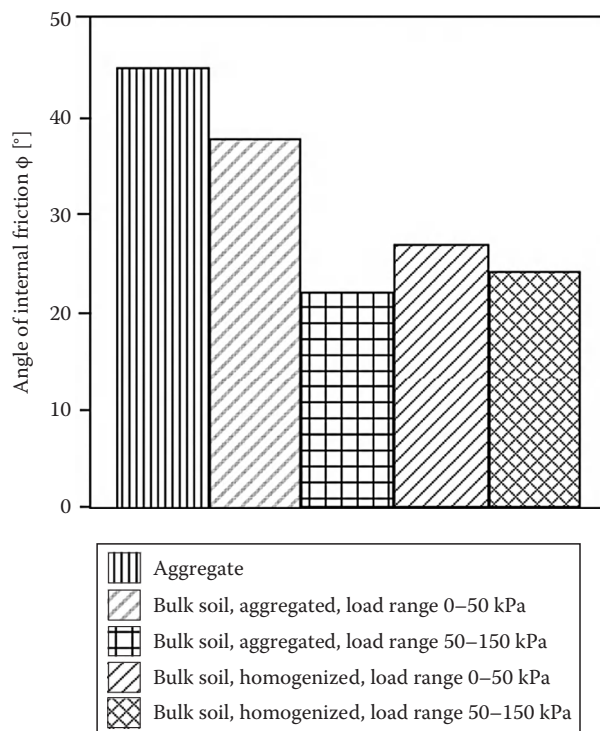
FIGURE 3.11 Changes in shear strength at constant soil matric potential $\Psi_m = -6\,kPa$ for various applied stress ranges for single aggregates, bulk soil, and homogenized material. (Modified from Horn, R., T. Baumgartl, R. Kayser, and S. Baasch. 1995. Effect of aggregate strength on changes in strength and stress distribution in structured bulk soils, p. 31–52. *In* K.H. Hartge and R. Stewart (eds.) Soil structure: Its development and function. CRC Press, Boca Raton, FL.)

Angle of internal friction for a single aggregate (~45°) is very high compared to bulk soil samples. However, aggregated bulk soil reveals higher friction angles than the homogenized soil at the same bulk density. This shear strength gain is lost when stresses exceed the overall stability of the aggregated sample leading to a reduction in the angle of internal friction to values similar for homogenized soil. When the maximum value of stress exceeds the maximum strength such that structure is homogenized, only texture-dependent properties remain.

3.4.3 Soil Deformation under Dynamic Stresses

Standard mechanical stability indicators such as shear strength parameters (ϕ and c) and precompression stress (σ_p) are without dispute suitable for characterizing the mechanical strength/stability of soils and to assess if significant irreversible volumetric and shear deformation under given stress conditions are likely to occur. Horn et al. (2005), for example, have derived stability maps on various scales down to farm level where sensitive and less sensitive areas are outlined, which could help farmers to avoid harmful subsoil compaction. We must, however, consider more in detail the repeated short-term loading because multiple loading events cumulatively add to the total compression, although concerning precompression stresses no further compression maybe expected. Zapf (1997), for example, argued that the increase in irreversible deformation of arable soil by agricultural machinery is not only related to the increasing mass of machines but also related to the increase in wheeling frequency. Peth and Horn (2006b) estimated that, based on the calculations of Olfe (1995), for an average-sized wheat-production farm in a time period of 5 years, the number of load repetitions may add up to 50 events for 85% of the field and up to 100 events for permanent wheeling tracks.

Cyclic compression tests allow estimating soil deformation due to repeated loading and hence assessing the effect of long-term field traffic on subsoil compaction. Similar to other strength parameters, cyclic compressibility (see coefficient C_n in Equation 3.19) for homogenous soils at the same bulk density is strongly texture dependent (Figure 3.12). The finer the texture and especially the higher the clay content, the higher is the sensitivity against cyclic compression (higher C_n values). Porosity changes with increasing number of loading cycles are significant and accumulate to up to ~3.0 vol% after 100 cycles (for sandy loam, Figure 3.12). Assuming that mostly larger pores are affected by compression, this would mean a reduction in air capacity by additional 3.0 vol%, although the applied load (40 kPa) was relatively low. Cyclic compressibility tests on undisturbed field samples reflect also the influence of structure and tillage management (Figure 3.13). Correlation between bulk density and C_n was higher for conventionally tilled soils and only weak for the conservation tillage system where C_n values were on an average lower at similar bulk densities. The lack of a strong correlation between bulk density and C_n suggests a stabilizing effect of structure. Some of the aggregated soils show during cyclic loading a stepwise compression curve with a few quasistable plateaus followed by stronger settlements until deformation is leveling off
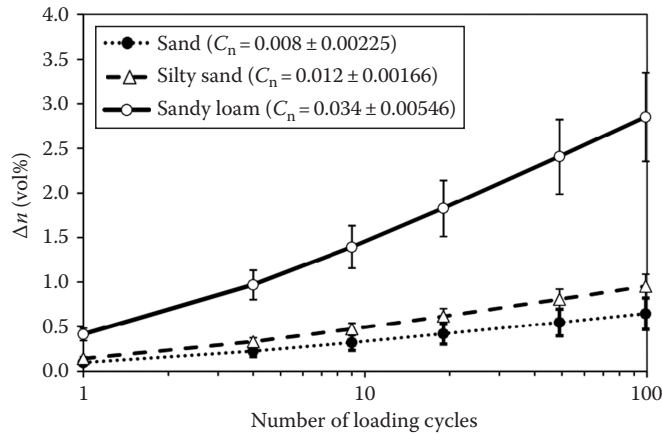
FIGURE 3.12 Changes in porosity for different soil textures (homogenized samples) due to cyclic loading with a load of 40 kPa at the same initial bulk density (1.62 g cm⁻³). Samples have been equilibrated to a standard matric potential of −6 kPa prior to the tests. Note that the change in porosity after the first loading–unloading event is omitted here (for sand = 1.28, for silty sand = 1.22, and for sandy loam = 2.12 vol%). Error bars indicate standard deviation ($n = 14$). Corresponding C_n values are given in brackets (mean ± SD). (From Peth, S., and R. Horn. 2006b. The mechanical behavior of structured and homogenized soil under repeated loading. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 169:401–410.)
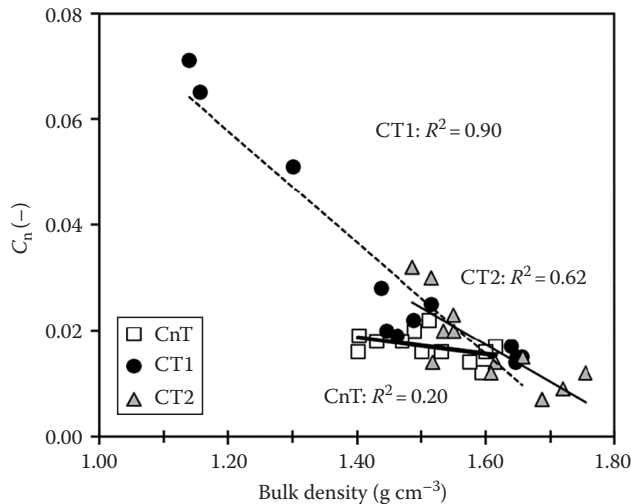
FIGURE 3.13 Correlation between initial soil bulk density and cyclic compressibility coefficients (C_n) of a haplic luvisol (Marienstein/Germany) with silt loam texture for different tillage systems: CT1 and CT2 are two different conventionally tilled sites; CnT is conservation tillage. (From Peth, S., and R. Horn. 2006b. The mechanical behavior of structured and homogenized soil under repeated loading. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 169:401–410.)

again (Figure 3.14). This behavior could be interpreted as stable aggregate units breaking under repeated loading just like steel may break due to material wear. Although these results still have preliminary character and certainly more systematic tests are needed, they demonstrate the importance of considering dynamic stress paths in soil testing in order to derive realistic estimates of compaction under field loading conditions.
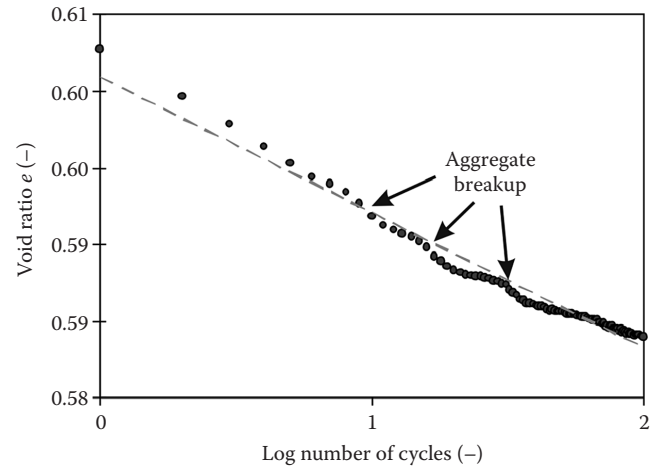
FIGURE 3.14 Aggregate breakup during cyclic loading ("cyclic wear of aggregates"). (After Peth, S., J. Rostek, A. Zink, A. Mordhorst, and R. Horn. 2010. Soil testing of dynamic deformation processes of arable soils. Soil Till. Res. 106:317–328.)

3.5 External Mechanical Forces and Stress Distribution in Soils

3.5.1 Stress Application and Variations in the Contact Area

Any load applied at the surface is transmitted to the soil in three dimensions by the solid, liquid, and gas phases. If air permeability is high enough to allow immediate deformation of the air-filled pores, soil settlement is mainly affected by fluid flow. However, fluid flow may be delayed because changes in water content or pore water pressure depend on the hydraulic conductivity, gradient, and pore continuity. Thus, the intensity and form of pressure transmission are again affected by soil strength. In the following discussion, stress distribution in both homogenized soils and aggregated systems will be defined. Based on the theory of Boussinesq (1885) who solved the problem of a point load acting perpendicular on an infinite half-space consisting of homogeneous, isotropic, and linear elastic material, Fröhlich (1934) introduced a textural-dependent concentration factor (v_k) to give more "flexibility" to the original Boussinesq solution for determining stress propagation in soils. Under saturated conditions, v_k ranges between 3 for very strong material (which is defined as linear elastic and corresponds to the Boussinesq solution) and 9 for soft soil becoming higher with increasing clay content. In weak soils with high concentration factor, stresses are transmitted to greater depth but remain concentrated around the perpendicular line of the load center. On the other hand, in strong soils with low values of the concentration factor, stresses are transmitted more horizontally and shallower.

At the same contact pressure, stresses are transmitted deeper when the contact area is larger. Furthermore, stress distribution patterns in the soil are not only different for tire lugs and the intervening area but are also affected by the stiffness of the carcass (Horn et al., 1987). Thus, there are no well-defined equipotential stress lines in soils, but the concentration factor values must be related
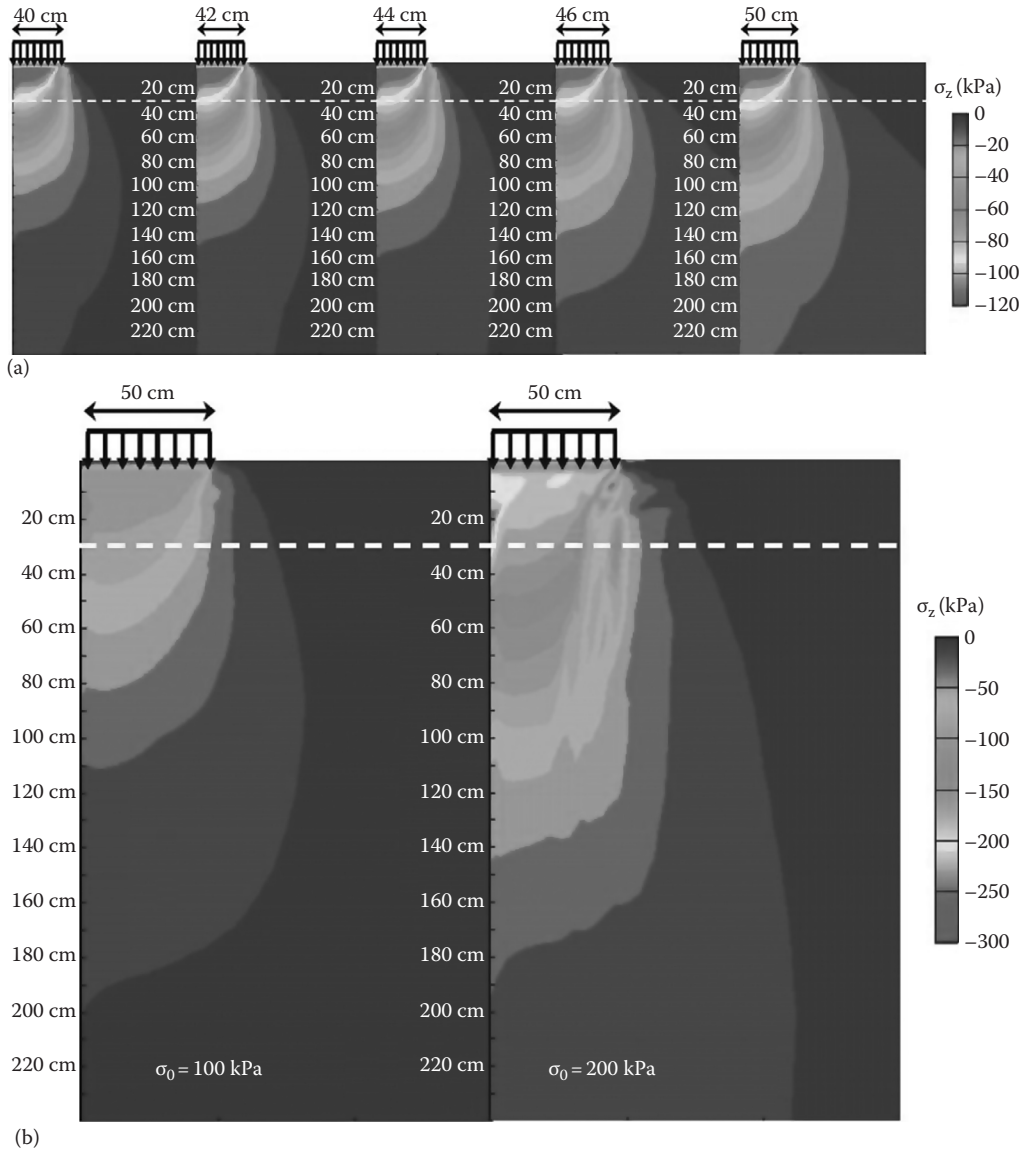
FIGURE 3.15 (a) Stress distribution under a tire with the same ground contact pressure of 100 kPa but increasing contact area (tire size increases from left to right) and (b) for the same tire size (100 cm diameter) but increasing ground contact pressure (σ_0) due to higher wheel load, respectively. Vertical stresses σ_z (kPa) have been modeled with an FEM approach assuming rotational symmetry of the contact area (i.e., circular tire-soil contact shape). Note that only a part of the model domain is shown. The width of the model domain was 4.5 m to reduce boundary effects. The horizontal white-dashed line indicates the usual plowing depth of 30 cm. Mechanical properties for the topsoil and the subsoil were assumed to be the same. Following mechanical parameters were used for the simulation: cohesion $c = 10$ kPa, angle of internal friction $\phi = 20°$, stress-dependent bulk modulus $K(\sigma)\,\text{kPa} = 800 \times \sigma_m^{0.5} + 50$, stress-dependent shear modulus $G(\sigma)\,\text{kPa} = (270 \times \sigma_m^{0.5})(1 - (\tau/\tau_f)^{0.37}) + 50$, where the yield stress $\tau_f(\sigma_n) = c + \sigma_n \tan \phi$. For further details on material functions and parameters, see Section 3.2.3 and Richards and Peth (2009).

to precompression stresses in relation to applied stress and contact area for a given texture (DVWK, 1995). If additionally applied stresses do not exceed the internal soil strength defined by the precompression stress, the concentration factor is smaller as compared to the situation when the stresses result in a plastic deformation. The latter is defined by higher values (Berli et al., 2003).

The speed effect of wheeling, as well as the kind of stress application (tire, rubber belt tire-inflation pressure), affects the stress path at the soil surface and also causes enormous differences in the stress propagation inside the soil volume. It is well

understood that at a given surface load, the stress propagation reaches the deeper the larger the contact area and that at a given contact area stresses are transmitted deeper for higher ground contact pressures (Peth and Horn, 2004; Figure 3.15a and b).

When we consider the stress vector field during wheeling, the tangential redirection of stresses increases with stronger wheel slip and less rigid tire or rubber belt (Figure 3.16a and b). Keller (2004), Arvidsson and Keller (2007), Cui et al. (2007), and Keller et al. (2007a) proved the effect of various tire forms and tire-inflation pressure on soil stress distribution and not
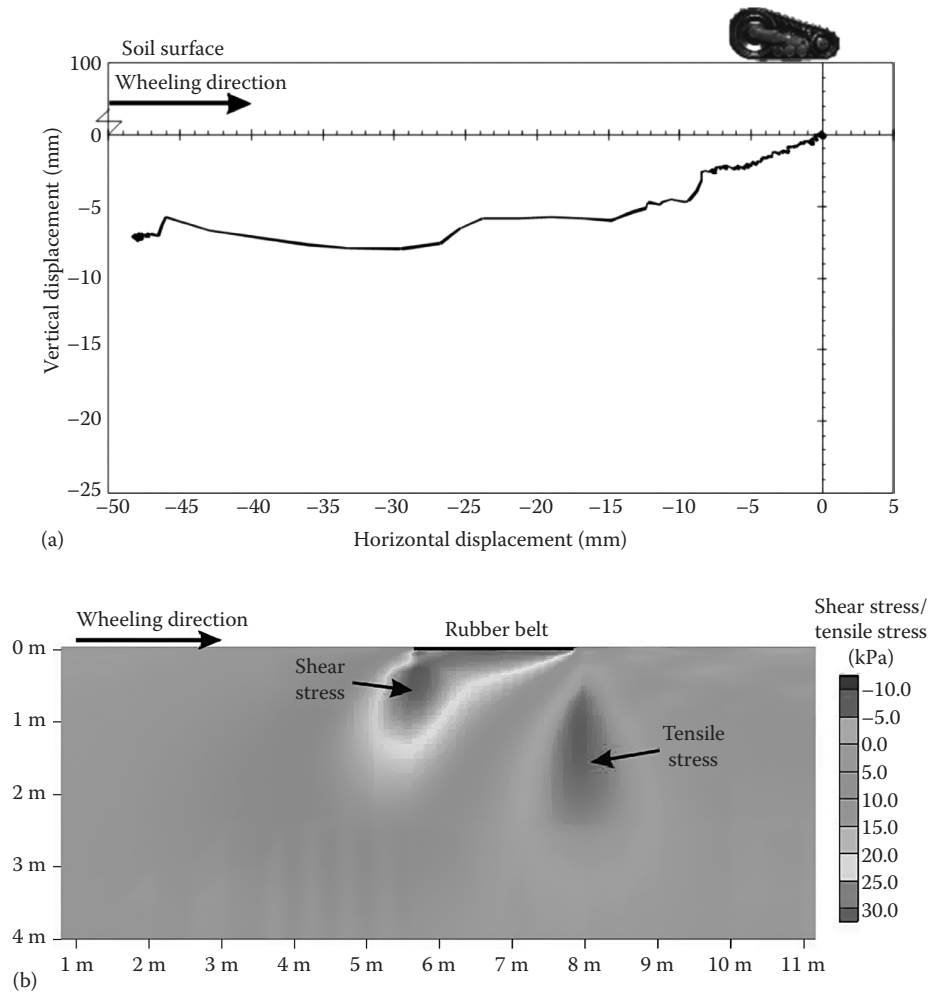
FIGURE 3.16 (a) Vertical and horizontal movement of a displacement sensor installed in situ at 10 cm depth during wheeling with a rubber belt tractor. (b) Modeled (FEM) distribution of shear/tensile stresses under a rubber belt tractor during wheeling. Note that material properties used for the FEM are not the same as for the site where soil displacement has been measured in situ (a). Simulated stress distribution has been conducted for a hypothetical case. Material properties and boundary conditions used in (b) were as follows: belt length = 240 cm, tractor weight = 12.6 Mg, ground contact pressure σ_0 = 67 kPa, horizontal stress transmitted via the belt = 32.5 kPa, cohesion c = 10 kPa, angle of internal friction ϕ = 20°, stress-dependent bulk modulus $K(\sigma)\,\text{kPa} = 800 \times \sigma_m^{0.5} + 50$, stress-dependent shear modulus $G(\sigma)\,\text{kPa} = (270 \times \sigma_m^{0.5})(1 - (\tau/\tau_f)^{0.37}) + 50$, where the yield stress $\tau_f\,(\sigma_n) = c + \sigma_n \tan \phi$. For further details on material functions and parameters, see Section 3.2.3 and Richards and Peth (2009).

only showed the heterogeneity of the stress distribution but also quantified the effect of changes in the tire forms and inflation pressure on stress distribution. Burt et al. (1989) already described the enormous heterogeneity in the stress distribution within the contact area that differed up to 300% and was higher beneath the lugs and the lowest in between them close to the middle part of the area while due to carcass effects it reached extremely high values at the outer edge of the tire.

The question about the effect of traffic on stress distribution and its consequences on soil properties remains, even after more than three decades of intensive discussion, still controversial, which is most often caused by different scientific views or interests in the subject. It is to be understood that the machine industry is mostly interested in the development of more effective and more powerful machines resulting in heavier traffic with higher wheel

load and wheel contact area. The benefits of new and smaller more intelligent self-steering GPS-controlled agricultural machines are still under debate, although it is a positive way to minimize the still increasing and deeper stress transmission into soils taking place when using conventional (heavier) machines. On the other hand, when soil scientists deal with traffic–soil interactions, they are mostly concerned about the effects of stress application on soil properties and the future effects for crop production, water infiltration, filtering and buffering, and groundwater recharge. Recently, enhanced greenhouse gas emissions have been more often associated with the interaction of soil stress and soil deformation, which results in changes in aeration and microbial activity. Legislation finally has to link both interests where stakeholders have to decide based on their administrative perspectives and given laws. The debate on soil use

and soil protection has resulted in Germany in the German Soil Protection Law (BBodSchG, 1998), while at European level such a law is still under discussion. However, the liability varies to a great extent even in between different provinces within Germany and the definition of main soil parameters (indicators) is not finalized (Horn and Fleige, 2009).

3.5.2 Stress Effects in the Soil

3.5.2.1 Internal Parameters

At a given bulk density and pore water pressure, applied stress at the soil surface will be transmitted deeper as silt and clay contents increase while stress attenuation will be greater in a smaller volume as the soil dries. For a given particle-size distribution, water content, and bulk density, stress attenuation will increase with increasing SOM content. Based on many stress distribution measurements in the field and in monoliths, a general correlation scheme involving texture, precompression stress, and tire contact area to derive the concentration factor v_k has been used to approximately predict the stress distribution in soils (DVWK, 1995); for more information see also Horn and Fleige (2003).

3.5.2.2 Structure Effects on Stress Attenuation

The effect of aggregation on stress distribution and its consequences for ecological parameters are well understood. Under in situ conditions for soils with the same internal parameters, stress attenuation is greater the more aggregated soils are. Concentration factor values expressed as precompression stress are smaller for better aggregated and drier soils (DVWK, 1995). Not only the pressure but also the size and shape of the contact area affect stress distribution in unsaturated structured soils.

Stress distribution calculations in a luvisol derived from loess under natural forest and its comparison with the identical computation under arable management show the effect of soil strength on stress attenuation (Figure 3.17a and b). Under natural forest, especially the clay-enriched Bt horizon with a prismatic–blocky structure (matric potential ~ −30 kPa) attenuated the applied stresses of the tractor and both combine harvesters while the topsoil (Ah- and the clay-depleted Al horizons), respectively, with a crumbly or coherent structure was deformed.

Under arable conditions, approximately 1/4th of the pores within the top 100 cm were lost compared with the forest site, which also resulted in higher precompression stress at all depths
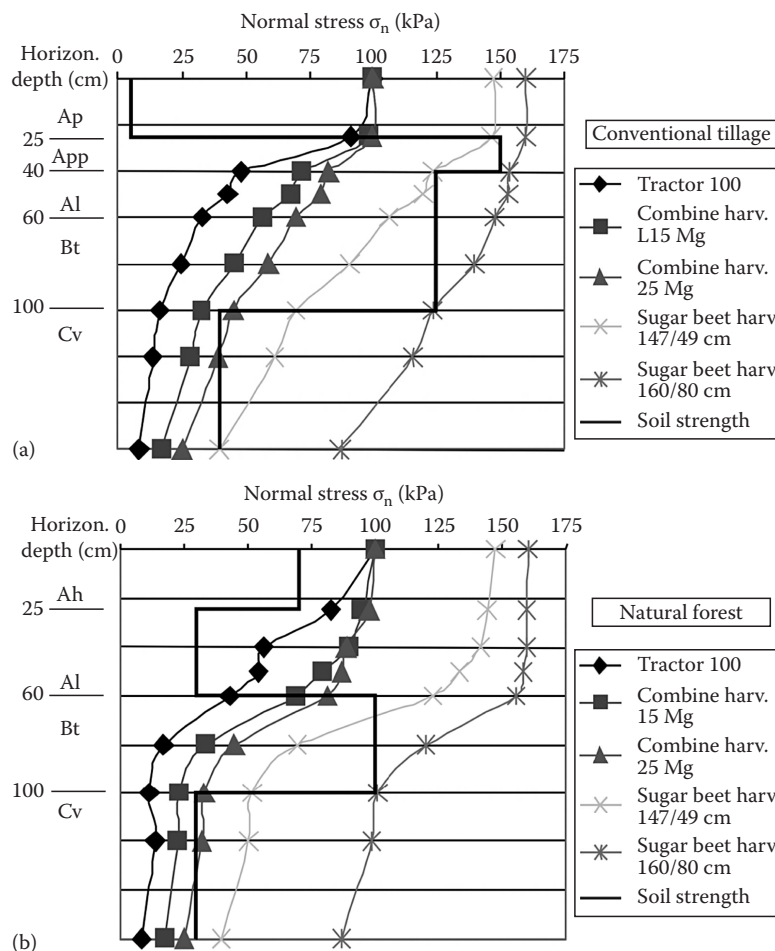
FIGURE 3.17 Modeled stress distribution in a luvisol derived from loess under forest (b) and under conventional tillage (a). The concentration factors were defined according to the degree of aggregation.
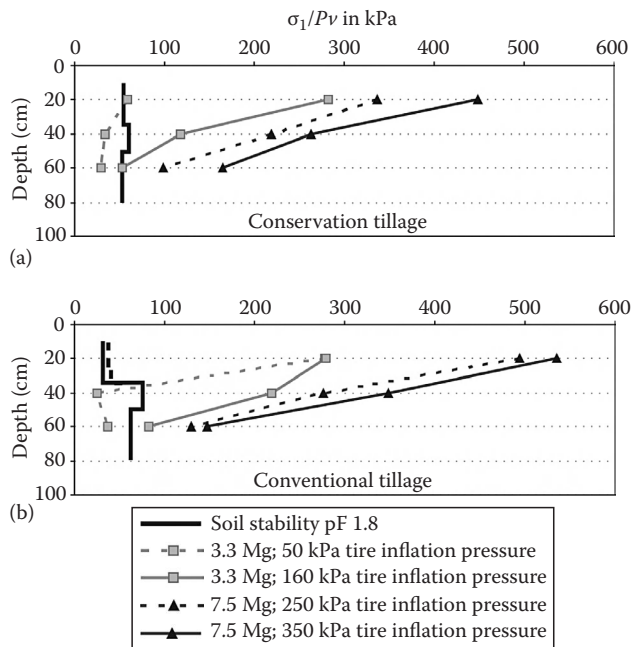
FIGURE 3.18 Comparison of stress attenuation for different wheel loads and tire-inflation pressures in a luvisol derived from loess under conservation (a) and conventional tillage (b). Dashed and solid lines with symbols indicate stresses derived from SST (stress rate transducer) measurements; solid line without symbol indicates horizon-specific precompression stresses. (From Zink, A. 2009. Bodenstabilität und Auswirkungen dynamischer Lasteinträge auf physikalische Eigenschaften von Ackerböden unter konservierender und konventioneller Bodenbearbeitung. Ph.D. Thesis. Schriftenreihe Institut für Pflanzenernährung und Bodenkunde, Germany, No. 84, 180 pp. ISSN: 0933-680X.)

except for the Ap horizon and a more complete stress attenuation apart from the homogenized seedbed. However, the first introduction of the heavier sugar beet harvesters caused a deeper and a more intense soil deformation while exceeding the corresponding actual soil strength (Peth et al., 2006). The effect of stress attenuation depending on soil structure and strength (as a function of depth) can be also derived from Figure 3.18 (Zink, 2009). The luvisol derived from loess under long-term conventional and conservation tillage management shows limited stress attenuation at given wheel loads with and without reduced inflation pressure, which causes irreversible soil deformation (stresses exceed the precompression stress).

Repeated short-term stress application results in an increase in the matric potential and a higher water saturation within the various horizons. Consequently, the repeated short-term loading not only weakens the soil structure by deformation but also pumps additional soil water from deeper depths (repeated elastic rebound creates a suction mechanism), which further enhances soil deformation. Semmel (1993) could prove that due to repeated wheeling-induced soil softening, horizontal minor stresses decreased while the major vertical stress increased during 50 repeated wheeling events with a 3.7 Mg tractor. The soil softening was also reflected by reduced soil strength and an increase in the concentration factor due to repeated wheeling. The same processes also occur in grassland ecosystems under arid conditions during sheep trampling. Krümmelbein et al. (2008) could show that in grassland soils, which are intensely trampled during winter, soil strength is reduced due to "kneading," which negatively affected hydraulic properties. Also, tree harvesting in forest soils causes an intense increase in the precompression stresses, formation of deeper ruts, reduced aeration, and a restricted plant growth as a result of dynamic soil loading (Vossbrink and Horn, 2004; Horn et al., 2007b). The controlled traffic lane concept could prevent large-scale soil deformation by restricting mechanical stresses to defined lanes, hence protecting arable and forest land even during intensive future land use (Alakukku et al., 2003; Chamen et al., 2003; Watts et al., 2005).

An interesting aspect of soil deformation caused by extremely heavy loads is shown in Figure 3.19. During a first time wheeling event with a sugar beet harvester (35 Mg), a formerly formed rigid plow pan layer lying on top of a weaker subsoil may even be destroyed by breaking the platy structure due to shear stress concentrations in the plow pan (Fazekas, 2005; Peth et al., 2006).
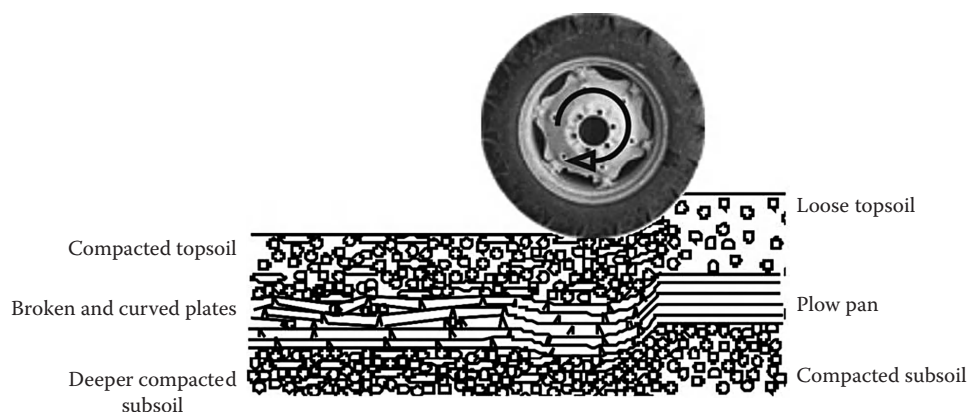
FIGURE 3.19 Schematic sketch of stress-induced changes in the plow pan layer as a consequence of the initial wheeling with a heavy sugar beet harvester. (Modified from Fazekas, O. 2005. Bedeutung von Bodenstruktur und Wasserspannung als stabilisierende Kenngrößen gegen intensive mechanische Belastungen in einer Parabraunerde aus Löss unter Pflug- und Mulchsaat. Ph.D. Thesis. Schriftenreihe des Institut für Pflanzenernährung und Bodenkunde, Germany, No. 67, 170 pp. ISSN: 0933-680X.)

Repeated wheeling of such a site may then result in deeper stress penetration and consequently aggravates subsoil compaction.

3.5.3 Effect of Stress Application on Soil Strain

If external stress is smaller than internal soil strength, no deformation results and vice versa. However, because the latter includes soil compaction and shear processes, both the rut depth and the vertical movement of a given soil volume below the rut must be known if we want to distinguish between both processes. The extent to which soil deformation occurs during traffic and the extent to which various tillage implements (conventional/conservation) deform a soil at a given pore water pressure are shown in Figure 3.20. In the conventional tillage treatment in a loessial luvisol, passage of a tractor (front/rear wheel) results in a pronounced vertical (up to 8 cm) and horizontal forward and backward (up to 2 cm) displacement. Under conservation tillage, the same tractor results in smaller soil deformations because of a higher internal

(a)

(b)

Wheel load: 5 t
Water suction: −60 hPa
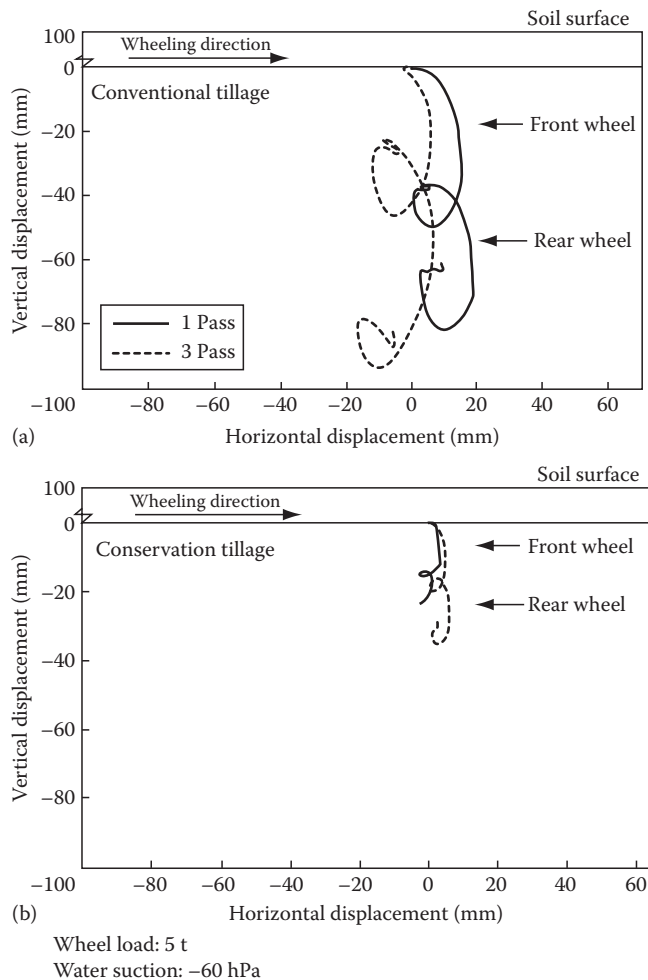
FIGURE 3.20 Strain distribution in structured soils due to traffic. Particle movement is more pronounced in the conventionally tilled luvisol (a) while soil deformation is less intensive in the conservation tillage plot (b). (From Wiermann, C., D. Werner, R. Horn, J. Rostek, and B. Werner. 2000. Stress/strain processes in a structured unsaturated silty loam Luvisol under different tillage treatments in Germany. Soil Till. Res. 53:117–128.)
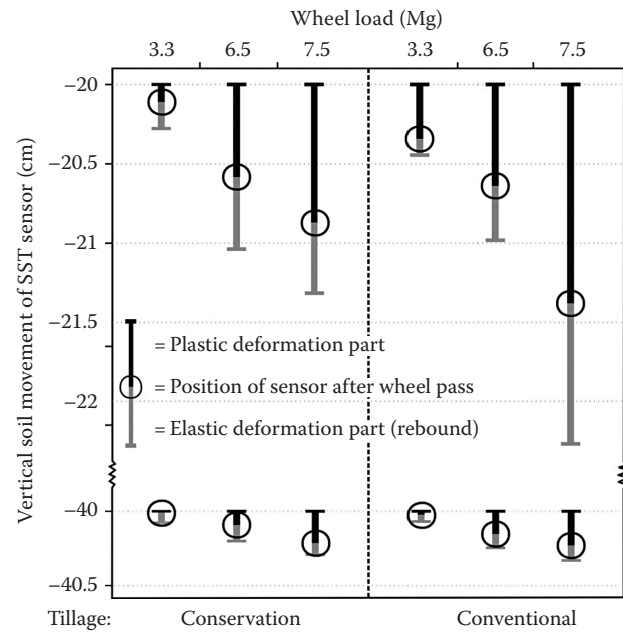
FIGURE 3.21 Vertical-elastic (reversible) and plastic (irreversible) soil displacement in a luvisol derived from loess for different wheel loads and management systems. Soil movement has been measured with a DTS/SST (displacement transmitter/stress state transducer system). (From Zink, A. 2009. Bodenstabilität und Auswirkungen dynamischer Lasteinträge auf physikalische Eigenschaften von Ackerböden unter konservierender und konventioneller Bodenbearbeitung. Ph.D. Thesis. Schriftenreihe Institut für Pflanzenernährung und Bodenkunde, Germany, No. 84, 180 pp. ISSN: 0933-680X.)

soil strength leading to a maximum vertical displacement of <4 cm after three traffic events and a much less pronounced horizontal displacement. With increasing aggregate development, soil strength increases and aggregate deterioration is less pronounced during displacement and alteration of the pore system due to the infilling of interaggregate pores by smaller particles. Nevertheless, all stresses that are not attenuated to levels below soil strength result in volume alterations, even if the applied stresses vary for different soil types, land uses, tillage systems, and environmental conditions. The intense effect of wheel load on soil strain down to 40 cm depth in luvisols derived from loess under conventional and conservation management depicts the sensitivity of stress application on changes in the pore system (Figure 3.21). With increasing wheel load at a given soil stress, the vertical height change is more intense under conventional than under long-term conservation tillage practice and affects also deeper soil depths irreversibly, which was also stated by Richards and Peth (2009).

3.6 Soil Deformation and its Effects on Physical and Physicochemical Properties

At mechanical failure, physical, chemical, and biological as well as physicochemical properties are affected. Stepniewski et al. (1994), Pagliai and Jones (2002), Lipiec and Hatano (2003), Lipiec et al.

(2004), and Horn et al. (2006) summed up the present knowledge and underline the detrimental effects of soil deformation on physical, chemical, and biological soil properties and functions. Larink et al. (2001) and Langmaack et al. (2002) emphasize the stress effect on biological activity and regeneration after stress application and observed an intense change in species abundances. Numerous papers defined in addition the effect of soil deformation of penetration resistance, which is often primarily linked with root growth, but these data can also be used to define in situ "stresses at rest" as a measure of the actual compression status of plots, fields, or even landscapes (Horn et al., 2007a).

Concerning changes in hydraulic properties under static loading, which are of main importance for numerous processes, it becomes obvious that both capacity properties like pore-size distribution (water retention characteristic) and intensity relations like the unsaturated hydraulic conductivity function show a severe change with applied stress (Figure 3.22; Horn et al., 1995).
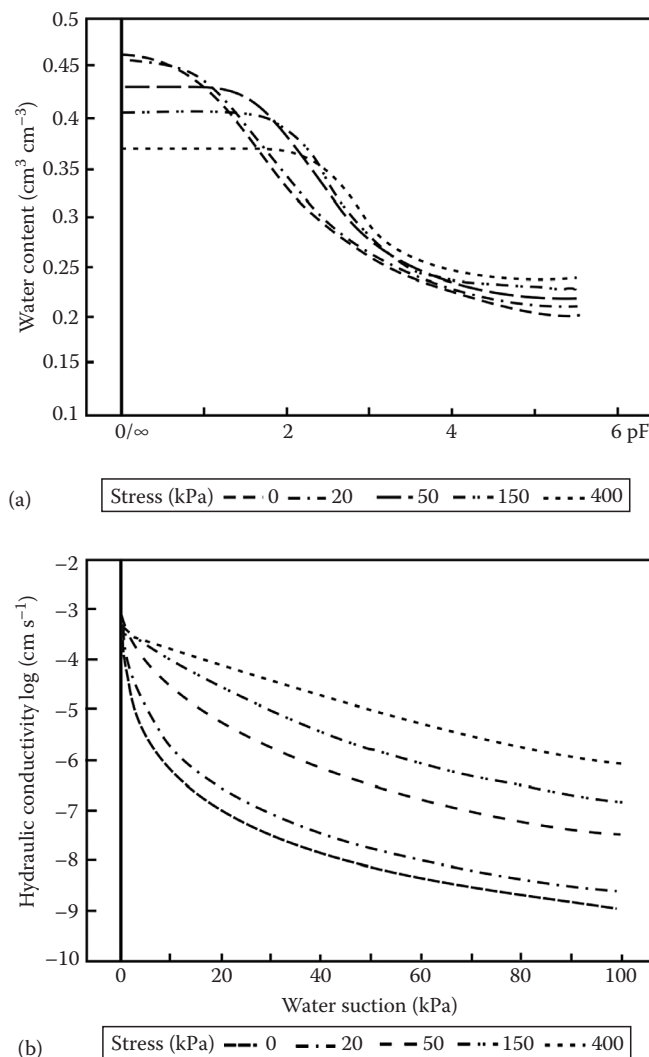
FIGURE 3.22 Change in water retention characteristic (a) and unsaturated hydraulic conductivity (b) as a function of applied mechanical stresses.
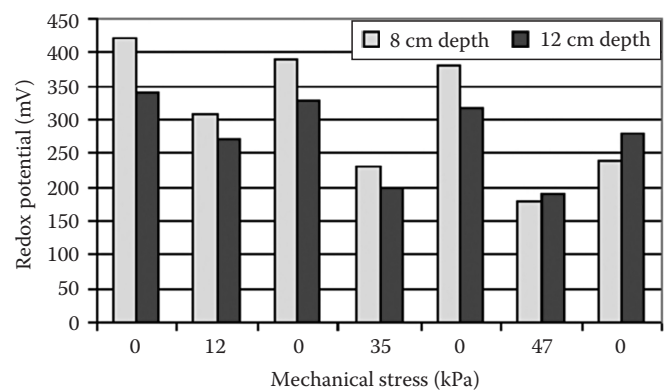
FIGURE 3.23 Changes in redox potential as a function of load in two soils depths. (Data from Horn, R. 1985a. Auswirkung mechanischer Belastungen auf die Redoxpotentiale von 3 Bodenmonolithen–ein Laborversuch. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 148:47–53.)

Identical diagrams are also prepared for gas diffusion (Glinski and Stepniewski, 1985), thermal properties, and also redox reactions where a more intense stress dependency with remaining smaller values is shown when the internal soil strength is exceeded (Figure 3.23). A direct link between the mechanical stress application and increased denitrification rates was presented by Flessa et al. (1998) and Liu et al. (2007).

However, not only these general changes, but also alterations in vector properties of, for example, the hydraulic conductivity (Doerner, 2005; Doerner and Horn 2009), reflect the strong effect of stress–strain processes on soil functions. Wheeling-induced formation of platy structures results in a pronounced horizontal anisotropy of the hydraulic conductivity and intensifies lateral water and solute movement, which enhances soil erosion and nutrient export. Particularly, shear-induced strain processes lead to changes in transport functions also with consequences for gas fluxes and compositions due to more pronounced flow path tortuosities and reduced pore connectivity, which can be observed down to the aggregate scale (Ball et al., 1999; Peth et al., 2008).

3.7 Modeling Soil Deformation and Coupled Processes

In unsaturated structured soils, hydraulic, thermal, and/or gas transport processes must be mutually linked in order to fully account for deformation-induced changes in soil functions. Especially, hydraulic and mechanical processes are tied up closely. This is evident for example in the action of hydraulic stresses during shrinking and swelling and their consequences for soil strength or in the role of pore water pressures on consolidation and compression (effective stress theory). On the other hand, soil deformation and reorientation of particles and aggregates influence pore-size distribution and pore-network geometries. Hence, hydraulic conductivity within the matrix and structural pores is changed, which finally affects, in turn, hydraulic stresses and subsequent deformation behavior.

The mathematical equations for water, gas, and heat transport in soils and the effects of tillage on changes in structure are available (Jury et al., 1991). However, often physical soil behavior is still investigated by treating either process in isolation rather than accounting for their interdependency (Kirby, 1994; Kirby et al., 1997; Keller et al., 2007b). Richards (1992) proposed a finite element model that was able to interactively couple load deformation and flow processes in unsaturated and swelling soils. The model was later further improved by Gräsle (1999) and tested by simulating standard mechanical laboratory tests. The modeling concepts Richards developed over decades are summarized in Richards and Peth (2009), including some model applications for civil engineering and soil science problems. A detailed conceptual framework for modeling coupled soil mechanical and hydraulic processes in soils is represented in Figure 3.24 (Gräsle et al., 1995). O'Sullivan and Simota (1995)
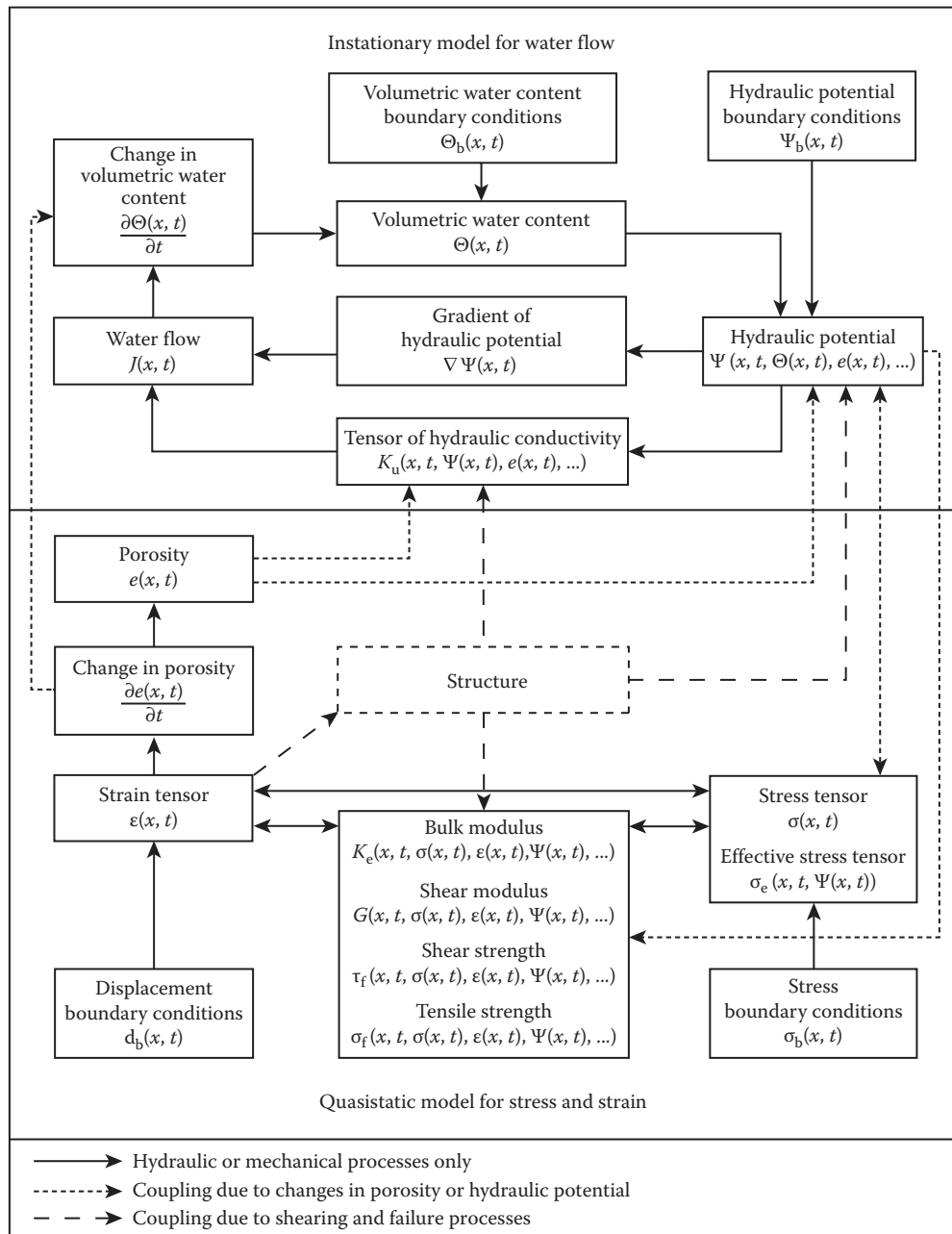
FIGURE 3.24 Schematic representation of a coupled hydraulic elastoplastic model to define and quantify coupled processes in soils. (From Gräsle, W. T. Baumgartl, R. Horn, and B.G. Richards. 1995. Interaction between soil mechanical properties of structured soils and hydraulic process—Theoretical fundamentals of a model. Balkema 2, p. 719–724. *In* E.E. Alonso and P. Delage (eds.) Proc. First Int. Conf. Unsaturated Soils/ UNSAT'95, September 6–8, Paris, France. Balkema, Rotterdam, the Netherlands.)

and Defossez and Richard (2002) have reviewed some soil compaction models and discussed problems when combining them with environmental impact and crop production models.

3.7.1 Deformation and Failure

This process describes the load/deformation/failure response of soil to a change in stress, strain, and/or displacement with all other factors held constant. It includes (1) nonlinear elastic behavior of the material as a function of stress and its history and matric potential (Richards, 1978); (2) changes in initial stress or strain as a result of swell/shrink behavior caused by changes in soil water suction, solute content, or temperature (Richards, 1986); (3) stress/strain path dependency or hysteretic behavior in load-deformation response (Richards, 1979); and (4) shear and tensile failure with dilatance or compression and with strain softening.

3.7.2 Coupling Soil Mechanical and Hydraulic Processes

During any kind of soil deformation, changes in hydraulic properties will occur, which is an important factor influencing the further deformation process as pointed out in the effective stress equation of *Terzaghi* Equation 3.1. In addition, changes in soil stress due to swelling or shrinkage will also cause changes in soil water potential. This effect of stress on water potential is sometimes referred to as the stress potential or the stress component of field measurements of soil water suction (Richards, 1986). Load-deformation analysis must therefore be coupled with changes in soil water potential, including subsequent water flow (Figure 3.24). The displacements can also be used to calculate the new geometry and material velocities for the analysis of water flow in soils undergoing strain. Such analysis can be also extended to 3D flow problems, which are of interest in the plow pan and in well-structured horizons below.

3.7.3 Example Application of the Finite Element Model

Modeling soil stiffness for various horizons with differing mechanical properties and the overall stress–strain relationships of a complete soil profile requires the derivation of stress-dependent moduli (bulk K and shear G modulus, see Section 3.2.3). The mechanical material properties defined by K and G may be determined by back-analyzing consolidation and shear test data (Richards et al., 1997). In this approach, the soil test is simulated with an FEM using an initial estimate of the material parameters, which are subsequently modified until a good agreement between modeled and measured test data is achieved. An example where a consolidation test has been simulated is shown in Figure 3.25.

Such stress-dependent mechanical soil properties may be used to calculate stress distribution under loads within a soil profile more accurately accounting for differences in soil stiffness and
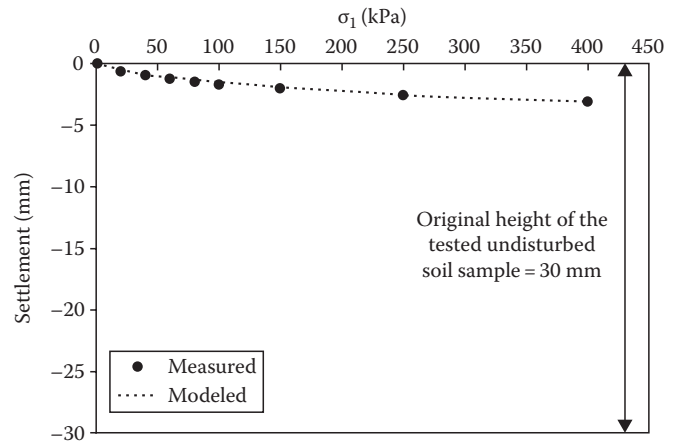
FIGURE 3.25   Measured and simulated settlement of a consolidation test (oedometer). The ordinate has been scaled to the original height of the tested undisturbed soil sample (30 mm). (From Richards, B.G., and S. Peth. 2009. Modelling soil physical behaviour with particular reference to soil science. Soil Till. Res. 102:216–224.)

strength with depth including the mechanical history of the soil (precompression stress). A comparison of measured stress (stress state transducer [SST]) data from a traffic experiment on a conservation tillage plot (shallow chiseling up to 8 cm, luvisol derived from glacial till) with stress distributions simulated by FEM using the mechanical parameters derived for three soil horizons of the corresponding soil profile is shown in Figure 3.26. Simulations including and excluding the plow pan layer, which was encountered in 20–40 cm soil depth, indicate that no good agreement between the measured and simulated stresses was obtained when the plow pan was not included in the FEM simulation. In contrast, if the corresponding mechanical data of the plow pan were included in the simulation, then a reasonably good agreement between the stress measurements and model prediction would be obtained (Figure 3.26). Similar examples of the simulation of stress distributions under tires using this FEM approach in conjunction with parameter estimations from mechanical laboratory testing are given in Peth et al. (2006) and Richards and Peth (2009).

3.8 Concluding Remarks

1. Mechanical soil properties are useful indicators for the development and physical quality of soil structure.
2. The determination of soil strength requires the measurement of volumetric stress and strain.
3. Structure development in arable and forest soils always results in increased soil strength. With increased aggregation, strength increases and the characteristics of the water retention curve become more pronounced.
4. In structured soils, hydraulic properties of inter- and intra-aggregate pores are different. Differences in hydraulic conductivity, plant available water, and air capacity become more pronounced as structure development increases.
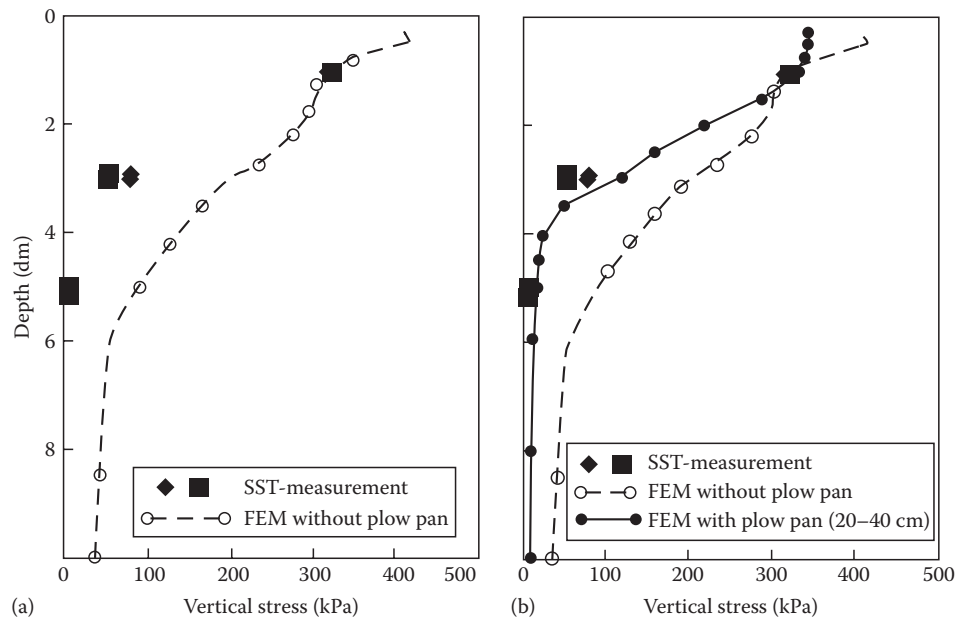
FIGURE 3.26 Determined and modeled vertical stress versus depth in a luvisol derived from glacial till under conservation tillage (type Horsch). The properties of a plow pan layer, which had been created until 1989 and which had a thickness of 20 cm (from 20 to 40 cm were not (a) and were (b) included in the model). (Data from Kühner, S. 1997. Simultane Messung von Spannungen und Bodenbewegungen bei statischen und dynamischen Belastungen zur Abschätzung der dadurch induzierten Bodenbeanspruchung. Ph.D. Thesis. Schriftenreihe des Institut für Pflanzenernährung und Bodenkunde, Germany, 39, 190 pp. ISSN: 0933-680X.)

In addition, root penetrability is reduced into aggregate interiors but may be enhanced in cracks of the interaggregate pore space.

5. The increase in strength with decreasing pore water pressure depends on the pattern of the water retention curve for single aggregates and bulk soil.

6. With increasing degree of aggregation, pore water pressure at the same water content in the bulk soil becomes more negative. Single aggregates are always stronger than the bulk soil or the homogenized material. This is related to smaller increases in pore water pressure during loading and higher aggregate bulk densities. Shear strength parameters (angle of internal friction and cohesion) are always higher for single aggregates than for bulk soil. However, as soon as applied stress exceeds the internal soil strength, single aggregate and aggregated bulk soils will become homogenized accompanied by a decline in strength.

7. Soil strength is strongly affected by spatial differences in pore water pressure. Hydraulic properties of single aggregates and their arrangement in the bulk soil are therefore most important for predicting soil strength and stress distributions in soils. Stress attenuation, stress distribution, and soil deformation due to loading under running wheels or belts can vary to a great extent depending on in situ soil conditions, machinery parameters, and frequency of loading. Soil deformation should therefore be considered as a dynamic rather than a static process.

8. Structure development always involves reduced accessibility to exchange sites for soil solutions, retarded exchange processes, and stronger chemical disequilibrium. Only at the lowest entropy (at an increased accessibility), structured soils become macroscopically homogeneous. With reduced pore water potential (reduced flux rate), the exchange process becomes more complete. At higher saturations, anaerobic soil volumes in aggregates become more extensive for longer periods. Thus, soil physicochemical properties vary intensely throughout the year.

9. The coupled analysis of multifunctional and interactive processes in unsaturated structured soils is an important topic for future soil research but very complex. Long-term soil stability involves materials, which are unsaturated, nonlinear hysteretic, and the soil composition changes with time. During tillage operations, many physical and mechanical properties are altered requiring a more precise definition of the actual conditions and their expected temporal changes.

10. Measurements and modeling of coupled ion and water fluxes through unsaturated structured soils are most important for future work because aggregation affects ion diffusion, dispersion, and mass flow. Changes in pore systems and functions are also affected by mechanical stresses. Thus, model coupling is required and could enhance the estimate of time-dependent effects on nutrient, heavy metal, or pesticide transport through structured soils also including the effect of drying or wetting or deformation (compaction) processes.

References

Alakukku, L., P. Weisskopf, W.C.T. Chamen, F.G.J. Tijink, J.P. van der Linden, S. Pires, C. Sommer, and G. Spoor. 2003. Prevention strategies for field traffic-induced subsoil compaction: A review. Part 1. Machine/soil interactions. Soil Till. Res. 73:145–160.

Arvidsson, J., and T. Keller. 2007. Soil stress as affected by wheel load and tyre inflation pressure. Soil Till. Res. 96:284–291.

Atterberg, A. 1911. Die plastizität der tone. Int. Mitt. Bodenkd. 1:10–43.

Bailey, A.C., C.E. Johnson, and R.L. Schafer. 1986. A model for agricultural soil compaction. J. Agr. Eng. Res. 33:257–262.

Bailey, A.C., R.L. Raper, C.E. Johnson, and E.C. Burt. 1992. An integrated approach to soil compaction modelling. Proc. Int. Agric. Eng. Conf. June 1–4, Uppsala, Sweden.

Ball, B.C., D.J. Campbell, and E.A. Hunter. 2000. Soil compactibility in relation to physical and organic properties at 156 sites in UK. Soil Till. Res. 57:83–91.

Ball, B.C., A. Scott, and J.P. Parker. 1999. Field N_2O, CO_2 and CH_4 fluxes in relation to tillage, compaction and soil quality in Scotland. Soil Till. Res. 53:29–39.

Barré, P., and P.D. Hallett. 2009. Rheological stabilization of wet soils by model root and fungal exudates depends on clay mineralogy. Eur. J. Soil Sci. 60:525–538.

Baumgartl, T. 1991. Spannungsverteilung in unterschiedlich texturierten Böden und ihre Bedeutung für die Bodenstabilität. Ph.D. Thesis. Schriftenreihe Inst. Pflanzenernaehrung und Bodenkunde, Germany, No. 12, 150 pp. ISSN: 0933-680X.

Baumgartl, T. 2003. Koppelung von mechanischen und hydraulischen Bodenzustandsfunktionen zur Bestimmung und Modellierung von Zugspannungen und Volumenänderungen in porösen Medien. Habilitation Thesis. Schriftenreihe Inst. Pflanzenernaehrung und Bodenkunde, Germany, No. 62, 133 pp. ISSN: 0933-680X.

Baumgartl, T., and R. Horn. 1993. The determination of aggregate stability: A comparison of methods. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 156:385–391.

BBodSchG. 1998. Gesetz zum Schutz vor schädlichen Bodenveränderungen und zur Sanierung von Altlasten. Bundesgesetzblatt I, 502ff.

Berli, M., J.M. Kirby, S.M. Springman, and R. Schulin. 2003. Modelling compaction of agricultural subsoils by tracked heavy construction machinery under various moisture conditions in Switzerland. Soil Till. Res. 73:57–66.

Birkas, M., M. Jolankai, C. Gyuricza, and A. Percze. 2004. Tillage effects on compaction, earthworms and other soil quality indicators in Hungary. Soil Till. Res. 78:185–196.

Bishop, A.W. 1959. The principle of effective stress. Teknisk Ukeblad. 106:859–863.

Blackwell, J., R. Horn, N.S. Jayawardane, R. White, and P.S. Blackwell. 1989. Vertical stress distribution under tractor wheeling in a partially deep loosened typic paleustalf. Soil Till. Res. 13:1–12.

Boone, F.R., and B.W. Veen. 1994. Mechanisms of crop responses to soil compaction, p. 237–264. *In* D. Soane and C. van Ouwerkerk (eds.) Soil compaction in crop production. Elsevier, Amsterdam, the Netherlands.

Boussinesq, I. 1885. Applications des potentiels à l'étude de l'équilibre et du mouvement des solides élastiques. Gauthier-Villars, Paris, France.

Braudeau, E., J.P. Frangi, and R.H. Mohtar. 2004. Characterizing nonrigid aggregated soil-water medium using its shrinkage curve. Soil Sci. Soc. Am. J. 68:359–370.

Burt, E.C., R.K. Wood, and A.C. Bailey. 1989. Effects of dynamic load on normal soil-tire interface stresses. Trans. ASAE 32:1843–1846.

Casagrande, A. 1932. Research on Atterberg's limits of soils. Public Roads 13:121–130.

Casagrande, A. 1936. The determination of pre-consolidation load and its practical significance, p. 60–64. Proc. Int. Conf. Soil Mech. Found. Eng. Harvard University, June 22–26, Cambridge, MA.

Chamen, T., L. Alakukku, S. Pires, C. Sommer, G. Spoor, F. Tijink, and P. Weisskopf. 2003. Prevention strategies for field traffic-induced subsoil compaction: A review. Part 2. Equipment and field practices. Soil Till. Res. 73:161–174.

Cui, K., P. Defossez, and G. Richard. 2007. A new approach for modelling vertical stress distribution at the soil/tyre interface to predict the compaction of cultivated soils by using the PLAXIS code. Soil Till. Res. 95:277–287.

da Silva, V.R., J.M. Reichert, and D.J. Reinert. 2006. Soil temperature variation in three different systems of soil management in blackbeans crop. Rev. Bras. Cienc. Solo 30:391–399.

Defossez, P., and G. Richard. 2002. Models of soil compaction due to traffic and their evaluation. Soil Till. Res. 67:41–64.

Dexter, A.R., and B. Kroesbergen. 1985. Methodology for determination of tensile strength of soil aggregates. J. Agr. Eng. Res. 31:139–147.

Doerner, J. 2005. Anisotropie von Bodenstrukturen und Porenfunktionen in Böden und deren Auswirkungen auf Transportprozesse im gesättigten und ungesättigten Zustand. Ph.D. Thesis. Schriftenreihe des Institut für Pflanzenernährung und Bodenkunde, Germany, No. 68, 145 pp. ISSN: 0933-680X.

Doerner, J., and R. Horn. 2009. Direction-dependent behaviour of hydraulic and mechanical properties in structured soils under conventional and conservation tillage. Soil Till. Res. 102:225–232.

DVWK. 1995. Soil strength of arable soils. Part I (in German). Gefügestabilität ackerbaulich genutzter Mineralböden: Teil I: Mechanische Belastbarkeit. Vol. 234. Wirtschafts- und Verlagsgesellschaft Gas und Wasser mbH. ISBN: 3-89554-025-0.

Ellies, A., R. Grez, and Y. Ramirez. 1995. Cambios en las propiedades humectantes en suelos sometidos a diferentes manejos. Turrialba 45:42–48.

Emerson, W.W., R.D. Bond, and A.R. Dexter. 1978. Modification of soil structure. Wiley & Sons, Chichester, West Sussex, U.K.

Fazekas, O. 2005. Bedeutung von Bodenstruktur und Wasserspannung als stabilisierende Kenngrößen gegen intensive mechanische Belastungen in einer Parabraunerde aus Löss unter Pflug- und Mulchsaat. Ph.D. Thesis. Schriftenreihe des Institut für Pflanzenernährung und Bodenkunde, Germany, No. 67, 170 pp. ISSN: 0933-680X.

Fazekas, O., and R. Horn. 2005. Interaction between mechanically and hydraulically affected soil strength depending on time of loading. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 168:60–67.

Flessa, H., U. Wild, M. Klemisch, and J. Pfadenhauer. 1998. Nitrous oxide and methane fluxes from organic soils under agriculture. Eur. J. Soil Sci. 49:327–335.

Fredlund, D.G., and H. Rahardjo. 1993. Soil mechanics for unsaturated soils. John Wiley & Sons, New York.

Fröhlich, O.K. 1934. Druckverteilung im Baugrund. Springer, Wien, New York.

Ghezzehei, T.A., and D. Or. 2001. Rheological properties of wet soils and clays under steady and oscillatory stresses. Soil Sci. Soc. Am. J. 65:624–637.

Glinski, J., and W. Stepniewski. 1985. Soil aeration and its role for plants. CRC Press, Boca Raton, FL.

Gräsle, W. 1999. Numerische Simulation mechanischer, hydraulischer und gekoppelter Prozesse in Böden unter Verwendung der Finiten Elemente Methode. Ph.D. Thesis. Schriftenreihe des Institut für Pflanzenernährung und Bodenkunde, Germany, No. 48, 400 pp. ISSN: 0933-680X.

Gräsle, W. T. Baumgartl, R. Horn, and B.G. Richards. 1995. Interaction between soil mechanical properties of structured soils and hydraulic process—Theoretical fundamentals of a model. Balkema 2, p. 719–724. *In* E.E. Alonso and P. Delage (eds.) Proc. First Int. Conf. Unsaturated Soils/UNSAT'95, September 6–8, Paris, France. Balkema, Rotterdam, the Netherlands.

Groenevelt, P.H., and C.D. Grant. 2001. Re-evaluation of the structural properties of some British swelling soils. Eur. J. Soil Sci. 52:469–477.

Grozic, J.L.H., T. Lunne, and S. Pande. 2003. An oedometer test study on the preconsolidation stress of glaciomarine clays. Can. Geotech. J. 40:857–872.

Gudehus, G. 1981. Bodenmechanik. Enke-Verlag, Stuttgart, Germany.

Hadas, A. 1987. Dependence of "True" surface energy of soils on air entry pore size and chemical constituents. Soil Sci. Soc. Am. J. 51:187–191.

Hallett, P.D., and T.A. Newson. 2001. A simple fracture mechanics approach for assessing ductile crack growth in soil. Soil Sci. Soc. Am. J. 65:1083–1088.

Hallett, P.D., and T.A. Newson. 2005. Describing soil crack formation using elastic–plastic fracture mechanics. Eur. J. Soil Sci. 56:31–38.

Hartge, K.H., and R. Horn. 1984. Untersuchungen zur Gültigkeit des Hook'schen Gesetzes bei der Setzung von Böden bei wiederholter Belastung. Z. Acker Pflanzenbau 153:200–207.

Hartge, K.H., and R. Horn. 1992. Die physikalische Untersuchung von Böden. Ferdinand Enke, Stuttgart, Germany.

Hartge, K.H., and R. Horn. 1999. Einführung in die Bodenphysik. Ferdinand Enke, Stuttgart, Germany.

Hempfling, R., H.R. Schulten, and R. Horn. 1990. Relevance of humus composition to the physical mechanical stability of agricultural soils—A study by direct pyrolysis mass-spectrometry. J. Anal. Appl. Pyrolysis 17:275–281.

Hendrickx, J.M.H., and M. Flury. 2001. Uniform and preferential flow, mechanisms in the vadose zone, p. 149–187. *In* National Research Council (ed.) Conceptual models of flow and transport in the fractured vadose zone. National Academy Press, Washington, DC.

Hesterberg, D., and A.L. Page. 1993. Rheology of sodium and potassium illite suspensions in relation to colloidal stability. Soil Sci. Soc. Am. J. 57:697–704.

Horn, R. 1976. Mechanical strength changes due to aggregate formation of a mesozoic clay. PhD Thesis. TU Hannover, Germany.

Horn, R. 1981. Die Bedeutung der Aggregierung von Böden für die mechanische Belastbarkeit in dem für Tritt relevanten Auflastbereich und deren Auswirkungen auf physikalische Bodenkenngrößen. Habilitation Thesis. Schriftenreihe FB 14, TU Berlin, 200 pp. ISBN: 379830792X.

Horn, R. 1985a. Auswirkung mechanischer Belastungen auf die Redoxpotentiale von 3 Bodenmonolithen-ein Laborversuch. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 148:47–53.

Horn, R. 1985b. Die Bedeutung der Trittverdichtung durch Tiere auf physikalische Eigenschaften Alpiner Böden. Z. Kulturtech. Flurber. 26:42–51.

Horn, R. 1986. Auswirkung unterschiedlicher Bodenbearbeitung auf die mechanische Belastbarkeit von Ackerböden. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 149:9–18.

Horn, R. 1988. Compressibility of arable land. Catena Suppl. 11:53–71.

Horn, R. 1994a. The effect of aggregation of soils on water, gas, and heat transport, p. 335–361. *In* E.D. Schulze (ed.) Flux control in biological systems: From enzymes to populations and ecosystems. Academic Press, March, London, U.K.

Horn, R. 1994b. Stress transmission and recompaction in tilled and segmently disturbed soils under trafficking, p. 53–87. *In* N.S. Jayawardane and E. Stewart (eds.) Subsoil management techniques. Lewis Publishers, Boca Raton, FL.

Horn, R. 2000. Subsoil compaction processes—State of knowledge, p. 9–20. *In* T.W. Riley and J.M.A. Desbiolles (eds.) Proc. 4th Int. Conf. Soil Dyn. Adelaide, Australia.

Horn, R. 2003. Stress–strain effects in structured unsaturated soils on coupled mechanical and hydraulic processes. Geoderma 116:77–88.

Horn, R., T. Baumgartl, R. Kayser, and S. Baasch. 1995. Effect of aggregate strength on changes in strength and stress distribution in structured bulk soils, p. 31–52. *In* K.H. Hartge and R. Stewart (eds.) Soil structure: Its development and function. CRC Press, Boca Raton, FL.

Horn, R., M. Burger, M. Lebert, and G. Badewitz. 1987. Druckfortpflanzung in Böden unter langsam fahrenden Traktoren. Z. Kulturtech. Flurber. 28:94–102.

Horn, R., and A.R. Dexter. 1989. Dynamics of soil aggregation in an irrigated desert loess. Soil Till. Res. 13:253–266.

Horn, R., and H. Fleige. 2003. A method for assessing the impact of load on mechanical stability and on physical properties of soils. Soil Till. Res. 73:89–99.

Horn, R., and H. Fleige. 2009. Risk assessment of subsoil compaction for arable soils in Northwest Germany at farm scale. Soil Till. Res. 102:201–209.

Horn, R., H. Fleige, S. Peth, and X. Peng (eds.). 2006. Soil management for sustainability. Advances in geoecology. Vol. 38. Catena, Reiskirchen, Germany.

Horn, R., H. Fleige, F.H. Richter, E.A. Czyz, A. Dexter, E. Diaz-Pereira, E. Dumitru et al. 2005. SIDASS project: Part 5: Prediction of mechanical strength of arable soils and its effects on physical properties at various map scales. Soil Till. Res. 82:47–56.

Horn, R., K.H. Hartge, J. Bachmann, and M.B. Kirkham. 2007a. Mechanical stresses in soils assessed from bulk-density and penetration-resistance data sets. Soil Sci. Soc. Am. J. 71:1455–1459.

Horn, R., C. Johnson, H. Semmel, R. Schafer, and M. Lebert. 1992. Räumliche Spannungsmessungen mit dem Stress State Transducer (SST) in ungesättigten aggregierten Böden—Theoretische Betrachtungen und erste Ergebnisse. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 155:269–274.

Horn, R., and M. Kutilek. 2009. The intensity-capacity concept—How far is it possible to predict intensity values with capacity parameters. Soil Till. Res. 103:1–3.

Horn, R., M. Lebert, and N. Burger, 1989. Prediction of the mechanical strength of arable soils based on laboratory and in situ measurements (in German, with English summary and captures). Vorhersage der mechanischen Belastbarkeit von Böden als Pflanzenstandort auf der Grundlage von Labor- and in situ Messungen. Final Report Bayer. StMLU Bewilligungs-Nr. 6333-972-57238, 178 S.

Horn, R., J. Vossbrink, S. Peth, and S. Becker. 2007b. Impact of modern forest vehicles on soil physical properties. Forest Ecol. Manag. 248:56–63.

Janssen, I. 2008. Landnutzungsabhängige Dynamik hydraulischer und mechanischer Bodenstrukturfunktionen in Nassreisböden. Ph.D. Thesis. Schriftenreihe Institut für Pflanzenernährung und Bodenkunde, Germany, No. 76, 168pp. ISSN: 0933-680X.

Jayawardane, N.S., and E. Stewart (eds.). 1994. Subsoil management techniques. Lewis Publishers Inc., Boca Raton, FL.

Johnson, C.E., and A.C. Bailey. 1990. A shearing strain model for cylindrical stress states. Proc. ASAE 90:1085.

Junge, T., W. Gräsle, G. Bensel, and R. Horn. 2000. Einfluss des Porenwasserdruckes auf die Zugfestigkeit von Bodenproben. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 163:21–26.

Jury, W.A., W.R. Gardner, and W.H. Gardner. 1991. Soil physics. John Wiley & Sons, New York.

Karlen, D.L. 2004. Soil quality as an indicator of sustainable tillage practices. Soil Till. Res. 78:129–130.

Kay, B.D. 1990. Rates of change of soil structure under different cropping systems. Adv. Soil Sci. 12:1–41.

Keller, T. 2004. Soil compaction and soil tillage—Studies in agricultural soil mechanics. SLU Service/Repro, Uppsala, Sweden.

Keller, T., J. Arvidsson, and A.R. Dexter. 2007a. Soil structures produced by tillage as affected by soil water content and the physical quality of soil. Soil Till. Res. 92:45–52.

Keller, T., P. Defossez, P. Weisskopf, J. Arvidsson, and G. Richard. 2007b. SoilFlex: A model for prediction of soil stresses and soil compaction due to agricultural field traffic including a synthesis of analytical approaches. Soil Till. Res. 93:391–411.

Keren, R. 1988. Rheology of aqueous suspensions of sodium/calcium montmorillonite. Soil Sci. Soc. Am. J. 52:924–928.

Kézdi, Á. 1974. Handbook of soil mechanics: Soil physics. Elsevier, Amsterdam, the Netherlands.

Kirby, J.M. 1994. Simulating soil deformation using a critical-state model: I. Laboratory tests. Eur. J. Soil Sci. 45:239–248.

Kirby, J.M., B.G. Blunden, and C.R. Trein. 1997. Simulating soil deformation using a critical-state model: II. Soil compaction beneath tyres and tracks. Eur. J. Soil Sci. 48:59–70.

Koolen, A.J. 1994. Mechanics of soil compaction, p. 23–44. *In* D. Soane and C. van Ouwerkerk (eds.) Soil compaction in crop production. Elsevier, Amsterdam, the Netherlands.

Koolen, A.J., and H. Kuipers. 1983. Agricultural soil mechanics. Springer, Berlin, Germany.

Kretschmer, H. 1997. Körnung und Konsistenz, p. 1–46. *In* H.P. Blume, P. Felix-Henningsen, W. Fischer, H.G. Frede, R. Horn, and K. Stahr (eds.) Handbuch der Bodenkunde. Ecomed, Landsberg, Germany.

Krümmelbein, J., S. Peth, and R. Horn. 2008. Determination of pre-compression stress of a variously grazed steppe soil under static and cyclic loading. Soil Till. Res. 99:139–148.

Kühner, S. 1997. Simultane Messung von Spannungen und Bodenbewegungen bei statischen und dynamischen Belastungen zur Abschätzung der dadurch induzierten Bodenbeanspruchung. Ph.D. Thesis. Schriftenreihe des Institut für Pflanzenernährung und Bodenkunde, Germany, 39, 190 pp. ISSN: 0933-680X.

Lal, R. 2005. Soil erosion and carbon dynamics. Soil Till. Res. 81:137–142.

Langmaack, M., S. Schrader, U. Rapp-Bernhardt, and K. Kotzke. 2002. Soil structure rehabilitation of arable soil degraded by compaction. Geoderma 105:141–152.

Larink, O., D. Werner, D. Langmaack, and S. Schrader, 2001. Regeneration of compacted soil aggregates by earthworm activity. Biol. Fertil. Soils 33:395–401.

Lipiec, J., and R. Hatano. 2003. Quantification of compaction effects on soil physical properties and crop growth. Geoderma 116:107–136.

Lipiec, J., R. Walczak, and G. Józefacuik (eds.). 2004. Plant growth in relation to soil physical conditions. Institute of Agrophysics, Lublin, Poland.

Liu, X.J., A.R. Mosier, A.D. Halvorson, C.A. Reule, and F.S. Zhang. 2007. Dinitrogen and N_2O emissions in arable soils: Effect of tillage, N source and soil moisture. Soil Biol. Biochem. 39:2362–2370.

Markgraf, W., and R. Horn. 2007. Scanning electron microscopy-energy dispersive scan analyses and rheological investigations of South-Brazilian soils. Soil Sci. Soc. Am. J. 71:851–859.

Markgraf, W., and R. Horn. 2009. Rheological investigations in soil micro mechanics: Measuring stiffness degradation and structural stability on a particle scale. *In* L.P. Gragg and J.M. Cassell (eds.) Progress in management engineering. Nova Science Publishers, Hauppauge, New York, pp. 237–279. ISBN: 978-1-60741-310-3.

Markgraf, W., R. Horn, and S. Peth. 2006. An approach to rheometry in soil mechanics—Structural changes in bentonite, clayey and silty soils. Soil Till. Res. 91:1–14.

McCarthy, D.F. 2007. Essentials of soil mechanics and foundations. Prentice Hall, New York.

Mezger, T. 2006. The rheology handbook. Vincentz, Hannover, Germany.

Mitchell, J.K. 1993. Fundamentals of soil behaviour. Wiley & Sons, New York.

Nichols, T.A., A.C. Bailey, C.E. Johnson, and D. Grisso. 1987. A stress state transducer for soil. Trans. ASAE 30:1237–1241.

Oldeman, L.R. 1992. Global extent of soil degradation. Proc. Symp. Soil Resil. Sustainable Land Use, September 28–October 2, 1992. Budapest, Hungary. CAB International, New York.

Olfe, G. 1995. Zur Bodenbelastung durch den Schlepper- und Maschineneinsatz in der pflanzlichen Produktion. KTBL Schrift 362:12–28.

O'Sullivan, M.F., and C. Simota. 1995. Modelling the environmental impacts of soil compaction: A review. Soil Till. Res. 35:69–84.

Pagliai, M., and R. Jones (eds.). 2002. Sustainable land management-environmental protection. A soil physical approach. Advances in GeoEcology. Vol. 35. Catena, Reiskirchen, Germany.

Parry, R.H.G. 2004. Mohr circles, stress paths and geotechnics. Spon Press, London, U.K.

Peng, X., and R. Horn. 2005. Modeling soil shrinkage curve across a wide range of soil types. Soil Sci. Soc. Am. J. 69:584–592.

Peng, X., and R. Horn. 2008. Time-dependent, anisotropic pore structure and soil strength in a 10-year period after intensive tractor wheeling under conservation and conventional tillage. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 171:936–944.

Peth, S. 2004. Bodenphysikalische Untersuchungen zur Trittbelastung von Böden bei der Rentierweidewirtschaft an borealen Wald- und subarktisch-alpinen Tundrenstandorten–Auswirkungen auf thermische, hydraulische und mechanische Bodeneigenschaften. Ph.D. Thesis. Schriftenreihe Institut für Pflanzenernährung und Bodenkunde, Germany, No. 64, 130 pp. ISSN: 0933-680X.

Peth, S., and R. Horn. 2004. Zur Abschätzung von Bodenspannungen unter landwirtschaftlichen Nutzfahrzeugen. Landtechnik 59:268–269.

Peth, S., and R. Horn. 2006a. Consequences of grazing on soil physical and mechanical properties in forest and tundra environments, p. 217–243. *In* B.C. Forbes, M. Bölter, L. Müller-Wille, J. Hukkinen, F. Müller, N. Gunslay, and Y. Konstatinov (eds.) Reindeer management in Northernmost Europe. Vol. 184. Springer, Berlin, Germany.

Peth, S., and R. Horn. 2006b. The mechanical behavior of structured and homogenized soil under repeated loading. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 169:401–410.

Peth, S., R. Horn, F. Beckmann, T. Donath, J. Fischer, and A.J.M. Smucker. 2008. Three-dimensional quantification of intra-aggregate pore-space features using synchrotron-radiation-based microtomography. Soil Sci. Soc. Am. J. 72:897–907.

Peth, S., R. Horn, O. Fazekas, and B.G. Richards. 2006. Heavy soil loading and its consequences for soil structure, strength and deformation of arable soils. J. Plant Nutr. Soil Sci./Z. Pflanzenernahr. Bodenkd. 169:775–783.

Peth, S., J. Rostek, A. Zink, A. Mordhorst, and R. Horn. 2010. Soil testing of dynamic deformation processes of arable soils. Soil Till. Res. 106:317–328.

Pietola, L., R. Horn, and M. Yli-Halla. 2005. Effects of trampling by cattle on the hydraulic and mechanical properties of soil. Soil Till. Res. 82:99–108.

Proctor, R.R. 1933. Fundamental principles of soil compaction. Description of field and laboratory methods. Eng. News Rec. 111:286–289.

Raine, S.R., and H.B. So. 1994. Ultrasonic dispersion of soil in water—The effect of suspension properties on energy-dissipation and soil dispersion. Aust. J. Soil Res. 32:1157–1174.

Reich, J., H. Unger, H. Streitenberger, C. Mäusezahl, C. Nussbaum, and P. Steinert. 1985. Verfahren und Vorrichtung zur Verbesserung verdichteter Unterböden. Agrartechnik Berlin 41:57–62.

Reicosky, D.C., M.J. Lindstrom, T.E. Schumacher, D.E. Lobb, and D.D. Malo. 2005. Tillage-induced CO_2 loss across an eroded landscape. Soil Till. Res. 81:183–194.

Richards, B.G. 1978. Application of an experimentally based non-linear constitutive model to soils in the laboratory and field tests. Aust. Geomech. J. G8:20–30.

Richards, B.G. 1979. The method of analysis of the effects of volume change in unsaturated expansive clays on engineering structures. Aust. Geomech. J. G9:27–41.

Richards, B.G. 1986. The role of lateral stresses on soil water relations in swelling soils. Aust. J. Soil Res. 24:457–467.

Richards, B.G. 1992. Modelling interactive load-deformation and flow processes in soils, including unsaturated and swelling soils, p. 18–37. Proc. 6th Australian-New Zealand Conf. Geomech., Feb. 3–14, Christchurch, New Zealand.

Richards, B.G., T. Baumgartl, R. Horn, and W. Gräsle. 1997. Modelling the effects of repeated wheel loads on soil profiles. Int. Agrophys. 11:177–187.

Richards, B.G., and S. Peth. 2009. Modelling soil physical behaviour with particular reference to soil science. Soil Till. Res. 102:216–224.

Semmel, H. 1993. Auswirkungen kontrollierter Bodenbelastungen auf das Druckfortpflanzungsverhalten und physikalisch-mechanische Kenngrößen von Ackerböden. PhD Thesis. Institut für Pflanzenernährung und Bodenkunde, Germany.

Semmel, H., R. Horn, U. Hell, A.R. Dexter, and E.D. Schulze. 1990. The dynamics of soil aggregate formation and the effect on soil physical properties. Soil Technol. 3:113–129.

Skidmore, E.L., and D.H. Powers. 1982. Dry soil-aggregate stability: Energy-based index. Soil Sci. Soc. Am. J. 46:1274–1279.

Soane, B.D., and C. van Ouwerkerk. 1994. Soil compaction in crop production. Elsevier, Amsterdam, the Netherlands.

Soehne, W. 1958. Fundamentals of pressure distribution and soil compaction under tractor tires. Agr. Eng. 39:276–290.

Stepniewski, W., J. Glinski, and B.C. Ball. 1994. Effects of soil compaction on soil aeration properties, p. 167–190. *In* D. Soane and C. van Ouwerkerk (eds.) Soil compaction in crop production. Elsevier, Amsterdam, the Netherlands.

Terzaghi, K. 1936. The shearing resistance of saturated soils, p. 54–56. Proc. 1st Int. Conf. Soil Mech. Found. Eng. Harvard University, June 22–26, Cambridge, Massachusetts.

Toth, G., L. Montanarella, and E. Rusco. 2008. Threats to soil quality in Europe. JRC Scientific and Technical Reports EUR 23438 EN. European Commission Joint Research Centre Institute for Environment and Sustainability, Ispra, Italy.

van den Akker, J., J. Arvidsson, and R. Horn (eds.). 1998. Experiences with the impact and prevention in the European Community. Report 168. ISSN: 0927-4499.

van der Ploeg, R.R., W. Ehlers, and R. Horn. 2006. Schwerlast auf dem Acker. Spektrum der Wissenschaft August, 80–88. ISSN: 0170-2971.

Van Oost, K., G. Govers, S. de Alba, and T.A. Quine. 2006. Tillage erosion: A review of controlling factors and implications for soil quality. Prog. Phys. Geog. 30:443–466.

Vossbrink, J., and R. Horn. 2004. Modern forestry vehicles and their impact on soil physical properties. Eur. J. Forest. Res. 123:259–267.

Warrick, A. (ed.). 2002. Soil physics companion. CRC Press LLC, Boca Raton, FL.

Watts, C.W., L.J. Clark, W.C.T. Chamen, and A.P. Whitmore. 2005. Adverse effects of simulated harvesting of short-rotation willow and poplar coppice on vertical pressures and rut depths. Soil Till. Res. 84:192–199.

Wiermann, C., D. Werner, R. Horn, J. Rostek, and B. Werner. 2000. Stress/strain processes in a structured unsaturated silty loam Luvisol under different tillage treatments in Germany. Soil Till. Res. 53:117–128.

Wulfsohn, D., B.A. Adams, and D.G. Fredlund. 1998. Triaxial testing of unsaturated agricultural soils. J. Agr. Eng. Res. 69:317–330.

Yasutomi, R., and S. Sudo. 1967. A method of measuring some physical properties of soil with a forced oscillation viscometer. Soil Sci. 104:336–341.

Yong, R.N., and B.P. Warkentin. 1966. Introduction to soil behaviour. Macmillan Co., New York.

Zapf, R. 1997. Mechanische Bodenbelastung durch die landwirtschaftliche Pflanzenproduktion in Bayern. BLBP, Freising, Germany.

Zink, A. 2009. Bodenstabilität und Auswirkungen dynamischer Lasteinträge auf physikalische Eigenschaften von Ackerböden unter konservierender und konventioneller Bodenbearbeitung. Ph.D. Thesis. Schriftenreihe Institut für Pflanzenernährung und Bodenkunde, Germany, No. 84, 180 pp. ISSN: 0933-680X.

4

Soil Water Content and Water Potential Relationships

Dani Or
Eidgenössische Technische Hochschule Zürich

Jon M. Wraith
University of New Hampshire

David A. Robinson
Centre for Ecology & Hydrology

Scott B. Jones
Utah State University

4.1 Introduction

Water in soil occupies pore spaces that arise from the physical arrangement of the particulate solid phase, competitively and often concurrently with the soil gas phase (Chapters 1, 2, and 8). While hidden from casual view, substantial volumes of water are commonly stored in soils. For example, one hectare of medium-textured soil 1 m deep and having field capacity water content of 20% by volume may store sufficient water to fill 4000 200 L barrels. This reservoir serves as a resource pool for plant growth and a substantial buffer in areas having scattered or sporadic precipitation. Other soil organisms, many of which are beneficial, especially to biogeochemical cycling, also rely heavily on the water-holding characteristics of soils for their existence. On the other hand, soil water is a highly dynamic entity, exhibiting substantial variation in both time and space (Western et al., 2002). This is particularly true near the soil surface and in the presence of active plant roots. Changes in soil water content and its energy status affect many soil mechanical properties including strength, compactibility, and penetrability, and may cause changes in the bulk density of swelling soils. The liquid-phase characteristics affect the soil gaseous phase and the rates of exchange between these phases, as well as other important soil properties such as the hydraulic conductivity that governs the rate of water and soluble chemical flow.

The purpose of this chapter is to introduce the basic concepts related to the amount and energy state of water in soil. These concepts are prerequisite to quantify and manage soil water storage, to obtain predictions concerning rates and directions of water flow and solute transport, to utilize soils as building or foundation materials, and for many other purposes. The term soil water is used here to represent the soil liquid phase, which is typically a water solution containing dissolved salts, organic substances, and gases.

The water status in soils is defined by (1) the amount of water in the soil or soil water content (θ), and (2) the force by which water is held in the soil matrix, soil energy content or soil water potential (ψ). These soil water attributes are related to each other through a function known as the soil water characteristic (SWC).

4.2 Soil Water Content

Many agronomic, hydrologic, and geotechnical practices require knowledge of the amount of water contained in a particular soil volume (Robinson et al., 2008b; Vereecken et al., 2008). Some of the most common methods used to characterize and determine soil water content, on a mass or volume basis, in the laboratory or in situ will be described.

4.2.1 Definitions

4.2.1.1 Soil Water Content on Mass Basis (Gravimetric)

Mass or gravimetric soil water content is expressed relative to the mass of oven dry soil according to

$$\theta_m = \frac{\text{water mass}}{\text{dry soil mass}} = \frac{(\text{wet soil mass}) - (\text{dry soil mass})}{\text{dry soil mass}}, \qquad (4.1)$$

and has units of kg kg^{-1} or other consistent mass units.

4.2.1.2 Soil Water Content on Volume Basis

It is often desirable to express water content on a volume basis. The volumetric water content (θ_v) is defined as the volume of water per bulk volume of soil:

$$\theta_v = \frac{\text{volume of water}}{\text{bulk volume of soil}} = \frac{(\text{mass of water/density of water})}{\text{sample volume}}. \qquad (4.2)$$

It also represents the depth ratio of soil water (i.e., the depth of water per unit depth of soil). The conversion between gravimetric and volumetric water contents requires knowledge of the soil dry bulk density (ρ_b), which is the ratio of oven dry soil mass to its original volume, and the density of water. The conversion formula is given by the following:

$$\theta_v = \theta_m \frac{\rho_b}{\rho_w}, \qquad (4.3)$$

where ρ_w is the density of water (1000 kg m^{-3} at 20°C). Alternatively, a soil sample of known original volume may be processed as in the gravimetric method, and θ_v determined from Equation 4.2.

4.2.1.3 Water Content on Relative Saturation Basis

An additional means of characterizing the soil water content is in terms of the degree of saturation:

$$\Theta = \frac{(\text{volume of water-filled pore space})}{\text{total volume of soil pore space}} = \frac{\theta_v}{\theta_{vs}}, \qquad (4.4)$$

where θ_{vs} is volumetric soil water content under completely water-saturated conditions. This index ranges from zero in completely dry soil to unity in a saturated soil. The degree of saturation is also commonly termed effective saturation or relative water content.

4.2.1.4 Soil Water Storage

It is often convenient to express the quantity of soil water in a specific soil depth increment in terms of soil water storage or equivalent depth of soil water (units of length). Equivalent depth of soil water, D_e (m), is calculated as $D_e = \theta_v \times D$, where D is the soil depth increment (m) having volumetric water content θ_v. This quantity is useful to relate aboveground water dimensions of rainfall, irrigation, or evaporation (L) to belowground dimensions (θ_v, m^3 water m^{-3} soil). For example, one may wish to quantify changes in soil water content arising from addition of water by rainfall or loss of water by evaporation (Chapter 6). A useful rule of thumb is that 1 L of water delivered to 1 m^2 of soil is equivalent to a 1 mm depth increment of water. For suitable accuracy, it is often necessary to sum the equivalent depth relationship over discrete soil depth layers having distinct water contents:

$$D_e = \sum_{i=1}^{n} \theta_{vi} D_i, \qquad (4.5)$$

where i denotes depth increments.

4.2.2 Measurement of Soil Water Content

4.2.2.1 Thermogravimetry

This is a direct and destructive method whereby a soil sample is obtained by augering or coring into the soil; its volume need not be known. The sample is weighed at its initial wetness and then dried to remove interparticle absorbed water, and some, but not always all structural water trapped within clay lattices known as crystallization water. The conventional protocol is to oven dry the samples at 105°C until the soil mass becomes stable; this usually requires 24–48 h or more, depending on the sample size, wetness, and soil characteristics (texture, aggregation, etc.). The difference between the wet and dry weights is the mass of water held in the original soil sample (Section 4.2.1.2). Despite the somewhat arbitrary specification of a standard oven temperature (105°C) and the variable drying period (depending on specific conditions), the gravimetric method is considered the standard against which many indirect techniques are calibrated. The method is not without bias and/or error, however, and Gardner (1986) discussed some sources of these including the potential for water loss between sampling time and initial weighing, precision of the three weights involved (wet, dry, and tare), and the unknown amount of soil texture-dependent residual water associated with the clay fraction.

The primary advantages of gravimetry are the direct and relatively inexpensive processing of samples. The shortcomings of this method are its destruction of the sample location, labor- and time-intensive nature, the time delay required for drying (although this may be shortened by use of a microwave rather than conventional oven, the methodology has not yet been standardized), and the fact that the method is destructive thereby prohibiting repetitive measurements within the same soil volume.

4.2.2.2 Neutron Scattering

This is a nondestructive but indirect method commonly used for repetitive field measurement of volumetric water content (Greacen et al., 1981). It is based on the propensity of H nuclei to slow (thermalize) high-energy fast neutrons. A typical neutron moisture meter consists of (1) a probe containing a radioactive source that emits high-energy (2–4 MeV) fast (1600 km s^{-1}) neutrons, as well as a detector of slow neutrons; (2) a scaler to electronically monitor the flux of slow neutrons; and optionally (3) a data logger to facilitate storage and retrieval of data (Figure 4.1a). The radioactive source commonly contains a mixture of ^{241}Am and Be at 10–50 mCi. The ^{241}Am emits β particles that strike the Be and cause emission of fast neutrons.

When the probe is lowered into an access tube, fast neutrons are emitted radially into the soil where they collide with various atomic nuclei. Collisions with most nuclei are virtually elastic, causing only minor loss of kinetic energy by the fast neutrons. Collisions with H nuclei, which have similar mass of neutrons, cause a significant loss of kinetic energy and slow down the fast neutrons (consider a marble [neutron] colliding with a similarly sized ball bearing [H nucleus] vs. a stationary bowling ball [larger atomic nucleus]). When, as a result of repeated collisions, the speed of fast neutrons diminishes to those at ambient temperature (about 2.7 km s^{-1}), with corresponding energies of about 0.03 eV, they are called thermalized or slow neutrons. Thermalized neutrons rapidly form a cloud of nearly constant density near the probe, where the flux of the slow neutrons is measured by the detector. The average loss of the neutrons' kinetic energy, thus the relative number of slow neutrons, is
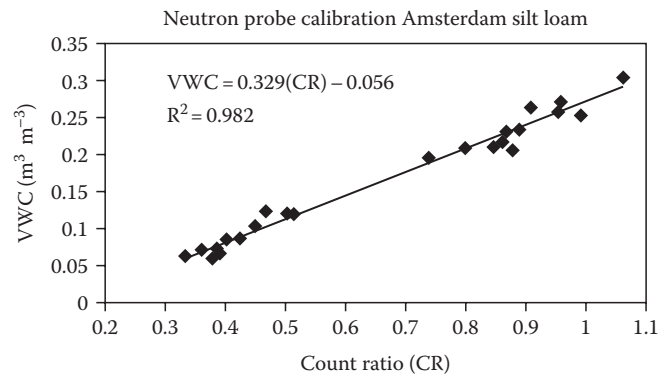
FIGURE 4.2 A neutron moisture meter calibration relationship.

therefore proportional to the amount of H nuclei in the surrounding soil. The primary source of H in soil is water; other sources of H in a given soil are assumed to be constant and are accounted for during calibration. Although several non-H substances including C, Cd, Bo, Cl, and Li, which may be present in trace amounts in some soils, may also thermalize fast neutrons, these may generally also be effectively compensated through soil-specific calibration.

Calibration of the neutron probe is therefore required to account for background H sources and other local effects (soil bulk density, trace neutron attenuators), and is conveniently achieved by paired measurements of soil water content and neutron probe counts. The calibration curve (Figure 4.2) is usually linear and relates θ_v to slow neutron counts or count ratio (CR):

$$\theta_v = a + b(CR), \tag{4.6}$$

where CR is the ratio of slow neutron counts at a specific location in the soil to a standard count obtained with the probe in its shield. For many soils, the calibration relation is approximately the same. Use of the count ratio rather than raw slow neutron counts compensates for the slow decay of the radioactive source over time.

The sphere of influence about the radiation source varies between about 15 cm (wet soil) and approximately 70 cm (very dry soil), depending on how far fast neutrons must travel in order to collide with a requisite number of H nuclei. An approximate equation for the radius of influence (r, in cm) as a function of soil wetness is (IAEA, 1970):

$$r = 15(\theta_v)^{-1/3}. \tag{4.7}$$

Thus, the neutron-scattering method is unsuitable for measurement near the soil surface because a portion of the neutrons may escape the soil. Advantages of this method include the ability to repeatedly measure θ_v at the same locations, averaging of the measured water content over a substantial soil volume, moderate equipment cost (generally about $4000–$6000), and the ability to measure soil water content at multiple depths and locations using the same equipment. Limitations or disadvantages include

(a)   (b)
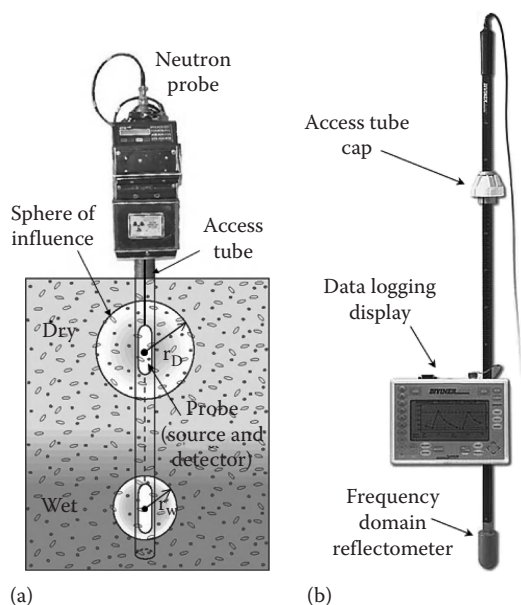
FIGURE 4.1 An illustration of (a) a neutron probe device for measuring soil water content; symbols r_D and r_W represent different radii of measurement in dry and wet soils and (b) an electromagnet-based sensor using frequency domain reflectometry (Diviner 2000, Sentek Sensor Technologies, Stepney, Australia, with permission) to determine soil water content in an access tube.

the radiation hazard and attendant licensing requirements, relatively poor (and uncertain) spatial resolution, unsuitability for near-surface measurements, and the soil-specific calibration requirement.

4.2.2.3 Electric and Dielectric Methods

Measurement methods based on changes in soil electrical properties, due to changes in soil water content, have been used for decades (Smith-Rose, 1934; Babb, 1951), mostly in the area of agriculture and exploration geophysics. Background information concerning such applications may be found in a number of sources, including Olhoeft (1985), who addressed low-frequency electrical properties of porous media, Selig and Mansukhani (1975), who discussed early attempts to use resistance and capacitance techniques for soil water content measurements, and Hoekstra and Delaney (1974) concerning soil dielectric properties at very high frequencies. Presently, the most common electrical approaches for soil water content measurement may be grouped according to (1) time domain dielectric methods (Topp et al., 1980; Robinson et al., 2003), (2) capacitance methods (Dean et al., 1987; Paltineanu and Starr, 1997), (3) fixed frequency methods (Blonquist et al., 2005a), (4) resistance methods (Spaans and Baker, 1992; Leib et al., 2003), (5) geophysical electrical methods (Robinson et al., 2008a), and (6) remote-sensing methods (Moran et al., 2004).

Determination of dielectric properties is the physical principle that underlies many of these methods to determine water content. In the megahertz region of the electromagnetic spectrum, the dielectric values for solid, water, and air are relatively frequency independent (Figure 4.3). As the dielectric value of water is ~80 and that of solid and air is ~5 and 1, respectively, the proportion of water in a porous material has a dominant effect on the effective dielectric value (ε_{eff}). Two basic approaches have

been used to establish the relationship between ε_{eff} and θ_v. The first approach is empirical, whereby mathematical expressions are simply fitted to observed data without using any particular physical model. For example, Topp et al. (1980) found good agreement using a third-order polynomial between apparent dielectric constant (κ_a) and θ_v for a range of soils:

$$\theta_v = -5.3 \times 10^{-2} + 2.92 \times 10^{-2} \kappa_a - 5.5 \times 10^{-4} \kappa_a^2 + 4.3 \times 10^{-6} \kappa_a^3,$$

(4.8)

where κ is used to differentiate a measured value from the intrinsic material value ε_{eff}. This relationship provides adequate description for many mineral soils and for the water content range $\theta < 0.5$ (which covers the entire range of interest in most mineral soils), with an estimation error of about 0.013 for θ_v. However, Equation 4.8 fails to adequately describe the $\kappa_a - \theta_v$ relationship for water contents exceeding 0.5, and for organic rich soils, mainly because the Topp et al. calibration was based on experimental results for mineral soils and concentrated in the range of $\theta_v < 0.5$.

The second approach uses dielectric mixing models and the volume fraction of each of the soil phases to derive a predictive relationship between ε_{eff} and θ_v. In many soils $\kappa_a = \varepsilon_{eff}$, but as bulk soil electrical conductivity (ECa) increases so κ_a may diverge from ε_{eff} on an instrument-dependent basis (Kelleners et al., 2005). Such physically based approaches vary in their degree of complexity and have been presented by a number of authors (Dobson et al., 1985; Friedman, 1998; Cosenza et al., 2003).

Of the physically based models, Robinson et al. (2005) presented a layer model for coarse-textured soils. The result was a formula for determining water content based on ε_{eff}, the measured porosity φ, and the saturated (ε_{sat}) and dry (ε_{dry}) dielectric values.
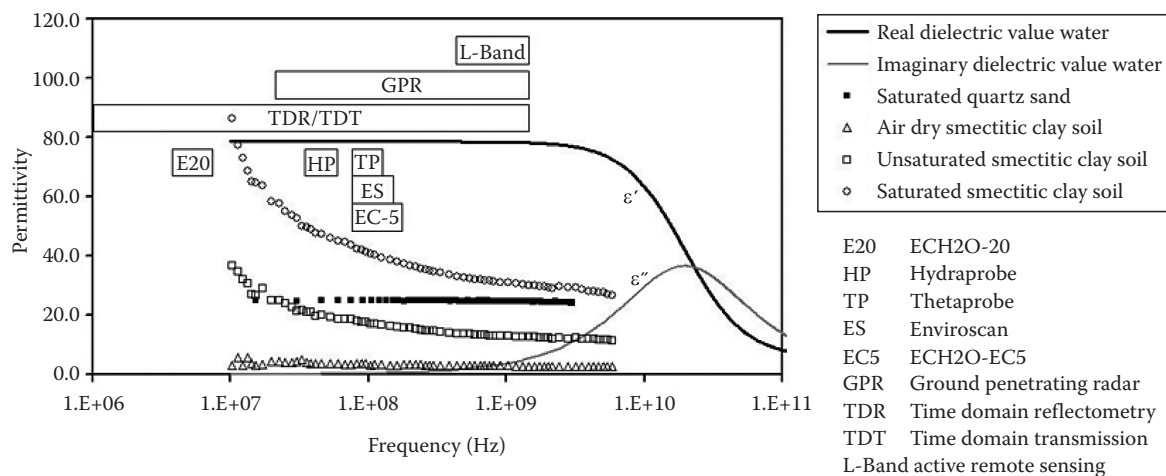
FIGURE 4.3 A frequency domain diagram showing the permittivity/dielectric spectrum of water, the real part (ε') indicating energy "storage," and the imaginary part (ε'') indicating energy "loss." Data are presented for sand and for clay with varying degrees of saturation. The change in clay permittivity response with frequency is called dielectric dispersion. The boxes with symbols indicate the approximate frequency and bandwidth for the corresponding sensing methods.

Assuming a value of 2.8 for ε_{dry}, the model only requires that the porosity and ε_{sat} be known to determine the water content.

$$\theta = \varphi \left(\frac{\sqrt{\varepsilon_{eff}} - \sqrt{\varepsilon_{dry}}}{\sqrt{\varepsilon_{sat}} - \sqrt{\varepsilon_{dry}}} \right). \tag{4.9}$$

A more sophisticated version of the model was presented in Robinson et al. (2005), where additional mixing models were used to calculate ε_{sat} and ε_{dry}, resulting in a physically based model with no fitting parameters that accurately described the $\varepsilon_{eff} - \theta_v$ relationship in media with different solid-phase dielectric values. The advantages of the mixing model approach are that it is based on physics and is therefore instrument or sensor independent and it provides the flexibility to include variable mineral phase dielectric values, temperature and frequency effects, and geometry and porosity effects. This is illustrated for a range of porosity in Figure 4.4.

Clay minerals, especially the 2:1 smectites and vermiculites, can give a frequency dependent dielectric response (Figure 4.3). As a result, they can give $\varepsilon_{eff} - \theta_v$ responses that diverge from standard empirical models such as Equation 4.8, generally giving lower ε_{eff} values for a respective value of θ_v. De Loor (1968) presented a model to describe the effective dielectric value (ε_{eff}) of the soil, which assumes a consistency of isotropically mixed plate-like particles, suitable for describing clays. The model can take two forms, a simplified version including solid, water, and air phases:

$$\theta = \frac{3(\varepsilon_s - \varepsilon_{eff}) + 2\varphi_s(\varepsilon_a - \varepsilon_s) - \varepsilon_{eff}\varphi_s(\varepsilon_s/\varepsilon_a - 1)}{\varepsilon_{eff}(\varepsilon_s/\varepsilon_{fw} - \varepsilon_s/\varepsilon_a) + 2(\varepsilon_a - \varepsilon_{fw})}, \tag{4.10}$$
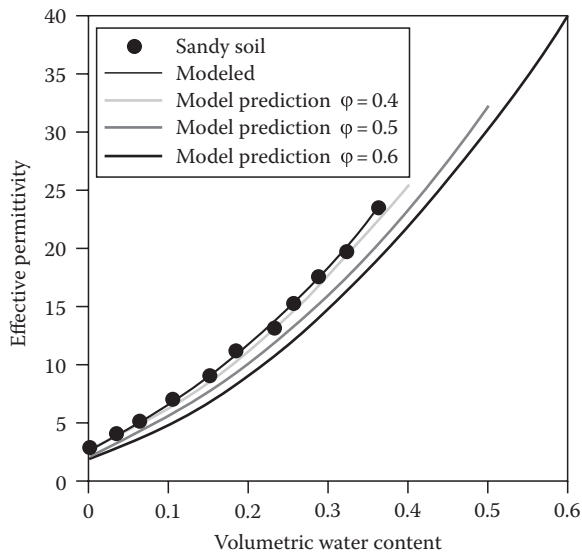
FIGURE 4.4 The dielectric water content relationship modeled using the approach described in Robinson et al. (2005). The different lines indicate the change in calibration due to the change in porosity.

and the one applied by Dobson et al. (1985) and Dirksen and Dasberg (1993), which adds a "bound" water phase, written as

$$\theta = \frac{\begin{array}{c}(3(\varepsilon_s - \varepsilon_{eff}) + 2\theta_{bw}(\varepsilon_{bw} - \varepsilon_{fw}) + 2\varphi_s(\varepsilon_a - \varepsilon_s) \\ + \varepsilon_{eff}\theta_{bw}(\varepsilon_s/\varepsilon_{fw} - \varepsilon_s/\varepsilon_{bw}) - \varepsilon_{eff}\varphi_s(\varepsilon_s/\varepsilon_a - 1))\end{array}}{(\varepsilon_{eff}(\varepsilon_s/\varepsilon_{fw} - \varepsilon_s/\varepsilon_a) + 2(\varepsilon_a - \varepsilon_{fw}))}, \quad \theta > \theta_{bw}, \tag{4.11}$$

where

φ_s is the solid volume fraction

subscripts of ε define the solid phase, s, gas phase, g, free water phase, fw, and bound water phase, bw (~4.2)

The volumetric bound water content, θ_{bw}, assumes a monomolecular layer of tightly bound water of a thickness $\delta = 3 \times 10^{-10}$ m spread over the soil-specific surface area, S ($cm^2\ g^{-1}$), for a bulk density, ρ_b (g cm^{-3}) (i.e., $\theta_{bw} = \delta \times \rho_b \times S$).

Due to the broad applicability of these models for understanding measurements made using time domain reflectometry (TDR), insertion probes, ground penetrating radar (GPR), and remote sensing, significant effort has been made in trying to improve our understanding of the dielectric properties of porous media. Recent attempts to better understand the dielectric behavior of soils and porous media include, particle-size distribution (Robinson and Friedman, 2001), particle geometry (Friedman and Robinson, 2002; Chen and Or, 2006), soil temperature (Or and Wraith, 1999), frequency dependence (Chen and Or, 2006), and dielectric phase configuration in aggregated media (Miyamoto et al., 2003; Blonquist et al., 2006). For a long time, bound water was proposed as the physical reason for the lower dielectric responses of clay soils (Jones and Or, 2003), but more recent literature suggests that particle shape, orientation (Jones and Friedman, 2000), and phase configuration (Blonquist et al., 2006) must also be considered.

4.2.2.4 Time Domain Reflectometry and Time Domain Transmissometry

Progress in devices for accurate measurement of reflected electromagnetic signals traveling along transmission lines led to the introduction of several measurement methods for soil dielectric properties using transmission lines embedded in the soil, such as TDR (Topp and Ferré, 2002; Robinson et al., 2003). TDR measures the apparent dielectric constant of the soil (κ_a) surrounding a TDR probe, at microwave frequencies from megahertz to gigahertz (Figure 4.3). The propagation velocity (v) of an electromagnetic wave along a transmission line (TDR probe) of length L embedded in the soil is determined from the time response of the system to a pulse generated by the TDR cable tester (Figure 4.5). The propagation

Time domain reflectometry unit
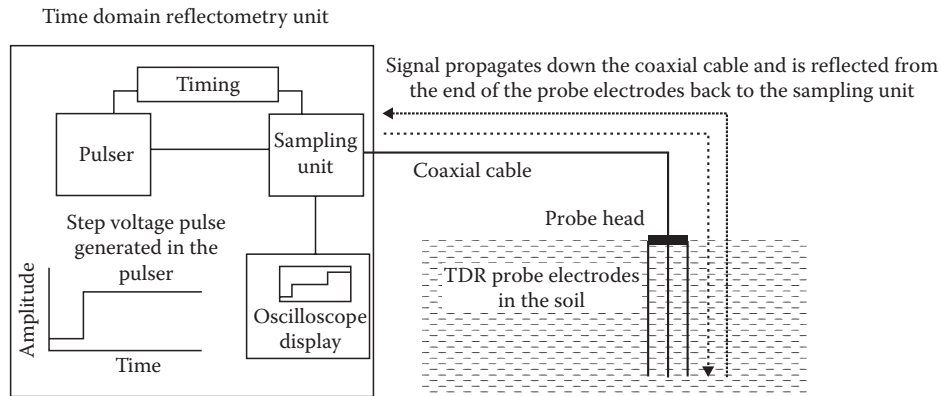
FIGURE 4.5 Schematic diagram of a TDR instrument attached via a coaxial cable to a three-rod TDR probe.

velocity (v = 2L/t) is a function of the apparent dielectric response of the soil according to

$$\kappa_a = \left(\frac{c}{v}\right)^2 = \left(\frac{ct}{2L}\right)^2, \tag{4.12}$$

where

  c is the velocity of electromagnetic waves in vacuum (3×10^8 m s^{-1})

  t is the travel time for the pulse to traverse the length of the embedded TDR probe in both directions (down and back)

A comprehensive body of literature now exists on TDR, and especially the electrode sampling volume (Ferre et al., 1998).

The main advantages of TDR over other methods for repetitive soil water content measurement are (1) superior accuracy to within 1% or 2% of θ_v, (2) minimal calibration requirements (in many cases soil-specific calibration is not needed), (3) no radiation hazards associated with neutron probe or gamma attenuation techniques, (4) excellent spatial and temporal resolution, and (5) simple to obtain continuous soil water measurements through automation and multiplexing (Baker and Allamaras, (1990)). Limitations of the TDR method include relatively high equipment expense, limited applicability under highly saline conditions due to signal attenuation, and the soil-specific calibration required for soils with high clay or organic matter contents. Severe attenuation of TDR waveforms in the presence of high salinity and/or some clays having high surface area and surface charge may interfere with, or even preclude, water content measurements.

Time domain transmission (TDT) sensors have more recently emerged (Blonquist et al., 2005b) as low cost, stand-alone sensors that have comparable measurement characteristics to TDR. These sensors are analogous to TDR, but rather than measuring the reflected signal, they measure the travel time around an electrode loop buried in the soil. Utilizing SDI-12 digital technology, they represent a new generation of multifunctional soil moisture sensors, also capable of determining bulk soil ECa and soil temperature.

4.2.2.5 Capacitance Probes

The electrical capacitance of two electrodes inserted in the soil is dependent upon the soil dielectric constant. The basic relationship between the soil dielectric constant and the electrical capacitance between two parallel plates of area A and spacing d is

$$C = \frac{A\varepsilon^* \varepsilon_0}{d}, \tag{4.13}$$

where

  ε^* is the relative complex dielectric constant of the soil, which contains both real (ε') and imaginary (ε'') components (Figure 4.3), with $\varepsilon^* = \varepsilon' - i\varepsilon''$, $i = \sqrt{-1}$

  ε_0 the dielectric constant of free space = 8.854×10^8 F m^{-1}

Note that the relative dielectric constant (ε^*) is the ratio of the complex dielectric constant to that of free space = $\varepsilon^*/\varepsilon_0$. Capacitance devices for field water content measurement are often based on annular electrode design rather than parallel plates (Kutilek and Nielsen, 1994; Fares and Polyakov, 2006) to facilitate depth measurements down boreholes (Dean et al., 1987; Figure 4.1b).

Commercial capacitance soil water gauges often use a resonant LC circuit relating changes in resonant frequency of the circuit to changes in the soil capacitance (Kelleners et al., 2004). Some gauges use a more direct capacitance bridge method to determine the unknown soil capacitance. Among the common capacitance-based measurement systems capable of soil water content profiling are the Sentry 200-AP (Troxler, NC) evaluated by Evett and Steiner (1995), which uses access tubes and a permanently installed modular bank of sensors (EnviroSCAN, Sentek, Australia) evaluated by Paltineanu and Starr (1997).

Advantages of capacitance methods include their lack of radiation hazard and lower expense than transmission line approaches such as TDR. However, research indicates greater sensitivity to bulk soil ECa and salinity (Kelleners et al., 2005). They also share the neutron probe's variable and uncertain measurement volume, and annular air gaps around sensors that utilize access tubes can cause substantial measurement errors.

Hence, buried probe designs seem to perform more reliably at present than those inserted into soil access tubes. Finally, capacitance methods share similar issues of relating the measured dielectric constant to soil water content as do other dielectric-based approaches.

4.2.2.6 Fixed Frequency Insertion Probes

Several new sensors and measurement methods are based on combinations of capacitive, reflective, and frequency-shift principles, all of which are governed by the soil dielectric properties. This trend appears highly promising for the development of accurate and cost-effective sensors for soil water content measurement. Examples of such stand-alone sensors (Figure 4.6) are the ThetaProbe (Gaskin and Miller, 1996) operating at about 100 MHz (Delta-T Devices Ltd, Cambridge, U.K.); the 50 MHz Hydra Probe (Campbell, 1990; Seyfried and Murdock, 2004; Stevens, Beaverton, OR); the 5 MHz ECH2O EC-20 probe (McMichael and Lascano, 2003; Decagon Devices, Pullman, WA), and newer sensors such as the Delta-T SM200 and WET sensors (Delta-T Devices Ltd, Cambridge, U.K.) both of which operate around 20 MHz and an array of Decagon Devices sensors including the EC-5, 10-HS, EC-TM, and 5TE, which all operate at 70 MHz frequency (Decagon Devices, Pullman, WA).

4.2.2.7 Electrical Resistivity Methods

Changes in the electrical resistivity of soils with changes in water content (and with soluble ionic constituents) have been used to develop simple and inexpensive sensors to infer soil water status. These sensors usually consist of concentric or parallel electrodes embedded in a porous matrix and connected to lead wires for measurement of electrical resistance within the sensor's porous matrix. The commonly used term gypsum block arises from early models that were in fact made of gypsum (Bouyoucos and Mick, 1940), and from the practice of saturating the matrix of many sensors made from alternative materials with gypsum to buffer local soil ionic effects. The sensor is embedded in the soil and allowed to equilibrate energetically with the soil solution. The matric potential of water in the sensor is estimated from the measured electrical resistance through previously determined calibration of the sensor itself (electrical resistance vs. matric potential). Under equilibrium conditions, the sensor matric potential is equal to the soil water matric potential; however, the sensor water content may be different than that of the soil. Hence, these measurements are often used to infer soil water matric potential from which the soil water content may be estimated based on a known relationship between these quantities (Gardner, 1986). With appropriate calibration for a particular soil (sensor electrical resistance vs. soil water content), the sensor can also be used to infer soil water content directly (Kutilek and Nielsen, 1994). The main advantages of electrical resistance sensors are their low cost and simple measurement requirements. Measurements may be obtained using a simple resistance meter or more conveniently acquired automatically using a data logger. On the other hand, the usual requirement for specific calibration of each individual sensor and for each soil to obtain acceptable accuracy, and lack of sensitivity under wet conditions, render this measurement method appropriate mostly as a qualitative indicator of soil water status (Spaans and Baker, 1992; Leib et al., 2003).

4.2.2.8 Geophysical Methods

The need for water content measurement at intermediate scales, such as the field to small watershed scale (Western et al., 2002), is driving an interest in using geophysical methods to bridge the gap between point measurements, using sensors such as TDR, and spatial measurements, using remote-sensing methods (Robinson et al., 2008b). Methods such as GPR, electrical resistivity imaging (ERI), and electromagnetic induction (EMI) have all been used to determine the spatial patterns of water content. Unlike many point sensors, where the operation and interpretation have become routine, geophysical methods still require a firm degree of user skill in both operation and interpretation. However, they offer the advantages of providing more spatially dense data and are non- or minimally invasive. Nevertheless, both ERI and EMI measure bulk soil ECa, with all the associated problems accompanied by a determination of water content from approximate direct-current (DC) electrical measurements. In addition, GPR uses unguided waves that can make soil water content interpretation more difficult. As a result, the data from geophysical imaging are often termed "soft data" and used as an aid to help interpret more spatially sparse point measurements.

GPR measures electromagnetic wave propagation time through the ground (Huisman et al., 2003a), utilizing the transmission and reflection of high frequency (1 MHz–1 GHz) electromagnetic waves within the soil. It offers the advantage
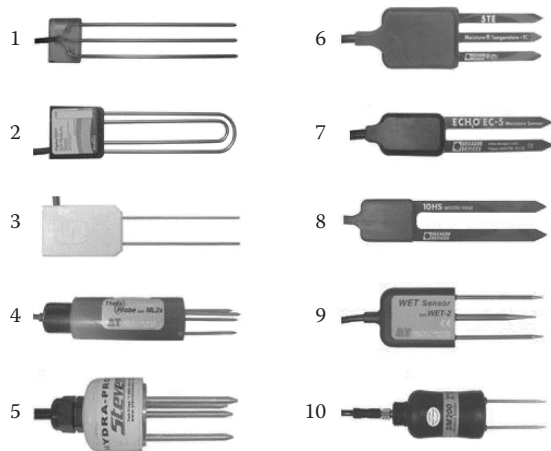
FIGURE 4.6 Array of different electromagnet-based sensors, which use dielectric measurement techniques to determine soil water content. Sensors include (1) TDR probe, (2) Acclima TDT, (3) Campbell Scientific 650 sensor, (4) Delta T ML2x sensor (Theta probe), (5) Stevens Water Hydra Probe sensor, (6) Decagon Devices 5TE sensor, (7) Decagon Devices EC-5 sensor, (8) Decagon Devices 10HS sensor, (9) Delta T Devices WET sensor, and (10) Delta T Devices SM200 sensor. Measurement techniques include travel-time analysis (1, 2, and 3), capacitance measurements (4, 6, 7, 8, 9, and 10), and impedance measurements (5).

of obtaining rapid on-the-go measurement, with penetration depths dependent on operation frequency and bulk soil ECa. Low frequencies usually equate to greater depths of penetration, while soil ECa values >1 dS m^{-1} (smectitic and saline soils) may attenuate the signal to such an extent that no valuable information is obtained (Weihermuller et al., 2007). A variety of methods have been adopted to determine water content from GPR data and are reviewed in Huisman et al. (2003a) and Annan (2005a, 2005b). Methods for determining water content include estimation of ground wave velocity (Hubbard et al., 2002; Huisman et al., 2003b); common offset profiling (Lunt et al., 2005); common midpoint measurements (Greaves et al., 1996); and surface reflectivity (Redman et al., 2002; Serbin and Or, 2004).

Both ERI (Samouelian et al., 2005) and EMI (Sheets and Hendrickx, 1995) determine water content by exploiting changes in soil ECa brought about by changes in water content. However, soil ECa is temperature, texture, and solute dependent, which can make interpretation difficult. Skilled users try to minimize certain factors, such as temperature effects by measuring at the same time during the day on repeated days, and by avoiding measurements at the hottest times. Determination of water content in soils with few, or no, retained cations, such as many highly weathered tropical or coarse-textured soils, often remains beyond the limits of these techniques.

The ERI method uses a low-frequency alternating-current to determine the electrical resistivity. Electrodes are evenly spaced in soil, often 0.25 to more than 5 m apart along a transect, and the resistivity is measured between different pairs of electrodes (Samouelian et al., 2005). By increasing the spacing between the measurement electrodes, greater depths of exploration can be achieved. By measuring multiple electrodes simultaneously, a 2D transect can be rapidly obtained. Successful application has been demonstrated using ERI for water content determination along transects by Michot et al. (2003) and Amidu and Dunbar (2007).

EMI instruments are configured with a receiver at one end and a transmitter loop at the other. The transmitter is energized and creates magnetic field loops that penetrate into the ground. These magnetic loops produce electrical field loops that in turn create a secondary magnetic field. At low induction numbers, the combined primary and secondary magnetic fields measured in the receiver are proportional to the bulk soil ECa (McNeill, 1980). As a result, EMI is noninvasive and can be linked to a GPS and field computer for the collection of spatial field data. The use of EMI to determine water content was initially reported by Kachanoski et al. (1988). Due to the instrument sensitivity to water and clay contents, Abdu et al. (2008) have proposed using the instrument to estimate soil moisture storage in small watersheds.

4.2.2.9 Remote-Sensing Methods

Remote-sensing methods may be vehicle, aircraft, or satellite based. They offer the ability to measure a large spatial area that is often more appealing to hydrologists in view of the scales at which they work (Moran et al., 2004). Airborne or spaceborne

measurements also tend to be cost effective and allow measurements in isolated terrain. Dielectric remote-sensing methods can be divided into "passive" and "active" methods.

Passive methods rely on the sun illuminating the ground surface and measure the naturally emitted radiation from the ground and do not work when it is cloudy. The radiation is measured using an instrument called a radiometer (Njoku and Entekhabi, 1996), and of growing interest are the systems that can be building or vehicle mounted. The soil emits radiation that increases with temperature and is termed the soil emissivity. In the microwave region, the dielectric value of the soil, as well as the surface geometry or roughness, affects how closely the surface resembles a perfect emitter or blackbody. As a result of the dependence of emissivity on dielectric properties, water content has a strong impact on measurements (Schmugge et al., 1974; Hallikainen et al., 1985). The amount of surface soil contribution to the emission varies with wavelength or frequency, becoming thinner as frequency increases. Measuring at a frequency of 1.4 GHz, the emitted radiation is considered sensitive to the upper 3–5 cm (Laymon et al., 2001).

Active sensing methods usually have a source to illuminate the soil surface and measure the response. Measurements using active methods are also depth dependent, according to measurement frequency, with L-band (1–2 GHz) being considered optimal for soil moisture determination (Entekhabi et al., 2004). Like passive for similar frequencies, the L-band active methods are considered to represent a response from the upper 5 cm of soil. One of the major advantages for the active method is that it does not rely on the soil surface being illuminated by the sun and can operate in almost all weathers. This has led to greater investigation of ground-based methods using horn antennas (Lambot et al., 2004, 2006; Serbin and Or, 2004).

Common to both measurement methods is the use of the data to determine soil moisture, and increasingly data assimilation methods are being developed to determine dynamic soil moisture behavior (Ni-Meister, 2008). Data assimilation allows the merging of observations with a system-dynamic model so that an improved estimate of the soil moisture state can be achieved. The direct data insertion approach (Jackson et al., 1981) takes remotely sensed soil moisture data and inserts them into the dynamic model when available. However, it does assume that remotely sensed observations give the "true" moisture content. More recent developments include the use of the extended Kalman filter, which is a sequential data assimilation tool, providing an optimal estimate for current and future states of a linear system. Hydrological processes are often nonlinear and so efforts have been made to linearize hydrological processes so that soil moisture data can be used (Walker and Houser, 2001).

4.2.2.10 Heat Pulse Probes

Determination of water content using soil thermal properties is becoming increasingly popular. Water content can be determined from the volumetric heat capacity (C_{soil}) or from the soil thermal conductivity (λ_{soil}). The soil volumetric heat capacity can be determined from the sum of the heat capacities of soil

solid, liquid, and gaseous phases (De Vries, 1963). Campbell (1985) introduced a simplified model for mineral soils C_{soil}, which disregards the heat capacity of air:

$$C_{soil} = C_s(1 - \varphi) + C_w\theta, \quad (4.14)$$

where

C is the heat capacity with subscripts w and s for water and solid phases respectively

φ is the porosity

Given the measurement of C_{soil} using a sensor, θ can be determined given the values of C_s and C_w from calorimetric methods (Kluitenberg, 2002). Water content determined from thermal conductivity, like ECa and dielectric properties, is dependent on the pore-scale geometry of the granular material (De Vries, 1963; Campbell, 1985), and is thus soil specific.

Campbell et al. (1991) proposed the dual-probe heat pulse (DPHP) method as a way of determining soil thermal properties. The method can measure the C, λ, or thermal diffusivity, α, from which θ can be determined. A short heat pulse is applied from a heater in one of the sensor needles and the temperature response is measured in a second needle separated by a short distance. The method has now been tested extensively in the laboratory (Bristow et al., 1993; Bristow, 1998; Basinger et al., 2003) and in field soils (Tarara and Ham, 1997; Heitman et al., 2003).

4.2.2.11 Emerging and Other Technologies

Laboratory investigations have resulted in a number of techniques emerging that are based on sound and light transmission. Brutsaert and Luthin (1964) and Mack and Brach (1966) were the first to report the sensitivity of acoustic wave propagation to soil water content. Recent developments include work by Adamo et al. (2004) and Blum et al. (2004) with acoustic tomography to study unsaturated flow. Recently, Lu (2007) proposed that the acoustic properties are more closely related to the matric potential of the soil than to the water content and thus provide a means to determine matric potential over a wide range. The use of light, in particular, measurements of changes in the refractive index, has proved a valuable laboratory technique for measuring 2D water flow in Hele-Shaw cells. Tidwell and Glass (1994) demonstrated that using a light source behind a cell and measuring the transmission the spatial distribution of water content in the cell can be determined.

Nuclear magnetic resonance (NMR) was first demonstrated in 1946 (Purcell, 1952; Bloch, 1953). Gummerson et al. (1979) were quick to adapt this to imaging unsaturated water flow in porous media. Bottomley et al. (1986) later used the technology to image the water content in root systems. The imaging method utilizes the fact that the hydrogen nucleus (^{1}H) possesses the strongest nuclear magnetic moment. These nuclei are randomly aligned in the absence of a strong magnetic field. On the application of an external magnetic field, the magnetic moments align with the field, antiparallel or parallel. An excess of antiparallel alignment provides a detectable macroscopic magnetization.

The modulus of the magnetization vector aligned with the external field is proportional to the number of ^{1}H nuclei in the sample. Assuming that water content is proportional to the number of ^{1}H nuclei, the proportion of water can be determined. Hinedi et al. (1993) demonstrated this linear relationship for granular media in the laboratory and subsequently the technique has been increasingly applied to soils by a number of researchers (Bird et al., 2005). Field NMR instruments have been developed for industrial applications and are described in Wolter and Krus (2005). The one-sided access (OSA)-NMR system (Perlo et al., 2005) is noninvasive and can be placed flat on a surface, measuring water content to a depth of 5–30 mm. Recent advances have also been made by the oil industry with the development of borehole NMR (Kleinberg and Jackson, 2001). The advantage of these methods is the ability to obtain additional information on the structure of the porous media (Song et al., 2000).

A critical constraint to advancing field-based process hydrology is the lack of spatially dense data to monitor changes in water content in time and space. Wireless sensor networks are developing as a new technology that allows sensors to be distributed in space, and wirelessly beam back measurements over radio links to a computer, so that real-time monitoring in time and space can be obtained (Akyildiz et al., 2002; Martinez et al., 2004). Soil moisture monitoring networks are emerging. Cardell-Oliver et al. (2005) presented a novel reactive soil moisture network that captures data with greater frequency in response to rainfall events. Some of the challenges faced in design and implementation are considered in Robinson et al. (2008b).

Other methods for soil water measurement include gamma ray attenuation techniques using dual-probe apparatus for bulk density and water content, x-ray computed tomography, and field techniques such as distributed fiber optic determination and the geophysical use of gravity. Additional information on these and other methods may be found in Dane and Topp (2002), Baker (1990), Campbell and Mulla (1990), Carter (1993), and Robinson et al. (2008a).

4.2.3 Applications of Soil Water Content Information

4.2.3.1 The Water Balance

A primary use of soil water content information is for evaluation of the hydrologic water balance described by

$$P + I = ET + D + R - \Delta W, \quad (4.15)$$

where

P is precipitation

I is irrigation

ET is evapotranspiration (evaporative water loss from the soil and from plants)

D is drainage or deep percolation

R is surface runoff

ΔW is change in water storage within the profile (soil water depletion)
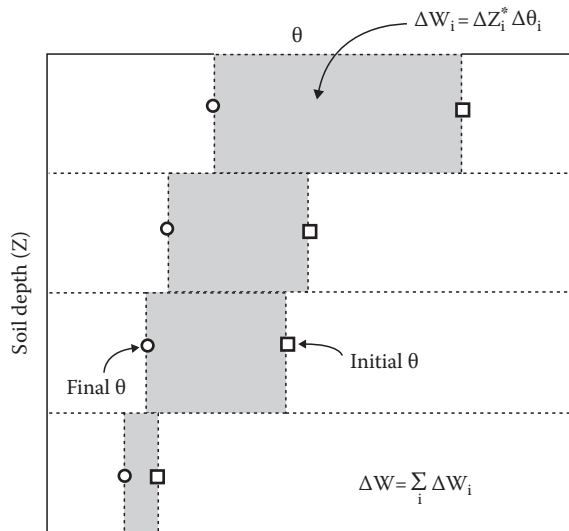
FIGURE 4.7 Schematic of soil water depletion calculations for a soil profile divided into four depth increments. Total change in water storage is the sum of depletion in each layer.

W is defined as the equivalent depth of water (D_e) stored in the soil profile under consideration and $\Delta W = (W_{initial} - W_{final})$ (Figure 4.7). These parameters are all associated with a given specific time interval. The convention used here is that inputs to the soil profile are taken as positive and outputs negative. The concept is based on conservation of mass (water) and is similar to the familiar exercise of balancing inputs and outgoings from a checking account, for example. Under typical conditions, ΔW is fairly significant over the short term (weeks to months), but generally evens out to about zero over one to several years. Measurement of water content for water balance studies requires the determination of profile soil moisture. An international study (IAEA, 2008) found that the neutron probe remains the most accurate and precise method for the determination of profile soil water content in a range of soil types. Of the sensors tested, they found that conventional TDR and the neutron probe method were the only indirect methods to have acceptable accuracy for field water balance studies.

4.2.3.2 Field Capacity, Wilting Point, and Plant-Available Soil Water

Observations of water content changes in the soil profile following wetting by irrigation or rainfall show that the rate of change decreases in time. In some cases, water content attains a nearly constant value within 1–2 days after wetting, following soil water redistribution in response to internal drainage. Field capacity is defined as the water content at which internal drainage becomes essentially negligible. The attainment of a near constant water content at field capacity (θ_{vFC}) is not always assured. It is dependent on (1) the depth of wetting and the antecedent (initial) water content of the soil profile (for a soil that is moist at the onset of wetting and for deep wetting, the rate of redistribution is slower and the apparent value of θ_{vFC} is higher) and (2) the

presence of impeding layers or a water table, which affect the rate and extent of water redistribution.

Another often misunderstood soil water content-related index is the permanent wilting point, which is defined as the water content at which plants can no longer extract soil water at a rate sufficient to meet physiological demands imposed by loss of water to the atmosphere, and thus irreversibly wilt and die. This water content (θ_{vWP}) is primarily dependent on the soil's ability to transmit water, but also to some degree on the plant's ability to withstand or mitigate drought. Though commonly taken as θ_v at −1.5 MPa (−15 bars) matric potential, there is substantial variation among plant species in their abilities and mechanisms to resist soil drought, with some plants, such as dryland shrubs and trees, surviving to potentials of approximately −8 MPa, well below the standard wilting point index. The permanent wilting point should not be confused with the phenomenon of transient wilting, which is commonly observed during the afternoon when evaporative demand is greatest. In this case, plants are able to rehydrate to some extent at night.

A primary practical use of field capacity and wilting point concepts is the determination of a plant-available soil water range (PASW; Figure 4.8). Soil water storage available for plant use is generally calculated as being between field capacity and wilting point ($\theta_{vFC} - \theta_{vWP}$), as water contents higher than θ_{vFC}, while usually plant-available, are generally not sustained for long periods of time except under specific circumstances. Plant-available soil water storage is an important factor in the determination of irrigation amounts for a cropped field or other soil–plant system. For practical purposes, irrigation amounts in excess of field capacity are lost to deep percolation, and thus, should be avoided in the interests of water resource efficiency as well as potential leaching of soluble chemicals (Chapter 7).

A useful rule of thumb is to estimate θ_{FC} as $\theta_s/2$, and θ_{WP} as $\theta_{FC}/2$; in other words, a soil exhibiting this property would
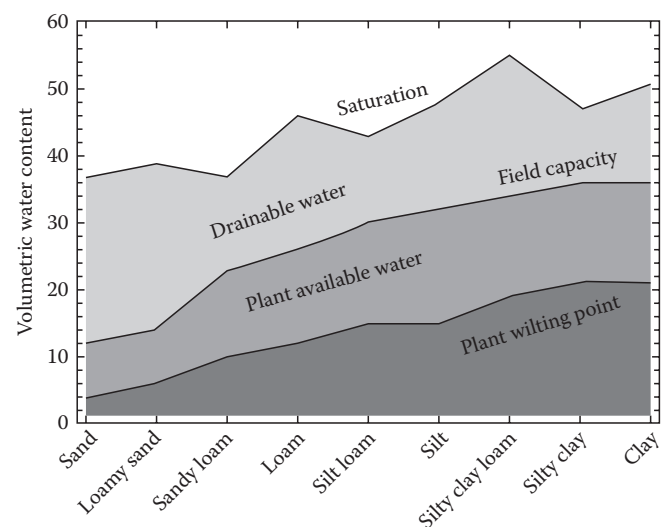
FIGURE 4.8 Schematic of estimated PASW content for a range of soil textural classes.

have lost 50% of its saturated water content at field capacity, and another 50% of the remaining water by the wilting point. Richards (1954) was probably the first to make this observation. Banin and Amiel (1970) and Dahiya et al. (1988) have also demonstrated that the assumed ratio is a good approximation, based on measured correlations on many soils.

4.2.3.3 Sources of Additional Information on Soil Water Content Indices

The information on these indices is fragmentary due to their dependency on other conditions in the soil profile (field capacity) and the plant in question (wilting point). The work of Slater and Williams (1965) provides a reasonable framework for estimating field capacity, and more recent information may be found on the Internet. A study by Waters (1980) is a good source of information regarding the wilting point. Dahiya et al. (1988) presented strong correlations among the saturation percentage, field capacity, wilting point, and available water based on data for 438 soils, with correlation coefficients in excess of 0.92. Banin and Amiel (1970) provided similar information concerning relationships among clay content, surface area, and water content at several index points. Petersen et al. (1996) also related soil surface area and other properties to their water-holding capacity.

4.2.3.4 The Role of Soil Moisture in Hydrology, Biogeochemistry, and Ecology

Soil moisture plays a critical role across a range of scales within hydrology. Entekhabi et al. (1996) demonstrated the mutual interaction between soil moisture states and atmospheric processes. Soil moisture becomes an important source of atmospheric water at continental scales, a fraction of which falls as precipitation back on the land surface downwind from the site of the original evapotranspiration. Soil moisture also plays an important role in determining the albedo of soils (Idso et al., 1974), this has become increasingly important in the global energy balance, as land-use change has occurred and more soils are cultivated and urbanization expands. At the watershed scale, the antecedent moisture conditions play a critical role in determining if rainfall will infiltrate or runoff, especially where saturation excess runoff processes dominate (Dunne and Black, 1970). As a result, significant effort has been placed on trying to determine water content across small watersheds (Grayson et al., 1997; Western and Grayson, 1998; Western et al., 1999).

Biogeochemical cycling describes the transformation of elements from one form to another. Soil moisture has been recognized from early in the history of soil science as a controlling factor over rates of weathering and soil formation (Jenny, 1941). In addition to controlling some weathering processes, such as hydrolysis and hydration, which are important for the release of nutrients, for example, phosphorus, perhaps more importantly moisture controls microbial activity in soils (Skopp et al., 1990). The microbial activity is the powerhouse behind many soil processes, for instance, nitrification and the evolution of CO_2 from microbial respiration (Stanford and Epstein, 1974; Schjonning et al., 2003; Turcu et al., 2005).

Ecohydrology has emerged as a distinct subdiscipline within hydrology and focuses in part on soil moisture (Rodriguez-Iturbe and Porporato, 2005). However, ecologists have been studying soil moisture and soil temperature for many years with regard to their role in determining plant community spatial organization and net primary productivity (Pastor and Post, 1986). Soil moisture provides the reservoir of water for plant growth and determining the available pool of water and its functionality is seen as increasingly critical in explaining and predicting vegetation response in drier climates (Ryel et al., 2008).

4.3 Soil Water Energy

As previously stated, water status in soils is characterized by both the amount of water present and by its energy state. Soil water is subjected to forces of variable origin and intensity, thereby acquiring different quantities and forms of energy. The two primary forms of energy of interest here are kinetic and potential. Kinetic energy is acquired by virtue of motion and is proportional to velocity squared. However, because the movement of water in soils is relatively slow (usually <0.1 m h^{-1}), its kinetic energy is negligible. The potential energy, which is defined by the position of soil water within a soil body and by internal conditions, is largely responsible for determining soil water status under isothermal conditions.

Like all other matter, soil water tends to move from where the potential energy is high to where it is low, in its pursuit of a state known as equilibrium with its surroundings. The magnitude of the driving force behind such spontaneous motion is a difference in potential energy across a distance between two points of interest. At a macroscopic scale, one can define potential energy relative to a reference state. The standard state for soil water is defined as pure and free water (no solutes and no external forces other than gravity) at a reference pressure, temperature, and elevation, and is arbitrarily given the value of zero (Bolt, 1976).

4.3.1 Total Soil Water Potential and Its Components

Soil water is subject to several force fields whose combined effects result in a deviation in potential energy relative to the reference state called the total soil water potential (ψ_T) and is defined as follows:

The amount of work that an infinitesimal unit quantity of water at equilibrium is capable of doing when it moves (isothermally and reversibly) to a pool of water at similar standard (reference) state (similar pressure, elevation, temperature, and chemical composition).

It should be emphasized, however, that there are alternative definitions of soil water potential using concepts of chemical potential or specific free energy of the chemical species in water (which is different than the soil solution, termed soil water here). Some of the arguments concerning the proper

definitions and their scales of application are presented by Corey and Klute (1985), Iwata et al. (1988), and Nitao and Bear (1996). Recognizing that these fundamental concepts are subject to ongoing debate, only simple and widely accepted definitions of these quantities, which are applicable at macroscopic scales and which yield an appropriate framework for practical applications, will be presented.

The primary forces acting on soil water held within a rigid soil under isothermal conditions can be conveniently grouped as (Day et al., 1967) (1) matric forces resulting from interactions of the solid phase with the liquid and gaseous phases, (2) osmotic forces owing to differences in chemical composition of the soil solution, and (3) body forces induced by gravitational and other inertial force fields (centrifugal).

The thermodynamic approach whereby potential energy rather than forces are used is particularly useful for equilibrium and flow considerations. Equilibrium would require the vector sum of these different forces acting on a body of water in different directions to be zero; this is an extremely difficult criterion to deal with in soils. On the other hand, potential energy defined as the negative integral of the force over the path taken by an infinitesimal amount of water, when it moves from a reference location to the point under consideration, is a scalar quantity. Consequently, the total potential can be expressed as the algebraic sum of the component potentials corresponding to the different fields acting on soil water:

$$\psi_T = \psi_m + \psi_s + \psi_p + \psi_z. \tag{4.16}$$

ψ_m is the matric potential resulting from the combined effects of capillarity and adsorptive forces within the soil matrix. Dominating mechanisms for these effects include (1) adhesion of water molecules to solid surfaces due to short-range London–van der Waals forces and extension of these effects by cohesion through H bonds formed in the liquid, (2) capillarity caused by liquid–gas (LG) and liquid–solid–gas interfaces interacting within the irregular geometry of soil pores, and (3) ion hydration and binding of water in diffuse double layers. There is some disagreement regarding the definition of this component of the total potential. Some consider all contributions other than gravity and solute interactions (at a reference atmospheric pressure). Others use a tensiometer (Section 4.3.4) to measure and provide a practical definition of the matric potential in a soil volume of interest (Hanks, 1992). The value of ψ_m ranges from zero when the soil is saturated to negative numbers when the soil is dry (note that $\psi_m = 0$ mm is $>\psi_m = -1000$ mm). Because it is often more convenient or intuitive to work with positive than negative quantities, the term matric suction or tension is commonly used. Each of these represents the absolute value of ψ_m.

ψ_s is the solute or osmotic potential determined by the presence of solutes in soil water, which lower its potential energy and its vapor pressure. The effects of ψ_s are important in the presence of (1) appreciable amounts of solutes and (2) a selectively permeable membrane or a diffusion barrier, which transmits water more readily than salts. The effects of ψ_s are otherwise negligible when only liquid water flow is considered and no diffusion barrier exists. The two most important diffusion barriers in the soil are (1) soil–plant root interfaces (cell membranes are selectively permeable) and (2) soil water–air interfaces; thus, when water evaporates, salts are left behind. In dilute solutions, the solute potential (also called the osmotic pressure) is proportional to the concentration and temperature according to

$$\psi_s = -RTC_s, \tag{4.17}$$

where

ψ_s is in kPa
R is the universal gas constant (8.314×10^{-3} kPa m^3 mol^{-1} K^{-1})
T is the absolute temperature (K)
C_s is solute concentration (mol m^{-3})

A useful approximation, which may be used to estimate ψ_s in kPa from the ECa of the soil solution at saturation (EC$_s$) in dS m^{-1}, is

$$\psi_s = -36EC_s. \tag{4.18}$$

ψ_p is pressure potential defined as the hydrostatic pressure exerted by unsupported water (i.e., saturating the soil) overlying a point of interest. Using units of energy per unit weight provides a simple and practical definition of ψ_p as the vertical distance from the point of interest to the overlying free water surface (unconfined water table elevation). The convention used here is that ψ_p is always positive below a water table or zero if the point of interest is at or above the water table. In this sense, nonzero magnitudes of ψ_p and ψ_m are mutually exclusive; either ψ_p is positive and ψ_m is zero (saturated conditions) or ψ_m is negative and ψ_p is zero (unsaturated conditions), or $\psi_p = \psi_m = 0$ at the free water table elevation. Another definition that is used in some quarters is to combine ψ_m and ψ_p as used here into a single component that adopts negative magnitude under unsaturated conditions and positive magnitude under saturated conditions.

ψ_z is gravitational potential, which is determined solely by the elevation of a point relative to some arbitrary reference point and is equal to the work needed to raise a body against the earth's gravitational pull from a reference level to its present position. When expressed as energy per unit weight, the gravitational potential is simply the vertical distance from a reference level to the point of interest. The numerical value of ψ_z itself is thus not important (it is defined with respect to an arbitrary reference level); what is important is the difference or gradient in ψ_z between any two points of interest. This value will not change with different reference point locations.

Soil water is at equilibrium when the net force on an infinitesimal body of water equals zero everywhere, or when the total potential is constant in the system. Though the last statement is a logical consequence of the definitions above, it is not strictly true as pointed out by Corey and Klute (1985). They argue that

constant total potential is a necessary but not a sufficient condition, and for thermodynamic equilibrium to prevail, three conditions must be met simultaneously: (1) thermal equilibrium or uniform temperature, (2) mechanical equilibrium meaning no net convection producing force, and (3) chemical equilibrium meaning no net diffusional transport or chemical reaction. In most practical applications, however, the macroscopic definition of the total potential and equilibrium conditions based on it is completely adequate (Kutilek and Nielsen, 1994).

The difference in chemical and mechanical potentials between soil water and pure water at the same temperature is known as the soil water potential (ψ_w):

$$\psi_w = \psi_m + \psi_s + \psi_p. \qquad (4.19)$$

Note that the gravitational component (ψ_z) is absent in this definition. Soil water potential is thus the result of inherent properties of soil water itself, and of its physical and chemical interactions with its surroundings, whereas the total potential includes the effects of gravity (an external and ubiquitous force field).

Total soil water potential and its components may be expressed in several ways depending on the definition of a unit quantity of water. Potential may be expressed as (1) energy per unit of mass, (2) energy per unit of volume, or (3) energy per unit of weight. A summary of the resulting dimensions, common symbols, and units are presented in Table 4.1.

Only μ has actual units of potential; ψ has units of pressure, and the units of h are in head of water (length). However, the above terminology (potential energy vs. units of potential) is widely used in a generic sense in the soil and plant sciences. The various expressions of soil water energy status are equivalent, with

$$\mu = \psi/\rho_w = gh, \qquad (4.20)$$

where
  ρ_w is density of water (1000 kg m^{-3} at 20°C)
  g is gravitational acceleration (9.81 m s^{-2})

4.3.2 Interfacial Forces and Capillarity

The matric potential is often the largest component of the total potential in partially saturated soils. To better understand the origins of this important potential, which attains nonzero values only under partial saturation when all the three phases (liquid, gas, and solid) are present, some of the properties of water in relation to porous media, which give rise to this phenomenon, must be discussed.

4.3.2.1 Surface Tension

At the interface between water and solids or other fluids such as air, water molecules are exposed to different forces than are molecules within the bulk water. For example, water molecules inside the liquid are attracted by equal cohesive forces to form H bonds on all sides, whereas molecules at the air–water interface feel a net attraction into the liquid because the density of water molecules on the air side of the interface is much lower and all H bonds are toward the liquid. The result is a membrane-like water surface having a tendency to contract; thus energy is stored in the form of surface tension (as in a stretched spring). Different liquids vary in their LG surface tension (σ_{LG}) expressed as energy per unit area = force per unit length. For example, water at 20°C: 72.7 mN m^{-1} = mJ m^{-2}; ethyl alcohol: 22 mN m^{-1} (=dyn cm^{-1} = erg cm^{-2}); and mercury: 430 mN m^{-1}.

4.3.2.2 Contact Angle

If liquid is placed in contact with a solid in the presence of a gas (three-phase system), the angle measured from the solid–liquid (SL) interface to the LG interface is the contact angle (γ). For a drop resting on a solid surface at equilibrium, the vector sum of the forces acting to spread the drop (outward) is equal to the opposing forces. This relationship is summarized by Young's equation (Adamson, 1990):

$$\sigma_{LG} \cos\gamma + \sigma_{SL} - \sigma_{GS} = 0, \qquad (4.21)$$

where σ is the respective interfacial surface tension. The equilibrium contact angle is therefore

$$\gamma = \cos^{-1}\left[\frac{\sigma_{GS} - \sigma_{SL}}{\sigma_{LG}}\right]. \qquad (4.22)$$

When liquid is more attracted to the solid (adhesion) than to other liquid molecules (cohesion), the angle is small and the solid is said to be wettable by the liquid. Conversely, when the cohesive exceeds the adhesive force, the liquid repels the solid and γ is large (Figure 4.9). The contact angle of water on clean glass is very small and is commonly taken as 0°. The contact

TABLE 4.1 Units, Dimensions, and Common Symbols for Potential Energy of Soil Water

Units	Symbol	Name	Dimensions	SI Units	cgs Units
Energy/mass	μ	Chemical potential	$L^2 t^{-2}$	J kg^{-1}	erg g^{-1}
Energy/volume	Ψ	Soil water potential, suction, or tension	$M (L t^2)^{-1}$	N m^{-2} (Pa)	erg cm^{-1}
Energy/weight	h	Pressure head	L	m	cm
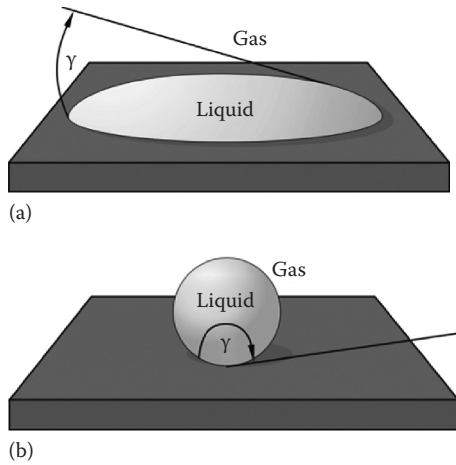
FIGURE 4.9 Liquid–solid–gas contact angles for (a) a small angle where the liquid wets the solid and (b) a large angle in which the liquid is repelled by the solid surface.

angle of soil water on soil minerals is also commonly assumed ~0°. The different contact angles of water and other liquids with soil solids, soil air, and with each other, are important contributors to the behavior of multiphase organic liquid/water/soil/air mixtures.

4.3.2.3 Curved Surfaces and Capillarity

Surface tension is associated with the phenomenon of capillarity. When the LG interface is curved rather than planar (flat), the resultant surface tension force normal to the LG interface creates a pressure difference across the interface. The pressure is greater at the concave side of the interface by an amount that is dependent on the radius of curvature and the surface tension of the fluid. For a hemispherical LG interface having radius of curvature R, the pressure difference is given by the Young–Laplace equation:

$$\Delta P = 2\sigma/R, \tag{4.23}$$

where

$\Delta P = P_{liq} - P_{gas}$ when the interface curves into the gas (water droplet in air)

$\Delta P = P_{gas} - P_{liq}$ when the interface curves into the liquid (air bubble in water, water in a small glass tube)

In many instances, a bubble may not be spherical or an element of liquid may be confined by irregular solid surfaces resulting in two (or more) different radii of curvature such as water in pendular rings between two spherical solid particles (Figure 4.10). The Young–Laplace equation for this case is given by:

$$\Delta P = \sigma\left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{4.24}$$

Note that (1) this equation reduces to Equation 4.23 for $R_1 = R_2$ and (2) the sign of R_i is negative for convex interfaces ($R_2 < 0$) and positive for concave interfaces ($R_1 > 0$).
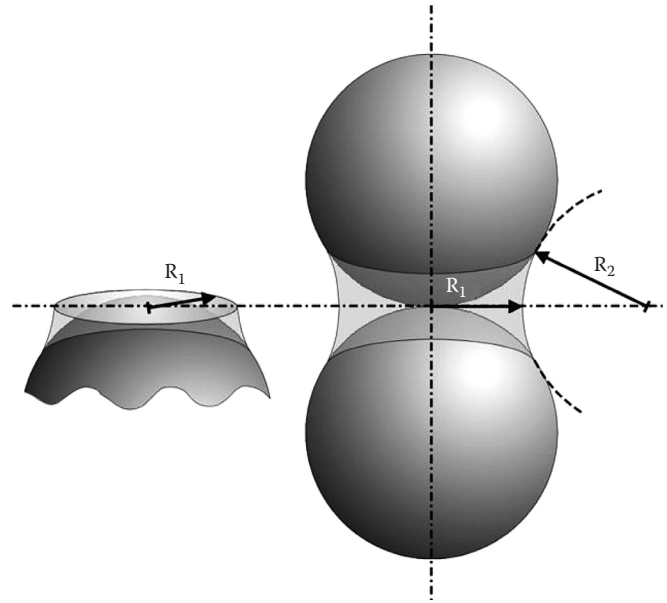
FIGURE 4.10 Radii of curvature for pendular ring of water between two spherical solids.

4.3.3 The Capillary Model

4.3.3.1 Capillary Rise

When a small cylindrical glass tube (capillary) is dipped in free water, a meniscus is formed as a result of the contact angle between the water and the walls of the tube and from consideration of minimum surface energy. The smaller the tube the larger the degree of curvature, resulting in a larger pressure difference across the air–water (gas–liquid) interface. The pressure at the water side (P_W) will be lower than atmospheric pressure (P_0). This pressure difference will cause water to rise into the capillary tube until the upward force across the water–air interface is balanced by the weight of water in the tube (Figure 4.11). Because the radius of meniscus curvature R = r/cos γ, where r is the tube radius, the height that water will rise in a capillary tube of radius r with γ contact angle is

$$h = \frac{2\sigma \cos \gamma}{\rho_w g r}. \tag{4.25}$$

Combining all the constants in Equation 4.25 (using typical values at room temperature) yields a simple and useful approximation:

$$h(m) = \frac{14.84}{r(\mu m)}. \tag{4.26}$$

When expressed as potential energy per unit weight, $\psi_m = -h$ and thus may be related to equivalent pore radius through Equation 4.25.
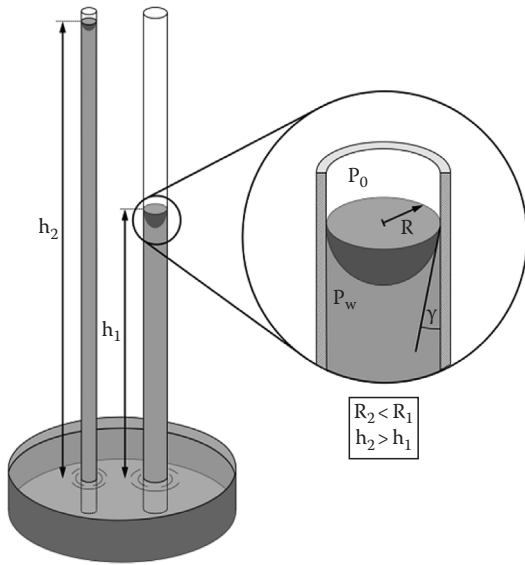
FIGURE 4.11 Capillary rise in cylindrical glass tubes.

4.3.3.2 Conceptual Models for Water in Soil Pore Space

The complex geometry of soil pore space creates numerous combinations of interfaces, capillaries, and wedges around which films of water are formed, resulting in a variety of air–water and solid–water contact angles. Water is thus drawn into and/or held by these interstices in proportion to the resulting capillary forces. In addition, water is adsorbed onto solid surfaces with considerable force at close distances. Due to practical limitations of the present measurement methods, no distinction is made between the various mechanisms affecting water in porous matrices (i.e., capillarity and surface adsorption). Common conceptual models for water retention in porous media and matric potential rely on a simplified picture of soil pore space as a bundle of capillaries. The primary conceptual steps made in such models are illustrated in Figure 4.12. The representation of soil pores as equivalent cylindrical capillaries greatly simplifies modeling and parameterization of soil pore space. The roles of water films at very low saturation levels and the unique contribution of surface adsorption to the matric potential are beyond the scope of this chapter. Interested readers are referred to reviews by Nitao and Bear (1996) and Parker (1986).

4.3.4 Measurement of Soil Water Potential Components

4.3.4.1 Tensiometer for Measuring Soil Matric Potential

A tensiometer consists of a porous cup, usually made of ceramic having very fine pores, connected to a vacuum gauge through a water-filled tube (Figure 4.13). The porous cup is placed in intimate contact with the bulk soil at the depth of measurement. When the matric potential of the soil is lower (more negative) than the equivalent pressure inside the tensiometer cup, water
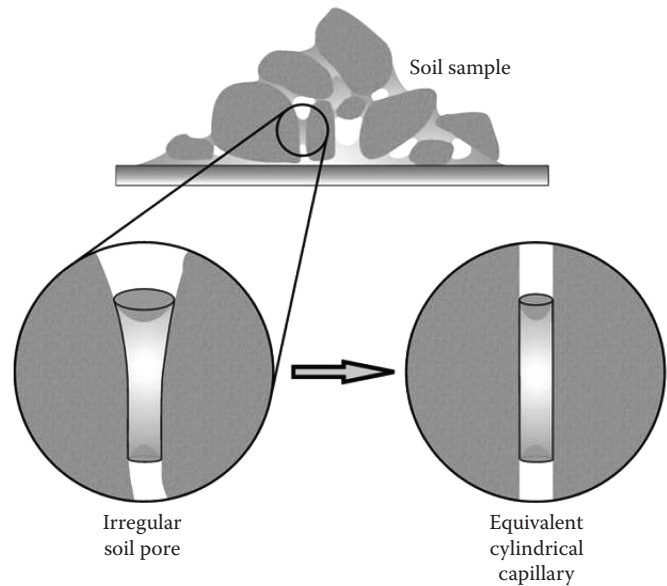
FIGURE 4.12 Concept of equivalent cylindrical capillary to represent soil pore spaces.

moves from the tensiometer along a potential energy gradient to the soil through the saturated porous cup, thereby creating suction sensed by the gauge. Water flow into the soil continues until equilibrium is reached and the suction inside the tensiometer equals the soil matric potential. When the soil is wetted, flow may occur in the reverse direction, that is, soil water enters the tensiometer until a new equilibrium is attained. The tensiometer equation is

$$\psi_m = \psi_{gauge} + (z_{gauge} - z_{cup}). \tag{4.27}$$

The vertical distance from the gauge plane (z_{gauge}) to the cup (z_{cup}) must be added to the matric potential measured by the gauge (expressed as a negative quantity) to obtain the matric potential at the depth of the cup, when potentials are expressed per unit of weight. This accounts for the positive head at the depth of the ceramic cup exerted by the overlying tensiometer water column.

Electronic sensors called pressure transducers often replace the mechanical vacuum gauges. The transducers convert mechanical pressure into an electric signal, which can be more easily and precisely measured. In practice, pressure transducers can provide more accurate readings than other gauges, and in combination with data logging equipment are able to supply continuous measurements of soil matric potential. Portable tensiometers that may be extended into boreholes for use at depths to several hundred meters have also been developed (Hubbell and Sisson, 1996). Achieving and maintaining adequate hydraulic contact between the porous cup and the soil at these depths was identified as an important issue, but these sensors may be highly useful in some mining, engineering, deep recharge, and hazardous waste applications.
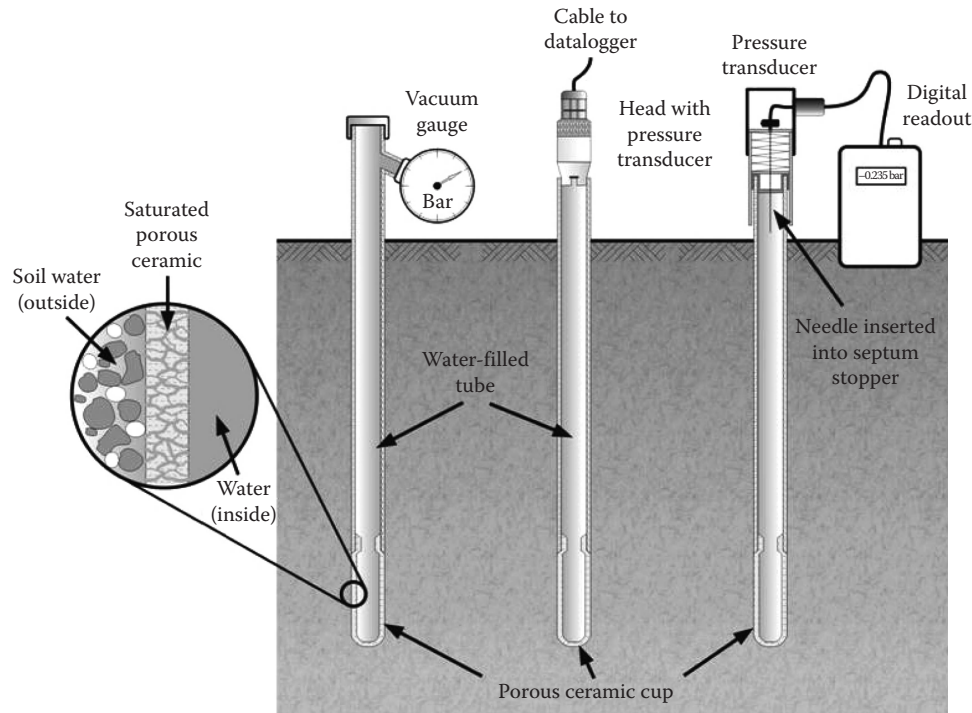
FIGURE 4.13   Pressure transducer and vacuum gauge tensiometers showing porous ceramic cup.

The standard tensiometer range is limited to suction values (absolute value of the matric potential) <100 kPa (1 bar or 10 m head of water) at sea level, and this value decreases proportionally with elevation. Thus, other means are needed to measure or infer soil matric potential under drier conditions.

Peck and Rabbidge (1969) described an osmotic tensiometer that relied on a confined aqueous solution behind a semipermeable membrane, rather than using free water as the reference state. A membrane highly impermeable to the confined solution allowed their device to cover the entire range of 0 to −1.5 MPa, unless soil solutes were excluded from the instrument by a vapor gap. Following the design of these osmotic tensiometers, polymer tensiometers have been developed, which utilize synthetic polymers (i.e., polyethylene glycol, polyacrylamide, and polyvinylpyrrolidone) in place of solutes to fill a small reservoir connected to the pressure transducer (Bakker et al., 2007). Keys to successful operation include reduction of thermal sensitivity, the prevention of polymer degradation or leakage from the reservoir using a secondary membrane, and maintenance of soil–ceramic contact for hydraulic continuity at low potentials.

4.3.4.2 Piezometer for Measuring Hydrostatic Pressure Potential

In a saturated soil such as below a water table, soil water is under positive hydrostatic pressure. The pressure potential (ψ_p) equals the vertical distance from a point in the soil to the surface of the free water table (recall that one expresses potential in terms of distance, length, or head when potential energy is expressed per unit of weight). The piezometer is a hollow tube placed in the soil to depths below the water table. It extends to the soil surface and is open to the atmosphere. The bottom of the piezometer is perforated to allow for soil water under positive hydrostatic pressure to enter the tube. Water enters the tube and rises to a height equal to that of the free water table. The water level within the piezometer may be determined using a variety of manual or automated measurement techniques.

4.3.4.3 Water Vapor Pressure for Measuring Water Potential

Under equilibrium conditions, the soil water potential is equal to the potential of water vapor in the surrounding soil air. A psychrometer measures the relative humidity (RH) of the water vapor, which is related to the water potential of the vapor (ψ_w) through the Kelvin equation:

$$\text{RH} = e/e_o = \exp^{[(M_w\psi_w)/(\rho_w RT)]}, \tag{4.28}$$

where
  e is water vapor pressure
  e_o is saturated vapor pressure at the same temperature
  M_w is the molecular weight of water (0.018 kg mol^{-1})
  R is the ideal gas constant (8.31 J K^{-1} mol^{-1} or 0.008314 kPa m^3 mol^{-1} K^{-1})
  T is absolute temperature (K)
  ρ_w is the density of water (1000 kg m^{-3} at 20°C)

Rearranging and taking a log transformation of Equation 4.28 yields

$$\psi_w = \frac{RT\rho_w}{M_w}\ln(e/e_o). \quad (4.29)$$

Equation 4.29 can be further simplified for the range of e/e_o near 1 often encountered in soils (the entire range of plant-available water is between $e/e_o = 1$ and 0.98):

$$\psi_w = \frac{RT\rho_w}{M_w}\left(\frac{e}{e_o} - 1\right) \approx 462\,T\left(\frac{e}{e_o} - 1\right), \quad (4.30)$$

for ψ_w in kPa. Note that because most salts are nonvolatile, the psychrometric measurement of ψ_w is the sum of the osmotic and matric potentials ($\psi_w = \psi_m + \psi_s$). The soil–air interface acts as a diffusion barrier in allowing only water molecules to pass from liquid to vapor state.

Understanding measurements of soil water potential using a determination of the soil air-phase humidity can be enhanced by a psychrometric chart because one can observe relationships among the different measurement techniques. A psychrometric chart depicts the physical properties of moist air computed for a fixed pressure or elevation. Thermophysical properties depicted in the chart may include dry-bulb and wet-bulb temperature, dew point temperature, vapor pressure, RH, humidity ratio, specific enthalpy, and specific volume (inverse air density). Measurements commonly carried out in soil access dry- and wet-bulb temperature using thermocouple psychrometry or dew point temperature using chilled mirror techniques, where matric potential is often related to the soil vapor pressure. In the following, we briefly describe the psychrometric chart and the interrelations among different thermophysical properties and then review the associated measurement techniques.

The psychrometric chart depicted in Figure 4.14 requires a pair of known or measured values in order to determine the unknown properties for a given condition. Typically dry-bulb (air) temperature is measured along with wet-bulb temperature. For example, assuming a thermocouple psychrometer measured a dry-bulb temperature of 25°C and a wet-bulb temperature of 23°C, the resulting RH according to this chart is 85%. This value is computed from the ratio of the vapor pressure to the saturation vapor pressure at 25°C, which are 20.2 and 24.8 mm Hg, respectively. The corresponding dew point temperature is observed to be 22.1°C.

4.3.4.4 Psychrometry for Measuring Soil Water Potential

A psychrometer measures the difference between dry-bulb and wet-bulb temperatures to determine the RH. The dry

FIGURE 4.14 Psychrometric chart showing relationships between dry-bulb (T_{db}), wet-bulb (T_{wb}), and dew point (T_{dp}) temperatures in addition to water vapor pressure (e) and saturation vapor pressure (e_o) for an elevation of 1500 m (85 kPa). In this example, a psychrometer measures dry-bulb (T_{db}) and wet-bulb (T_{wb}) temperatures of 25°C and 23°C, respectively, indicating a RH of 85%. From the chart, a corresponding dew point temperature (T_{dp}) is 22.1°C and the vapor pressure of 20.2 mm Hg and a saturation vapor pressure of 24.8 mm Hg (note: RH = e/e_o).

bulb is the temperature of the ambient air (nonevaporating surface), and the wet bulb is the temperature of an evaporating surface (generally lower than the dry-bulb temperature). The RH determines the rate of evaporation from the wet-bulb junction, and thus the extent of temperature depression below ambient.

A thermocouple psychrometer consists of a fine-wire chromel–constantan or other bimetallic thermocouple. A thermocouple is a double junction of two dissimilar metals. When the two junctions are subjected to different temperatures, they generate a voltage difference (Seebeck effect). Conversely, when an electrical current is applied through the junctions, it creates a temperature difference between the junctions by heating one while cooling the other, depending on the current's direction. For typical soil use, one junction of the thermocouple psychrometer is suspended in a thin-walled porous ceramic or stainless screen cup buried in the soil (Figure 4.15a), while another is embedded in an insulated plug to measure the ambient temperature at the same location. In psychrometric mode, the suspended thermocouple is cooled below the dew point by means of an electrical current (Peltier cooling) until pure water condenses on the junction. The cooling current then stops, and as water evaporates, it draws heat from the junction (heat of vaporization), depressing it below the temperature of the surrounding air until it attains a wet-bulb temperature. The warmer and drier the surrounding air, the higher the evaporation rate and the greater the wet-bulb depression. The difference in temperatures between the dry- and wet-bulb thermocouples

is measured and used to infer the RH (or relative vapor pressure) using the psychrometer equation:

$$\frac{e}{e_o} = 1 - \left[\frac{s + \gamma}{e_o}\right]\Delta T, \qquad (4.31)$$

where

s is the slope of the saturation water vapor pressure curve ($s = de_o/dT$)

γ is the psychrometric constant (about 0.067 kPa K^{-1} at 20°C)

ΔT is the temperature difference (K)

The slope (s) and e_o are functions of temperature only and can be approximated by closed-form expressions (Brutsaert, 1982).

A typical temperature depression measurable by a good psychrometer is on the order of 0.000085 C kPa^{-1}. This means that any errors in measuring wet-bulb depression can introduce large errors into psychrometric determinations. Thermal equilibrium is, therefore, a prerequisite to obtaining reliable readings, as any temperature difference between wet and dry sensors resulting from thermal gradients will be (erroneously) incorporated into the RH calculation.

Summarizing, psychrometric measurements of soil water potential are based on equilibrium between liquid soil water and water vapor in the ambient soil atmosphere. The drier the soil, the fewer water molecules escape into the ambient atmosphere resulting in lower RH (lower vapor pressure). When the

Measurement chamber overturned

Sample holder

Measurement chamber

WP4-T

Dry bulb junction

Wet bulb junction

5 mm

3 mm

Ceramic shield or screen

Peltier cooler

Chilling mirror

Detector

Circulation fan

Thermopile

Infrared temp.

(a) Thermocouple psychrometer

(b) Dewpoint potentiometer
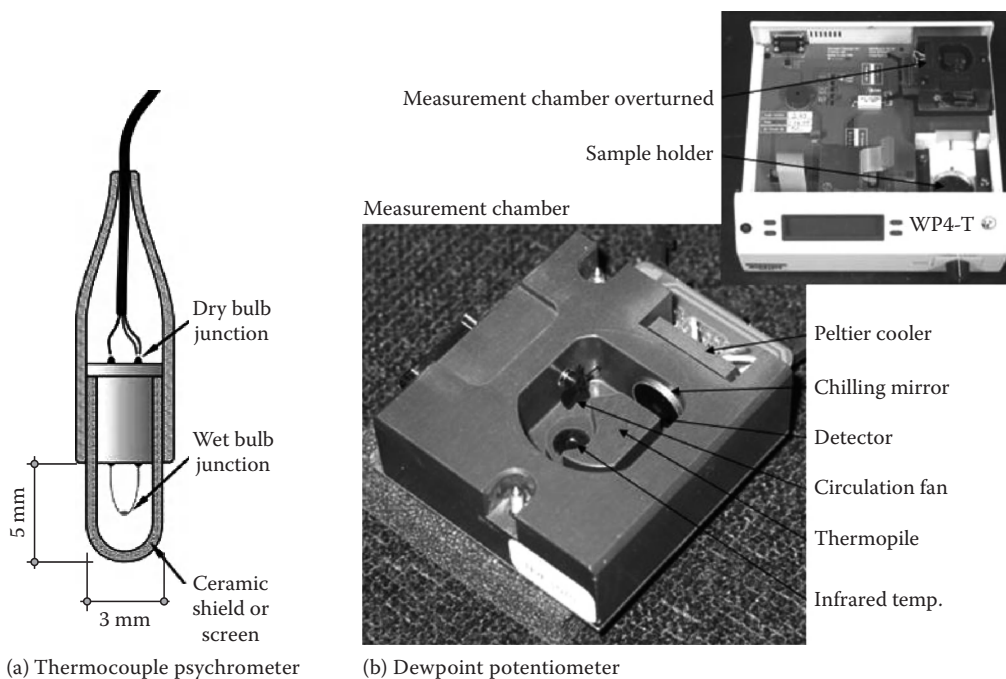
FIGURE 4.15 A field psychrometer (a) with porous ceramic shield. (Wescor, Inc. Logan, UT. With permission.) Dew point potentiometer (b) (Decagon Devices, Inc.) showing the internal components providing soil sample temperature measurement (dry bulb in Figure 4.14) and the chilled mirror measurement for dew point temperature determination. The system provides soil water potential calculations from measurements in the range of −0.1 down to −300 MPa.

osmotic potential is negligible, the soil water potential measured by a psychrometer is nearly equal to the soil matric potential. In principle, soil psychrometers may be buried in the soil and left for long periods, although corrosion is a problem in some environments.

4.3.4.5 Dew Point Potentiometer for Measuring Soil Water Potential

In a cooling process, the temperature at which water vapor condenses into water is known as the dew point temperature (T_{dp}) and occurs near 100% RH. This temperature is illustrated in Figure 4.15b, where T_{dp} is 22.1°C corresponding with the dry- and wet-bulb temperature intersection point in the previous example. Note that the dew point temperature is also found by drawing a horizontal line to the left intersecting the 100% humidity curve and dropping vertically to the corresponding air (dry bulb) temperature. In Figure 4.15b, a detailed measurement system from a Decagon WP4-T illustrates the electronics and hardware, which measure the soil water potential using the chilled mirror dew point technique.

4.3.4.6 Reference Porous Matrix Properties for Measuring Matric Potential

The interdependence of soil matric potential and water content facilitates indirect methods for inferring the energy of soil water from water content-dependent physical property measurements where a well-characterized and repeatable porous matrix can be produced. Measurement approaches have included thermal, electrical, dielectric measurements within porous media including plaster of Paris, ceramics, fiberglass, porous plastics, and natural soils (Wraith and Or, 2001). The measurable physical property (heat capacity, thermal conductivity, electrical resistivity, dielectric constant) of a porous matrix is affected by its water content and hence related to its matric potential. With the reference matrix sensor buried in the soil, changes in soil matric potential result in a gradient between the soil and the reference matrix that induces a water flux between the two materials until a new equilibrium potential is established. Reference matrix sensor physical properties being measured are influenced by these water content changes. In this manner, the measured sensor response yields quasi-direct information on the soil matric potential and indirect information on soil wetness where a water retention relationship is known. A common challenge of these sensors is quality mass production of the reference porous medium whose θ–ψ relationship exhibits low variance and whose hydraulic properties are compatible with the soil being measured. Accuracy, repeatability, and spatial resolution of specific sensors or methods are important considerations to their potential applications and in the analysis of soil matric potential measurements.

Heat dissipation sensors contain line- or point-source heating elements embedded in a rigid porous matrix with fixed pore space. The measurement is based on applying a heat pulse by passing a constant current through the heating element for a specified time and analyzing the temperature response

measured by a thermocouple fixed at a known distance from the heating source (Phene et al., 1971; Bristow et al., 1993). Sensors are individually or uniformly calibrated in terms of heat dissipation versus sensor wetness (i.e., matric potential). A typical useful matric potential range for such sensors is −10 to −1000 kPa.

A similar line-source sensor with a fine-wire heating element axially centered in a cylindrical ceramic matrix having a radius of 1.5 cm and length of 4.2 cm is depicted in Figure 4.16. A thermocouple is located adjacent to the heating element at mid-length. Both the heating wire and the thermocouple are contained in the shaft portion of a hollow needle. Because the thermocouple is located adjacent to the heating element, as the soil dries and water moves out of the ceramic, the magnitude of temperature change during a given period under constant heating current and duration will increase due to the reduced thermal conductivity of the porous matrix. The magnitude of the measured temperature increase and/or decrease is often linearly related to the natural logarithm of matric potential. The accuracy and operational limits of these sensors were evaluated by Reece (1996), who also presented an improved calibration method for between-sensor variability, based on normalizing sensor readings by oven dry sensor thermal conductivity. A primary advantage of these sensors is their wide range of applicability from about −0.1 to −12.0 MPa.

Electrical resistance sensors have been used in soils for matric potential determination since the work of Bouyoucos and Mick (1940) and continue to be developed to this day. Ongoing efforts seek to overcome limitations associated with temperature and salinity dependence on the resistance measurement (see Section 4.2.2.7). The measurement method is popular because sensors are inexpensive to produce. Measuring the dielectric of a reference porous medium can be used to infer soil matric potential. Early studies used TDR probes embedded in plaster of Paris or ceramic with a well-characterized θ–ψ relationship in the range of −30 to −1000 kPa (Noborio et al., 1999). Commercially available dielectric-based sensors provide a narrower measurement range of −10 to −500 kPa but are less expensive to produce. Some limitations of this method include possible effects from salinity
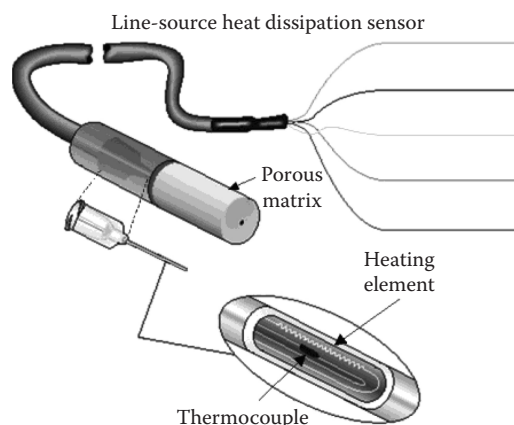
FIGURE 4.16 Schematic of CSI 229 heat dissipation sensor. (From Campbell Scientific, Inc., Logan, UT.)

on the dielectric measurement, especially at measurement frequencies below 100 MHz, and an altered $\theta-\kappa$ relationship occurring in frozen soils.

4.4 Soil Water Content—Energy Relationships

4.4.1 Soil Water Characteristic

An SWC curve describes the functional relationship between soil water content (θ_m or θ_v) and matric potential under equilibrium conditions. The SWC is an important soil property related to the distribution of pore space (sizes, interconnectedness), which is strongly affected by texture and structure, as well as related factors including organic matter content. The SWC is a primary hydraulic property required for modeling water flow, for irrigation management, and for many additional applications related to managing or predicting water behavior in the porous system. An SWC is a highly nonlinear function and is relatively difficult to obtain accurately. Because the matric potential extends over several orders of magnitude for the range of water contents commonly encountered in practical applications, the matric potential is often plotted on a logarithmic scale. Several SWC curves for soils of different textures demonstrating the effects on porosity (saturated water content) and on the slope of the relationships resulting from variable pore-size distributions are depicted in Figure 4.17.

4.4.2 Measurement of SWC Relationships

Several methods are available to obtain the measurements needed for SWC estimation. The basic requirement is for pairs of $\psi_m-\theta$ measurements over the wetness range of interest. Among the primary experimental problems in determining an SWC are (1) the limited functional range of the tensiometer, which is often used for in situ measurements, (2) inaccurate θ measurements in

some cases, (3) the difficulty in obtaining undisturbed samples for laboratory determinations, and (4) a slow rate of equilibrium under low matric potential (i.e., dry soils).

In situ methods are considered the most representative for determining SWCs, particularly when a wide range of $\psi_m-\theta_v$ values are obtained. An effective method to obtain simultaneous measurements of ψ_m and θ_v utilizes TDR probes installed in the soil at close proximity to transducer tensiometers and/or other matric sensors, with the changing values of each attribute monitored through time as the soil wetness varies. Large changes in ψ_m and θ_v can be induced under highly evaporative conditions near the soil surface or in the presence of active plant roots.

4.4.2.1 Laboratory Estimation Using Pressure Plate and Pressure Flow Cell

The pressure plate apparatus consists of a pressure chamber enclosing a water-saturated porous plate, which allows water but prevents air flow through its pores (Figure 4.18). The porous plate is open to atmospheric pressure at the bottom, while the top surface is at the applied pressure of the chamber. Sieved soil samples (usually <2 mm) are placed in retaining rubber rings in contact with the porous plate and left to saturate in water. After saturation is attained, the porous plate with the saturated soil samples is placed in the chamber and a known gas (commonly N_2 or air) pressure is applied to force water out of the soil and through the plate. Flow continues until equilibrium between the force exerted by the air pressure and the force by which soil water is being held by the soil (ψ_m) is reached.

Soil water retention in the low suction range of 0–10 m (=1 bar = 0.1 MPa) is strongly influenced by soil structure and its natural pore-size distribution. Hence, undisturbed intact soil samples (cores) are preferred over repacked samples for the wet end of the SWC. The pressure flow cell (Tempe cell) can hold intact soil samples encased in metal rings. The operation of the cell follows that of the pressure plate, except the pressure range is usually lower (0–10 m). The porous ceramic plates for both the pressure plate and the flow cell must be completely saturated, a process which may take a few days to achieve. Following equilibrium between soil matric potential and the applied air pressure, the soil samples are removed from the apparatus, weighed, and oven dried for gravimetric determination of water content. An estimate of the soil sample bulk density must be provided to convert θ_m to θ_v in the case of disturbed samples. Because of differences in pore sizes and geometry, the water content of repacked soils at a given matric potential should not be used to accurately infer θ of intact soils at the same ψ_m.

Several pressure steps may be applied to the same samples when using flow cells. The cells may be disconnected from the air pressure source and weighed to determine the change in water content from the previous step, then reconnected to the air pressure and a new (greater) pressure step applied. Water outflow from the cells may also be monitored and related by volume or mass to the change in sample water content.
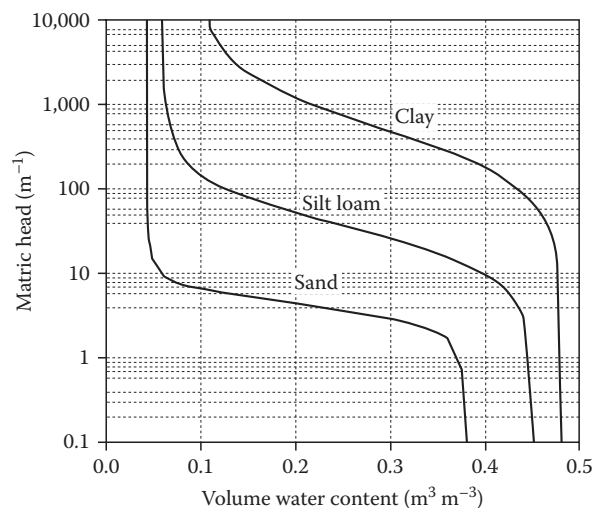
FIGURE 4.17 Example of soil water retention relationships for three soil textures.
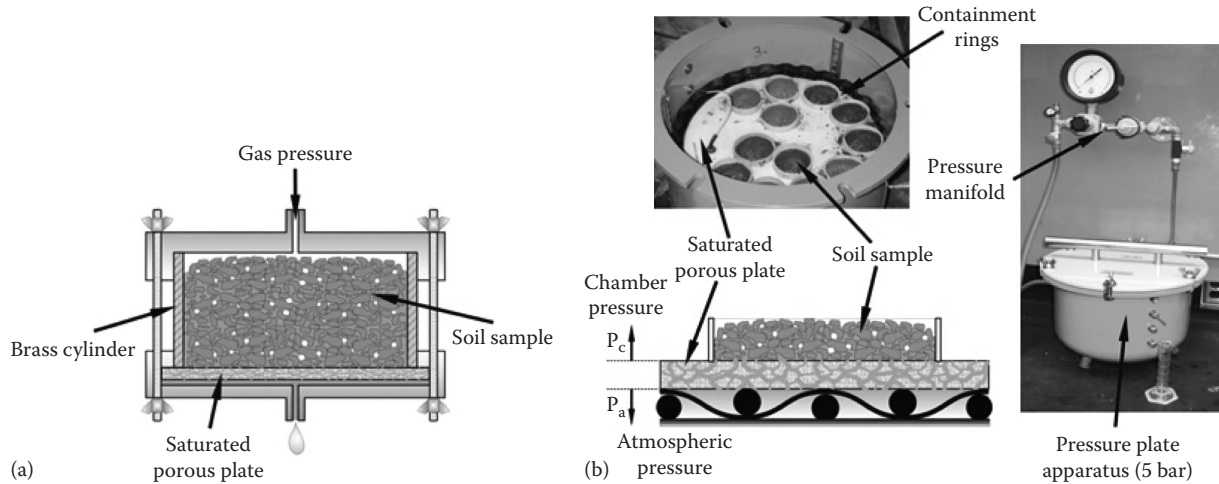
FIGURE 4.18 (a) Pressure flow cell (Tempe cell) and (b) pressure plate apparatus used to desaturate soil samples to desired matric potential.

4.4.2.2 Field Measurement Methods—Sensor Pairing

Despite the paramount importance of SWC determination in situ, suitable measurement techniques are severely lacking at present. The most common approach is to use paired sensors such as neutron moisture meter access tubes or TDR waveguides and tensiometers to determine water content and matric potential simultaneously and in the same soil volume. Single probes that combine TDR with tensiometry have also been proposed (Baumgartner et al., 1994). The limitations of most sensor pairing techniques stem from (1) differences in the soil volumes sampled by each sensor (e.g., large volume averaging by a neutron moisture meter vs. a small volume sensed by a heat dissipation sensor or a psychrometer), (2) the time required for matric potential sensors to reach equilibrium, and (3) limited ranges and deteriorating accuracy of different sensor pairs. This often results in limited overlap in SWC information measured using different techniques and problems with measurement errors within the range of overlap.

A summary of common methods available for matric potential measurement or inference and their range of application is presented in Figure 4.19. Note that most of the available techniques have a limited range of overlap (or do not overlap at all), and most are laboratory methods not suitable for field applications.

4.4.3 Fitting Parametric SWC Expressions to Measured Data

Measuring an SWC is laborious and time consuming. Measured θ–ψ pairs are often fragmentary, and usually constitute relatively few measurements over the wetness range of interest. For modeling and analysis purposes and for characterization and comparison between different soils and scenarios, it is therefore beneficial to represent the SWC in a continuous and parametric form (Fredlund and Xing, 1994). A parametric expression of an SWC model should (1) contain as few parameters as necessary to simplify its estimation and (2) describe the behavior of the SWC

FIGURE 4.19 Typical ranges of application for some common matric potential measurement or inference methods.

at the limits (wet and dry ends) while closely fitting the nonlinear shape of $\theta(\psi_m)$ data.

An effective and commonly used parametric model for relating water content to the matric potential was proposed by van Genuchten (1980) and is denoted as VG:

$$\Theta = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \left[\frac{1}{1 + (\alpha \, | \, \psi_m \, |)^n} \right]^m, \qquad (4.32)$$

where

θ_r and θ_s are the residual and saturated water contents, respectively

α, n, and m are parameters directly dependent on the shape of the $\theta(\psi)$ curve

A considerable simplification is gained by assuming that m = 1 − 1/n. Thus, the parameters required for estimation of the model are θ_r, θ_s, α, and n. θ_s is usually known and is easy to
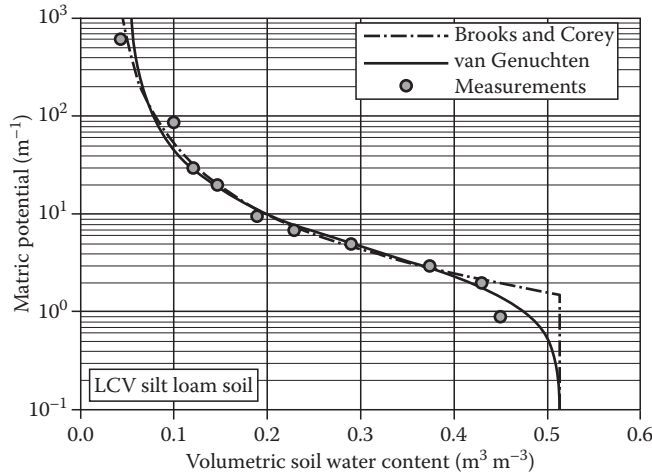
FIGURE 4.20 VG and BC parametric models fitted to measured water retention data.

obtain experimentally with good accuracy, leaving only three unknown parameters (θ_r, α, and n) to be estimated from the experimental data in many cases. Note that θ_r may be taken as $\theta_{-1.5\,MPa}$, $\theta_{air\,dry}$, or a similar value, though it is often advantageous to use it as a fitting parameter.

Another well-known parametric model was proposed by Brooks and Corey (1964) and is denoted as BC:

$$\Theta = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \left[\frac{\psi_b}{\psi_m}\right]^\lambda \quad \psi_m > \psi_b$$
$$\Theta = 1 \quad \psi_m \leq \psi_b, \tag{4.33}$$

where

ψ_b is a parameter related to the soil matric potential at air entry (subscript b represents bubbling pressure)
λ is related to the soil pore-size distribution

Matric potentials are expressed as positive quantities in both VG and BC parametric expressions.

Estimation of VG or BC parameters from experimental data requires (1) sufficient data points (at least 5–8 $\theta(\psi_m)$ pairs) and (2) a program for performing nonlinear regression. Recent versions of many computer spreadsheets provide relatively simple and effective mechanisms for performing nonlinear regression. Details of the computational steps required for fitting an SWC to experimental data using commercially available spreadsheet software are given in Wraith and Or (1998). In addition, computer programs for estimation of specific parametric models are also available such as the RETC code (van Genuchten et al., 1991).

Fitted parametric models of VG and BC to silt loam $\theta(\psi)$ data measured by Or et al. (1991) are presented in Figure 4.20. The resulting best fit parameters for the VG model are $\alpha = 0.417\,m^{-1}$; n = 1.75; $\theta_s = 0.513\,m^3\,m^{-3}$; and $\theta_r = 0.05\,m^3\,m^{-3}$ (with $r^2 = 0.99$). For the BC model, the best fit parameters are $\lambda = 0.54$; $\psi_b = 1.48\,m$; $\theta_s = 0.513\,m^3\,m^{-3}$; and $\theta_r = 0.03\,m^3\,m^{-3}$ (with $r^2 = 0.98$). Note that the most striking difference between the VG and the BC models is in the discontinuity at $\psi = \psi_b$ for BC.

The water retention character of multiporosity porous media can be described by superimposing multiple SWC curves (Durner, 1994) on measured data such as those illustrated in Figure 4.21 using the following expression.

$$\Theta = \sum_{i=1}^{k} w_i \left[\frac{1}{1 + (\alpha_i |\psi_m|)^{n_i}}\right]^{m_i}, \tag{4.34}$$

where

i is an index
k is the number of SWC curves to be fitted with weighting, w_i
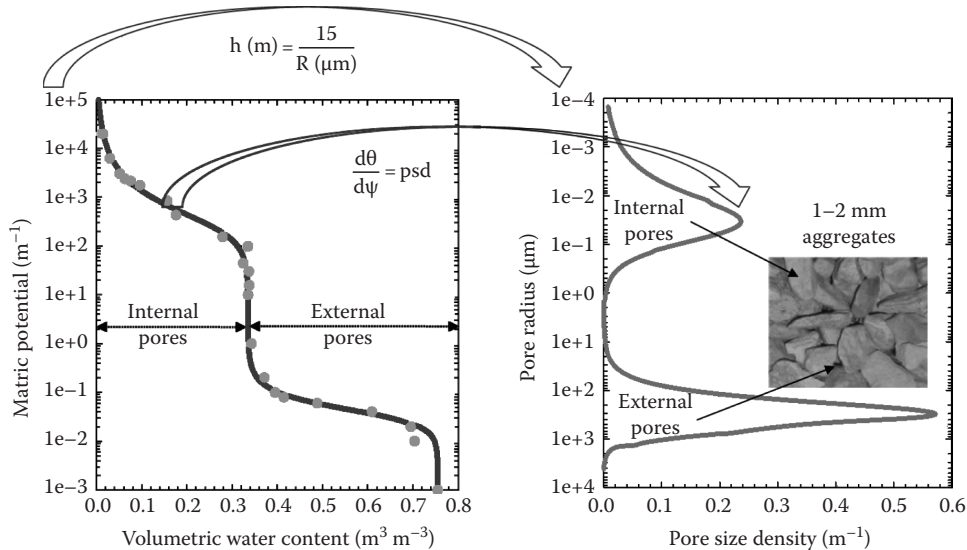
FIGURE 4.21 SWC of dual-porosity aggregates (1–2 mm) mapped to a corresponding pore-size distribution using Equation 4.25.

In this stable ceramic aggregate, the large disparity between internal and external pore domains results in a four-order-of-magnitude difference in the mean pore diameters mapped using Equation 4.26. The pore-size density (psd) obtained from the slope of the SWC ($d\theta/d\psi$) is informative for surface area estimation and understanding transport phenomena. Sources of additional experimental and parametric SWC information include (1) the Unsaturated Soil Hydraulic Database (UNSODA) (Leij et al., 1996), which is a computer database compiled by the US Salinity Laboratory, which contains an exhaustive collection of retention (SWC) and unsaturated hydraulic conductivity information for soils of different textures from around the world. While the authors/compilers have attempted to provide some indices of data quality or reliability, the users are advised (as always) to use their own experience and discretion in adapting others' data to their own applications; and (2) the regression studies by McCuen et al. (1981) and Rawls and Brakensiek (1989) provide a wealth of information on the BC parameter values for many soils including estimation of the hydraulic parameters based on other, often more easily available, soil properties. These estimates may be sufficiently accurate for some applications and could be used to obtain first-order approximations.

4.4.4 Hysteresis in the Soil–Water Characteristic Relation

Water content and the potential energy of soil water are not uniquely related because the amount of water present at a given matric potential is dependent on the pore-size distribution and the properties of air–water–solid interfaces. A $\theta(\psi)$ relationship may be obtained by (1) taking an initially saturated sample and applying suction or pressure to desaturate it (desorption) or by (2) gradually wetting an initially dry soil (sorption). These two pathways produce curves that in most cases are not identical; the water content in the drying curve is higher for a given matric potential than that in the wetting branch (Figure 4.22). This is called hysteresis, defined as the phenomenon exhibited by a system in which the reaction of the system to changes is dependent upon its past reactions to change.

The hysteresis in SWC can be related to several phenomena: (1) the ink bottle effect resulting from nonuniformity in shape and sizes of interconnected pores as illustrated in Figure 4.23a; drainage is governed by the smaller pore radius r, whereas wetting is dependent on the larger radius R; (2) different liquid–solid contact angles for advancing and receding water menisci (Figure 4.23b); (3) entrapped air in a newly wetted soil (e.g., pore doublet) (Dullien, 1992); and (4) swelling and shrinking of the soil under wetting and drying, which may alter the porosity and pore-size distribution. Based on early observations of the phenomenon by Haines (1930) as well as present day theories (Mualem, 1984; Kool and Parker, 1987), the role of individual factors remains unclear and is subject to ongoing research.
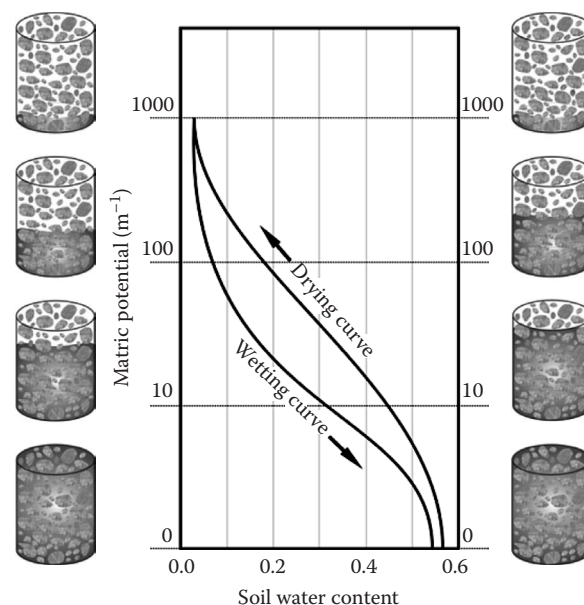
FIGURE 4.22 Concept of hysteresis in soil–water characteristic relationships.
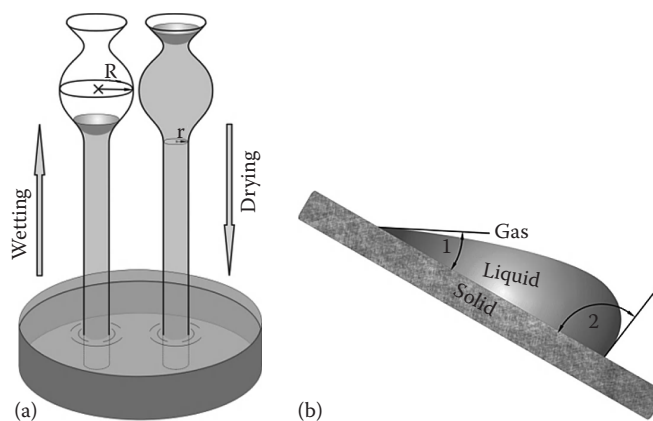
FIGURE 4.23 "Ink bottle" effect (a) and the contact angle effect (b).

4.5 Resources

A number of resources are available for readers interested in additional insight or information concerning soil water content and energy. The following suggestions are by no means inclusive but should serve augmenting and extending the discussions presented here, as well as providing a source of additional references. We advise readers to consult the references we have cited in relevant sections of the chapter, as well as various texts, monographs, and review chapters. Soil physics and related textbooks, which address soil water content and water energy include those of Childs (1969), Hanks (1992), Hillel (1998), Kirkham and Powers (1972), Kutilek and Nielsen (1994), Marshall et al. (1996), Hillel (1998), and Jury and Horton (2004). Several chapters in the monograph edited by Dane and Topp (2002), as well as that edited by Carter (1993), contain valuable information. A number

of excellent review papers or chapters are also available, including Baker (1990), Campbell and Mulla (1990), and Parker (1986).

Some Internet sites contain valuable information related to soil water content and energy. These include various individual or professional organization pages, discussion groups (e.g., the soil moisture group SOWACS), and others. However, Internet addresses tend to change rather frequently, and new sources may be added. This means it is generally more efficient to conduct keyword searches using one or more of the available search engines than to rely on potentially outdated web addresses. Extensive geospatial digital soil databases are being established such as the USDA Web Soil Survey and websites detailing worldwide resources can be found online (e.g., compendium of online soil survey), In addition, there are numerous numerical modeling tools available online, which are useful for simulating soil water dynamics (e.g., Hydrus 1-D).

Acknowledgments

References

Abdu, H., D.A. Robinson, M. Seyfried, and S.B. Jones. 2008. Geophysical imaging of watershed subsurface patterns and prediction of soil texture and water holding capacity. Water Resour. Res. 44:W00D18.

Adamo, F., G. Andria, F. Attivissimo, and N. Giaquinto. 2004. An acoustic method for soil moisture measurement. IEEE Trans. Inst. Meas. 53:891–898.

Adamson, A.W. 1990. The solid–liquid interface-contact angle, p. 379–420. *In* Physical chemistry of surfaces. 5th edn. John Wiley & Sons, New York.

Akyildiz, I.F., W. Su, Y. Sankarasubramaniam, and E. Cayirci. 2002. Wireless sensor networks: A survey. Comput. Netw. 38:393–422.

Amidu, S.A., and J.A. Dunbar. 2007. Geoelectric studies of seasonal wetting and drying of a Texas vertisol. Vadose Zone J. 6:511–523.

Annan, A.P. 2005a. GPR methods for hydrogeological studies, p. 185–213. *In* Y. Rubin and S. Hubbard (eds.) Hydrogeophysics, water science and technology library. Vol. 50. Springer, Dordrecht, the Netherlands.

Annan, A.P. 2005b. Ground-penetrating radar, p. 357–438. *In* D.K. Butler (ed.) Near surface geophysics, investigations in geophysics. Vol. 13. No. 4. Society of Exploration Geophysicists, Tulsa, OK.

Babb, A.T.S. 1951. A radio-frequency electronic moisture meter. Analyst 76:428–433.

Baker, J.M. 1990. Measurement of soil water content. Remote Sens. Rev. 5:263–279.

Baker, J.M., and R.R. Allmaras. 1990. System for automating and multiplexing soil moisture measurement by time-domain reflectometry. Soil Sci. Soc. Am. J. 54:1–6.

Bakker, G., M.J. van der Ploeg, G.H. de Rooij, C.W. Hoogendam, H.P.A. Gooren, C. Huiskes, L.K. Koopal, and H. Kruidhof. 2007. New polymer tensiometers: Measuring matric pressures down to the wilting point. Vadose Zone J. 6:196–202.

Banin A., and A. Amiel. 1970. A correlative study of the chemical and physical properties of a group of natural soils of Israel. Geoderma 3:185–198.

Basinger, J.M., G.J. Kluitenberg, J.M. Ham, J.M. Frank, P.L. Barnes, and M.B. Kirkham. 2003. Laboratory evaluation of the dual-probe heat-pulse method for measuring soil water content. Vadose Zone J. 2:389–399.

Baumgartner, N., G.W. Parkin, and D.E. Elrick. 1994. Soil water content and potential measured by hollow time domain reflectometry probe. Soil Sci. Soc. Am. J. 58:315–318.

Bird, N.R.A., A.R. Preston, E.W. Randall, W.R. Whalley, and A.P. Whitmore. 2005. Measurement of the size distribution of water-filled pores at different matric potentials by stray field nuclear magnetic resonance. Eur. J. Soil Sci. 56:135–143.

Bloch, F. 1953. The principle of nuclear induction. Science 118:425–430.

Blonquist, J.M., Jr., S.B. Jones, I. Lebron, and D.A. Robinson. 2006. Microstructural and phase configurational effects determining water content: Dielectric relationships of aggregated porous media. Water Resour. Res. 42:WO5424. doi:10.1029/2005WR004418.

Blonquist, J.M., Jr., S.B. Jones, and D.A. Robinson. 2005a. Standardizing characterization of electromagnetic water content sensors: Part 2. Evaluation of seven sensing systems. Vadose Zone J. 4:1059–1069.

Blonquist, J.M., Jr., S.B. Jones, and D.A. Robinson. 2005b. A time domain transmission sensor with TDR performance characteristics. J. Hydrol. 314:235–245.

Blum, A., I. Flammer, T. Friedli, and P. Germann. 2004. Acoustic tomography applied to water flow in unsaturated soils. Vadose Zone J. 3:288–299.

Bolt, G.H. 1976. Soil physics terminology. Int. Soc. Soil Sci. Bull. 49:16–22.

Bottomley, P.A., H.H. Rogers, and T.H. Foster. 1986. NMR imaging shows water distribution and transport in plant root systems in situ. Proc. Natl Acad. Sci. U S A 83:87–89.

Bouyoucos, G.J., and A.H. Mick. 1940. An electrical resistance method for the continuous measurement of soil moisture under field conditions. MI Agr. Exp. Stn. Tech. Bull. 172:1–38.

Bristow, K.L. 1998. Measurement of thermal properties and water content of unsaturated sandy soil using dual-probe heat-pulse probes. Agr. Forest Meteorol. 89:75–84.

Bristow, K.L., G.S. Campbell, and K. Calissendroff. 1993. Test of a heat-pulse probe for measuring changes in soil water content. Soil Sci. Soc. Am. J. 57:930–934.

Brooks, R.H., and A.T. Corey. 1964. Hydraulic properties of porous media. Hydrology Paper 3. Colorado State University, Fort Collins, CO.

Brutsaert, W. 1982. Evaporation into the atmosphere. D. Reidel Publishing Company, Dordrecht, the Netherlands.

Brutsaert, W., and J.N. Luthin. 1964. The velocity of sound in soils near the surface as a function of moisture content. J. Geophys. Res. 69:643–652.

Campbell, G.S. 1985. Soil physics with basic. Elsevier, New York.

Campbell, J.E. 1990. Dielectric properties and influence of conductivity in soils at one to fifty megahertz. Soil Sci. Soc. Am. J. 54:332–341.

Campbell, G.S., C. Calissendorff, and J.H. Williams. 1991. Probe for measuring soil specific heat using a heat-pulse method. Soil Sci. Soc. Am. J. 55:291–293.

Campbell, G.S., and D.J. Mulla. 1990. Measurement of soil water content and potential, p. 127–142. *In* B.A. Stewart and D.R. Nielsen (eds.) Irrigation of agricultural crops. ASA, Madison, WI.

Cardell-Oliver R., M. Kranz, K. Smettem, and K. Mayer. 2005. A reactive soil moisture sensor network: Design and field evaluation. Int. J. Distrib. Sens. Netw. 1:149–162.

Carter, M.R. (ed.). 1993. Soil sampling and methods of analysis. Lewis Publishers, Boca Raton, FL.

Chen, Y.P., and D. Or. 2006. Geometrical factors and interfacial processes affecting complex dielectric permittivity of partially saturated porous media. Water Resour. Res. 42:W06423.

Childs, E.C. 1969. An introduction to the physical basis of soil water phenomena. John Wiley & Sons Ltd., London, U.K.

Corey, A.T., and A. Klute. 1985. Application of the potential concept to soil water equilibrium and transport. Soil Sci. Soc. Am. J. 49:3–11.

Cosenza, P., C. Camerlynck, and A. Tabbagh. 2003. Differential effective medium schemes for investigating the relationship between high-frequency relative dielectric permittivity and water content of soils. Water Resour. Res. 39, 9:1230. doi:10.1029/2002WR001774.

Dahiya, I.S., D.J. Dahiya, M.S. Kuhad, and P.S. Karwasra. 1988. Statistical equations for estimating field capacity, wilting point, and available water capacity of soils from their saturation percentage. J. Agr. Res. 110:515–520.

Dane, J.H., and G.C. Topp. 2002. Methods of soils analysis: Part 4, physical methods. SSSA, Madison, WI.

Day, P.R., G.H. Bolt, and D.M. Anderson. 1967. Nature of soil water, p. 193–208. *In* R.M. Hagan, H.R. Haise, and T.W. Edminster (eds.) Irrigation of agricultural lands. ASA, Madison, WI.

Dean, T.J., J.P. Bell, and A.J.B. Baty. 1987. Soil moisture measurement by an improved capacitance technique. Part I. Sensor design and performance. J. Hydrol. 93:67–78.

De Loor, G.P. 1968. Dielectric properties of heterogeneous mixtures containing water. J. Microw. Power 3:67–73.

De Vries, D.A. 1963. Thermal properties of soils, p. 210–235. *In* W.R. van Wijk (ed.) Physics of plant environment. North Holland Publishing Co., Amsterdam, the Netherlands.

Dirksen, C., and S. Dasberg. 1993. Improved calibration of time domain reflectometry soil water content measurements. Soil Sci. Soc. Am. J. 57:660–667.

Dobson, M.C., F.T. Ulaby, M.T. Hallikainen, and M.A. El-Rayes. 1985. Microwave dielectric behavior of wet soil: II. Dielectric mixing models. IEEE Trans. Geosci. Remote Sens. GE-23:35–46.

Dullien, F.A.L. 1992. Porous media: Fluid transport and pore structure. 2nd edn. Academic Press, New York.

Dunne, T., and R.D. Black. 1970. An experimental investigation of runoff production in permeable soils. Water Resour. Res. 6:478–490.

Durner, W. 1994. Hydraulic conductivity estimation for soils with heterogeneous pore structure. Water Resour. Res. 30:211–223.

Entekhabi, D., E.G. Njoku, P. Houser, M. Spencer, T. Doiron, Y.J. Kim, J. Smith et al. 2004. The hydrosphere state (hydros) satellite mission: An earth system pathfinder for global mapping of soil moisture and land freeze/thaw. IEEE Trans. Geosci. Remote Sens. 42:2184–2195.

Entekhabi, D., I. Rodriguez-Iturbe, and F. Castelli. 1996. Mutual interaction of soil moisture state and atmospheric processes. J. Hydrol. 184:3–17.

Evett, S.R., and J.L. Steiner. 1995. Precision of neutron scattering and capacitance type soil water content gauges from field calibration. Soil Sci. Soc. Am. J. 59:961–968.

Fares, A., and V. Polyakov. 2006. Advances in crop water management using capacitive sensors. Adv. Agron. 90:43–77.

Ferre, P.T.A., J.H. Knight, D.L. Rudolph, and R.G. Kachanoski. 1998. The sample areas of conventional and alternative time domain reflectometry probes. Water Resour. Res. 34:2971–2979.

Fredlund, D.G., and A. Xing. 1994. Equations for the soil-water characteristic curve. Can. Geotech. J. 31:521–532.

Friedman, S.P. 1998. A saturation degree-dependent composite spheres model for describing the effective dielectric constant of unsaturated porous media. Water Resour. Res. 34:2949–2961.

Friedman, S.P., and D.A. Robinson. 2002. Particle shape characterization using angle of repose measurements for predicting the effective permittivity and electrical conductivity of saturated granular media. Water Resour. Res. 38:1236.

Gardner, W.H. 1986. Water content, p. 493–544. *In* A. Klute (ed.) Methods of soil analysis. Part 1. 2nd edn. ASA, Madison, WI.

Gaskin, G.J., and J.D. Miller. 1996. Measurement of soil water content using a simplified impedance measuring technique. J. Agr. Eng. Res. 63:153–159.

Grayson, R.B., A.W. Western, F.H.S. Chiew, and G. Blöschl. 1997. Preferred states in spatial soil moisture patterns: Local and non-local controls. Water Resour. Res. 33:2897–2908.

Greacen, E.L., R.L. Correll, R.B. Cunningham, G.G. Johns, and K.D. Nicolls. 1981. Calibration, p. 50–81. *In* E.L. Greacen (ed.) Soil water assessment by the neutron method. CSIRO, Melbourne, Australia.

Greaves, R.J., D.P. Lesmes, J.M. Lee, and M.N. Toksoz. 1996. Velocity variations and water content estimated from multi-offset, ground-penetrating radar. Geophysics 61:683–695.

Gummerson, R.J., C. Hall, W.D. Hoff, R. Hawkes, G.N. Holland, and W.S. Moore. 1979. Unsaturated water flow within porous materials observed by NMR imaging. Nature 281:56–57.

Haines, W.B. 1930. Studies in the physical properties of soil. V. The hysteresis effect in capillary properties, and the modes of moisture distribution associated therewith. J. Agr. Sci. 20:97–116.

Hallikainen, M.T., F.T. Ulaby, M.C. Dobson, M.A. El-Rayes, and L.K. Wu. 1985. Microwave dielectric behavior of wet soil—Part I: Empirical models and experimental observations. IEEE Trans. Geosci. Remote Sens. GE-23:25–34.

Hanks, R.J. 1992. Applied soil physics. 2nd edn. Springer-Verlag, New York.

Heitman, J.L., J.M. Basinger, G.J. Kluitenberg, J.M. Ham, J.M. Frank, and P.L. Barnes. 2003. Field evaluation of the dual-probe heat pulse method for measuring soil water content. Vadose Zone J. 2:552–560.

Hillel, D. 1998. Environmental soil physics. Academic Press, San Diego, CA.

Hinedi, Z.R., D.B. Borchard, and A.C. Chang. 1993. Probing soil and aquifer material porosity with nuclear magnetic resonance. Water Resour. Res. 29:3861–3866.

Hoekstra, P., and A. Delaney. 1974. Dielectric properties of soils at UHF and microwave frequencies. J. Geophys. Res. 79:1699–1708.

Hubbard, S., K. Grote, and Y. Rubin. 2002. Mapping the soil volumetric water content of a California vineyard using high-frequency GPR ground wave data. Lead. Edge 21:552–559.

Hubbell, J.M., and J.B. Sisson. 1996. Portable tensiometer for use in deep boreholes. Soil Sci. 161:376–381.

Huisman, J.A., S.S. Hubbard, J.D. Redman, and A.P. Annan. 2003a. Measuring soil water content with ground penetrating radar: A review. Vadose Zone J. 2:476–491.

Huisman, J.A., J.J.J.C. Snepvangers, W. Bouten, and G.B.M. Heuvelink. 2003b. Monitoring temporal development of spatial soil water content variation: Comparison of ground penetrating radar and time domain reflectometry. Vadose Zone J. 2:519–529.

IAEA. 1970. Neutron moisture gauges. Technical Report Series No. 112. International Atomic Energy Agency, Vienna, Austria.

IAEA. 2008. Field estimation of soil water content: A practical guide to methods, instrumentation and sensor technology. Training Course Series No. 30. International Atomic Energy Agency, Vienna, Austria.

Idso, S.B., R.J. Reginato, R.D. Jackson, B.A. Kimball, and F.S. Nakayama. 1974. The three stages of drying in a field soil. Soil Sci. Soc. Am. Proc. 38:831–837.

Iwata, S., T. Tabuchi, and B.P. Warkentin. 1988. Soil water interactions. Marcel Dekker, New York.

Jackson, T.J., T.J. Schmugge, A.D. Nicke, G.A. Coleman, and E.T. Engman. 1981. Soil moisture updating and microwave remote sensing for hydrological simulation. Hydrol. Sci. Bull. 26:305–319.

Jenny, H. 1941. Factors of soil formation: A system of quantitative pedology. Dover Publications Inc., New York.

Jones, S.B., and S.P. Friedman. 2000. Particle shape effects on the effective permittivity of anisotropic or isotropic media consisting of aligned or randomly oriented ellipsoidal particles. Water Resour. Res. 36:2821–2833.

Jones, S.B., and D. Or. 2003. Modeled effects on permittivity measurements of water content in high surface area porous media. Physica B 338:284–290.

Jury, W.A., and R. Horton. 2004. Soil physics. John Wiley & Sons, New York.

Kachanoski, R.G., and E. de Jong. 1988. Scale dependence and the temporal persistence of spatial patterns of soil water storage. Water Resour. Res. 24:85–91.

Kelleners, T.J., D.A. Robinson, P.J. Shouse, J.E. Ayars, and T.H. Skaggs. 2005. Frequency dependence of the complex permittivity and its impact on dielectric sensor calibration in soils. Soil Sci. Soc. Am. J. 69:67–76.

Kelleners, T.J., R.W.O. Soppe, J.E. Ayers, and T.H. Skaggs. 2004. Calibration capacitance probe sensors in a saline silty clay soil. Soil Sci. Soc. Am. J. 68:770–778.

Kirkham, D., and W.L. Powers. 1972. Advanced soil physics. R.E. Krieger Publishing Company, Malabar, FL.

Kleinberg, R.L., and J.A. Jackson. 2001. An introduction to the history of NMR well logging. Concepts Magn. Reson. 13:340–342.

Kluitenberg, G.J. 2002. Heat capacity and specific heat, p. 1201–1208. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4—Physical methods. SSSA Book Series No. 5, SSSA, Madison, WI.

Kool, J.B., and J.C. Parker. 1987. Development and evaluation of closed-form expressions for hysteretic soil hydraulic properties. Water Resour. Res. 23:105–114.

Kutilek, M., and D.R. Nielsen. 1994. Soil hydrology. Catena Verlag, Cremlingen-Destedt, Germany.

Lambot, S., M. Antoine, M. Vanclooster, and E.C. Slob. 2006. Effect of soil roughness on the inversion of off-ground monostatic GPR signal for noninvasive quantification of soil properties. Water Resour. Res. 42:W03403.

Lambot, S., J. Rhebergen, I. van den Bosch, E.C. Slob, and M. Vanclooster. 2004. Measuring the soil water content profile of a sandy soil with an off-ground monostatic ground penetrating radar. Vadose Zone J. 3:1063–1071.

Laymon, C.A., W.L. Crosson, T.J. Jackson, A. Manu, and T.D. Tsegaye. 2001. Ground-based passive microwave remote sensing observations of soil moisture at S-band and L-band with insight into measurement accuracy. IEEE Trans. Geosci. Remote Sens. 39:1844–1858.

Leib, B.G., J.D. Jabro, and G.R. Matthews. 2003. Field evaluation and performance comparison of soil moisture sensors. Soil Sci. 168:396–408.

Leij, F.J., W.J. Alves, M.Th. van Genuchten, and J.R. Williams. 1996. The UNSODA unsaturated hydraulic database. EPA/600/R-96/095. USEPA, Cincinnati, OH.

Lu, Z. 2007. The phase shift method for studying nonlinear acoustics in a soil. Acta Acoust. 93:542–554.

Lunt, I.A., S.S. Hubbard, and Y. Rubin. 2005. Soil moisture content estimation using ground-penetrating radar reflection data. J. Hydrol. 307:254–269.

Mack, A.R., and E.J. Brach. 1966. Soil moisture measurement with ultrasonic energy. Soil Sci. Soc. Am. Proc. 30:544–548.

Marshall, T.J., J.W. Holmes, and C.W. Rose. 1996. Soil physics. 3rd edn. Cambridge University Press, Cambridge, U.K.

Martinez, K., J.K. Hart, and R. Ong. 2004. Environmental sensor networks. Computer 37:50–56.

McCuen, R.H., W.J. Rawls, and D.L. Brakensiek. 1981. Statistical analysis of the Brook–Corey and Green–Ampt parameters across soil texture. Water Resour. Res. 17:1005–1013.

McMichael, B., and R.J. Lascano. 2003. Laboratory evaluation of a commercial dielectric soil water sensor. Vadose Zone J. 2:650–654.

McNeill, J.D. 1980. Electromagnetic terrain conductivity measurement at low induction numbers. Tech Note TN-6, Geonics Ltd., Ontario, Canada.

Michot, D., Y. Benderitter, A. Dorigny, B. Nicoullaud, D. King, and A. Tabbagh. 2003. Spatial and temporal monitoring of soil water content with an irrigated corn crop cover using surface electrical resistivity tomography. Water Resour. Res. 39:1138.

Miyamoto, T., T. Annaka, and J. Chikushi. 2003. Soil aggregate structure effects on dielectric permittivity of an andisol measured by time domain reflectometry. Vadose Zone J. 2:90–97.

Moran, M.S., C.D. Peters-Lidard, J.M. Watts, and S. McElroy. 2004. Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. Can. J. Remote Sens. 30:805–826.

Mualem, Y. 1984. A modified dependent domain theory of hysteresis. Soil Sci. 137:283–291.

Ni-Meister, W. 2008. Recent advances on soil moisture data assimilation. Phys. Geogr. 29:19–37.

Nitao, J.J., and J. Bear. 1996. Potentials and their role in transport in porous media. Water Resour. Res. 32:225–250.

Njoku, E.G., and D. Entekhabi. 1996. Passive microwave remote sensing of soil moisture. J. Hydrol. 184:101–129.

Noborio, K., R. Horton, and C.S. Tan. 1999. Time domain reflectometry probe for simultaneous measurement of soil matric potential and water content. Soil Sci. Soc. Am. J. 63:1500–1505.

Olhoeft, G.R. 1985. Low-frequency electrical properties. Geophysics 50:2492–2503.

Or, D., D.P. Groeneveld, K. Loague, and Y. Rubin. 1991. Evaluation of single and multi-parameter methods for estimating soil-water characteristic curves. Geotechnical Engineering Report No. UCB/GT/91-07. University of California, Berkeley, CA.

Or, D., and J.M. Wraith. 1999. Temperature effects on soil bulk dielectric permittivity measured by time domain reflectometry: A physical model. Water Resour. Res. 35:371–383.

Paltineanu, I.C., and J.L. Starr. 1997. Real-time soil water dynamics using multisensor capacitance probes: Laboratory calibration. Soil Sci. Soc. Am. J. 61:1576–1585.

Parker, J.C. 1986. Hydrostatics of water in porous media, p. 209–296. *In* D.L. Sparks (ed.) Soil physical chemistry. CRC Press, Boca Raton, FL.

Pastor, J., and W.M. Post. 1986. Influence of climate, soil moisture, and succession on forest carbon and nitrogen cycles. Biogeochemistry 2:3–27.

Peck, A.J., and R.M. Rabbidge. 1969. Design and performance of an osmotic tensiometer for measuring capillary potential. Soil Sci. Soc. Am. Proc. 33:196–202.

Perlo, J., V. Demas, F. Casanova, C. Meriles, J. Reimer, A. Pines, and B. Blumich. 2005. High resolution spectroscopy with a portable single-sided sensor. Science 308:1279.

Petersen, L.W., P. Moldrup, O.H. Jacobsen, and D.E. Rolston. 1996. Relations between specific surface area and soil physical and chemical properties. Soil Sci. 161:9–21.

Phene, C.J., G.J. Hoffman, and S.L. Rawlins. 1971. Measuring soil matric potential in situ by sensing heat dissipation within a porous body: I. Theory and sensor construction. Soil Sci. Soc. Am. Proc. 35:27–33.

Purcell, E.M. 1952. Research in nuclear magnetism, p. 1942–1946. *In* Nobel lectures, physics. Oslo, Norway.

Rawls, W.J., and D.L. Brakensiek. 1989. Estimation of soil water retention and hydraulic properties, p. 275–300. *In* H.J. Morel-Seytoux (ed.) Unsaturated flow in hydraulic modeling theory and practice. NATO ASI Series. Series C: Mathematical and physical science. Vol. 275. Kluwer, Dordrecht, the Netherlands.

Redman, J.D., J.L. Davis, L.W. Galagedara, and G.W. Parkin. 2002. Field studies of GPR air launched surface reflectivity measurements of soil water content. Proc. SPIE 4758:156–161.

Reece, C.F. 1996. Evaluation of line heat dissipation sensor for measuring soil matric potential. Soil Sci. Soc. Am. J. 60:1022–1028.

Richards, L.A. (ed.) 1954. Diagnosis and improvement of saline and alkali soils. USDA handbook 60. US Government Printing Office, Washington, DC.

Robinson, D.A., A. Binley, N. Crook, F.D. Day-Lewis, T.P.A. Ferre, V.J.S. Grauch, R. Knight et al. 2008a. Advancing process-based watershed hydrological research using near-surface geophysics: A vision for, and review of, electrical and magnetic geophysical methods. Hydrol. Processes 22:3604–3635.

Robinson, D.A., C.S. Campbell, J.W. Hopmans, B.K. Hornbuckle, S.B. Jones, R. Knight, F. Ogden, J. Selker, and O. Wendroth. 2008b. Soil moisture measurement for ecological and hydrological watershed scale observatories: A review. Vadose Zone J. 7:358–389.

Robinson, D.A., and S.P. Friedman. 2001. Effect of particle size distribution on the effective dielectric permittivity of saturated granular media. Water Resour. Res. 37:33–40.

Robinson, D.A., S.B. Jones, J.M. Blonquist, Jr., and S.P. Friedman. 2005. A physically derived water content/permittivity calibration model for coarse-textured, layered soils. Soil Sci. Soc. Am. J. 69:1372–1378.

Robinson, D.A., S.B. Jones, J.M. Wraith, D. Or, and S.P. Friedman. 2003. A review of advances in dielectric and electrical conductivity measurement in soils using time domain reflectometry. Vadose Zone J. 2:444–475.

Rodriguez-Iturbe, I., and A. Porporato. 2005. Ecohydrology of water-controlled ecosystems: Soil moisture and plant dynamics. Cambridge University Press, Cambridge, U.K.

Ryel, R.J., C.I. Ivans, M.S. Peek, and A.J. Leffler. 2008. Functional differences in soil water pools: A new perspective on plant water use in water limited ecosystems, p. 397–422. *In* U. Luttge (ed.) Progress in botany. Vol. 69. Springer, London, U.K.

Samouelian, A., I. Cousin, A. Tabbagh, A. Bruand, and G. Richard. 2005. Electrical resistivity survey in soil science: A review. Soil Till. Res. 83:173–193.

Schjonning, P., I.K. Thomsen, P. Moldrup, and B.T. Christensen. 2003. Linking soil microbial activity to water- and air-phase contents and diffusivities. Soil Sci. Soc. Am. J. 67:156–165.

Schmugge, T.J., P. Gloersen, T. Wilheit, and F. Geiger. 1974. Remote sensing of soil moisture with microwave radiometers. J. Geophys. Res. 79:317–323.

Selig, E.T., and S. Mansukhani. 1975. Relationship of soil moisture to the dielectric property. ASCE Geotech. Eng. Div. GT8:755–770.

Serbin, G., and D. Or. 2004. Ground-penetrating radar measurement of soil water content dynamics using a suspended horn antenna. IEEE Trans. Geosci. Remote Sens. 42:1695–1705.

Seyfried, M.S., and M.D. Murdock. 2004. Measurement of soil water content with a 50-MHz soil dielectric sensor. Soil Sci. Soc. Am. J. 68:394–403.

Sheets, K.R., and J.M.H. Hendrickx. 1995. Noninvasive soil-water content measurement using electromagnetic induction. Water Resour. Res. 31:2401–2409.

Skopp, J., M.D. Jawson, and J.W. Doran. 1990. Steady-state aerobic microbial activity as a function of soil water content. Soil Sci. Soc. Am. J. 54:1619–1625.

Slater, P.J., and J.B. Williams. 1965. The influence of texture on the moisture characteristics of soils. I. A critical comparison of techniques for determining the available-water capacity and moisture characteristic curve of a soil. J. Soil Sci. 16:1–15.

Smith-Rose, R.L. 1934. The electrical properties of soils for alternating currents at radio frequencies. Proc. R. Soc. Lond. 140:359.

Song, Y.Q., S. Ryu, and P.N. Sen. 2000. Determining multiple length scales in rocks. Nature 406:178.

Spaans, E.J.A., and J.M. Baker. 1992. Calibration of watermark soil moisture sensors for temperature and matric potential. Plant Soil 143:213–217.

Stanford, G., and E. Epstein. 1974. Nitrogen mineralization–water relations in soils. Soil Sci. Soc. Am. J. 38:103–107.

Tarara, J.M., and J.M. Ham. 1997. Measuring soil water content in the laboratory and field with dual-probe heat-capacity sensors. Agron. J. 89:535–542.

Tidwell, V.C., and R.J. Glass. 1994. X-ray and visible-light transmission for laboratory measurement of 2-dimensional saturation fields in thin-slab systems. Water Resour. Res. 30:2873–2882.

Topp, G.C., J.L. Davis, and A.P. Annan. 1980. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. Water Resour. Res. 16:574–582.

Topp, G.C., and P.A. Ferré. 2002. Thermogravimetric method using convective oven-drying, p. 8–12. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4—Physical methods. SSSA, Madison, WI.

Turcu, V.E., S.B. Jones, and D. Or. 2005. Continuous soil carbon dioxide and oxygen measurements and estimation of gradient-based gaseous flux. Vadose Zone J. 4:1161–1169.

van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44:892–898.

van Genuchten, M.Th., F.J. Leij, and S.R. Yates. 1991. The RETC code for quantifying the hydraulic functions of unsaturated soils. EPA/600/2-91/065. USEPA, Ada, OK.

Vereecken, H., J.A. Huisman, H. Bogena, J. Vanderborght, J.A. Vrugt, and J.W. Hopmans. 2008. On the value of soil moisture measurements in vadose zone hydrology: A review. Water Resour. Res. 44:W00D06.

Walker, J.P., and P.R. Houser. 2001. A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. J. Geophys. Res. 106:11761–11774.

Waters, P. 1980. Comparison of the ceramic and the pressure membrane to determine the 15 bar water content of soils. J. Soil Sci. 31:443–446.

Weihermuller, L., J.A. Huisman, S. Lambot, M. Herbst, and H. Vereecken. 2007. Mapping the spatial variation of soil water content at the field scale with different ground penetrating radar techniques. J. Hydrol. 340:202–216.

Western, A.W., and R.B. Grayson. 1998. The Tarrawarra data set: Soil moisture patterns, soil characteristics, and hydrological flux measurements. Water Resour. Res. 34:2765–2768.

Western, A.W., R.B. Grayson, and G. Blöschl. 2002. Scaling of soil moisture: A hydrologic perspective. Annu. Rev. Earth Planet. Sci. 30:149–180.

Western, A.W., R.B. Grayson, G. Blöschl, G.R. Willgoose, and T.A. McMahon. 1999. Observed spatial organization of soil moisture and its relation to terrain indices. Water Resour. Res. 35:797–810.

Wolter, B., and M. Krus. 2005. Moisture measuring with nuclear magnetic resonance (NMR), p. 492–515. *In* K. Kepfer (ed.) Electromagnetic aquametry, electromagnetic wave interaction with water and moist substances. Springer, New York.

Wraith, J.M., and D. Or. 1998. Nonlinear parameter estimation using spreadsheet software. J. Nat. Resour. Life Sci. Educ. 27:13–19.

Wraith, J.M., and D. Or. 2001. Soil water characteristic determination from concurrent water content measurements in reference porous media. Soil Sci. Soc. Am. J. 65:1659–1666.

5

Water Flow in Soils

David E. Radcliffe
University of Georgia

Jirka Šimůnek
University of California, Riverside

5.1 Introduction

Water movement in soils is a key process that affects water quantity and quality in the environment. Since the transport of solutes is closely linked with the soil water flux, any quantitative analysis of solute transport must first evaluate water fluxes into and through soil. Water typically enters the soil surface in the form of precipitation or irrigation or by means of industrial and municipal spills (Figure 5.1). Some of the rainfall or irrigation water may be intercepted by the plant canopy. If the rainfall or irrigation intensity is larger than the infiltration capacity of the soil, water will be removed by surface runoff or will accumulate at the soil surface until it evaporates back to the atmosphere or infiltrates into the soil. Part of the water that infiltrates is returned back to the atmosphere by evaporation. Water that infiltrates into the soil profile may also be taken up by plant roots and eventually returned to the atmosphere by plant transpiration. The processes of evaporation and transpiration are often combined into the single process of evapotranspiration. Only water that is not returned to the atmosphere by evapotranspiration may percolate to the deeper soil zone and eventually reach the groundwater. If the water table is close enough to the soil surface, the process of capillary rise may move water from the water table through the capillary fringe toward the root zone and the soil surface.

Water movement in soils occurs under both saturated and unsaturated conditions. Saturated conditions occur below the water table where pressure head (h) is positive and water movement is predominately horizontal, with lesser components of flow in the vertical direction. While unsaturated conditions and negative pressure heads generally predominate above the water table (in the *vadose zone*), localized zones of saturation can exist especially following precipitation or irrigation or above an impermeable zone. As a general rule, water movement in the vadose zone is vertical but can also have large lateral components.

Governing equations for water flow can be developed for either steady or transient conditions, but analytical solutions usually
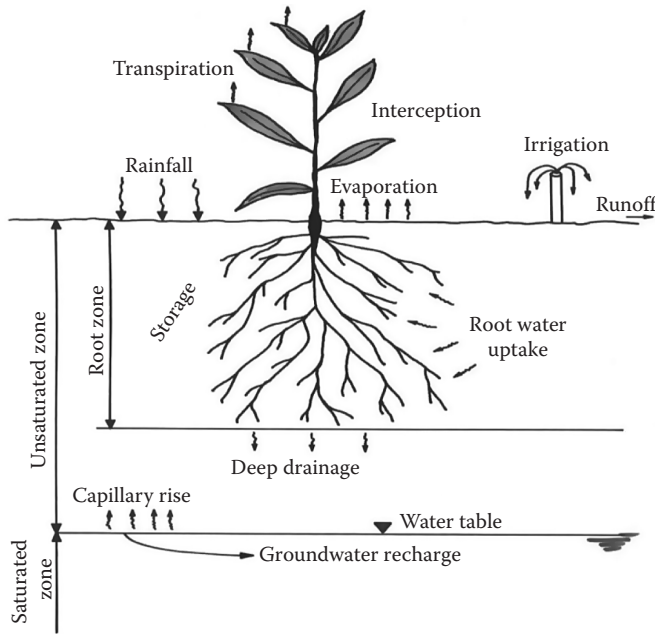
FIGURE 5.1 Water flow processes in soil. (From Šimůnek, J., M. Šejna, H. Saito, M. Sakai, and M.Th. van Genuchten. 2008. The HYDRUS-1D software package for simulating the movement of water, heat, and multiple solutes in variably saturated media, Version 4.0, HYDRUS Software Series 3. Department of Environmental Sciences, University of California Riverside, Riverside, CA.)

exist only for steady conditions. Although true steady flow may occur only rarely, transient flow usually approaches steady flow at long times, and the steady flow equations help to show what factors may be important under transient conditions. Numerical methods can be used to solve transient flow problems.

In the following sections, we start with steady water flow and describe the governing equations and show several examples using analytical solutions. Then we follow with transient water flow. The governing equations and initial and boundary conditions are described along with a brief introduction to numerical methods. Then, we describe in detail the important transient flow processes: infiltration, evaporation, redistribution, transpiration, preferential flow, and groundwater recharge. The HYDRUS numerical models (HYDRUS (2D/3D), Šimůnek et al., 2006; HYDRUS-1D, Šimůnek et al., 2008) are used to illustrate these processes. More detail on the model simulations can be found in Radcliffe and Šimůnek (2010).

5.2 Steady Flow in Saturated Soil

5.2.1 Darcy Equation

The equation that describes steady water flow through soil was developed by H. Darcy (Darcy, 1856; Hubbert, 1956). For 1D flow, it can be written as

$$J_w = -K_s \frac{\partial H}{\partial z}, \tag{5.1}$$

where
    J_w is the volumetric flow rate of water per unit cross-sectional area (L T^{-1})
    K_s is the saturated hydraulic conductivity (L T^{-1})
    H is the total head (L)
    z is the spatial coordinate (L)

J_w may be positive or negative depending on flow direction (the usual convention is positive for flow up or to the right). Darcy's law can be extended to 2D and 3D flow.

Darcy's equation takes the same form as several other important laws in science, among them, Fick's law for molecular diffusion, Ohm's law for electric current flow, and Fourier's law for heat conduction. It shows that for water to move there must be a difference in water potential (see Chapter 4 for definition), but the rate of water flow will depend on the hydraulic gradient and the hydraulic conductivity of the soil.

5.2.2 Saturated Flow Parameters

For saturated flow, the most important soil parameter is saturated hydraulic conductivity, which is a function of the fluid and soil properties:

$$K_s = \frac{k\rho_l g}{\eta}, \tag{5.2}$$

where
    k is the *intrinsic permeability* of the soil (L^2)
    ρ_l is the density of the fluid (M L^{-3})
    g is the gravitational acceleration (L T^{-2})
    η is the coefficient of dynamic viscosity of the fluid (M L^{-1} T^{-1})

It is apparent from this equation that temperature has an effect on K_s since it will affect ρ_l and η. It is generally true that while a given soil will have a constant k as long as the soil structure is not affected by the liquid, it will have a different K_s for different fluids.

Soils with low porosity, few large pores, and poor interconnectivity between pores have low K_s. Rawls et al. (1982) compiled values of K_s from 1323 soils collected over 32 states of the United States. Average K_s for 11 of the 12 United States Department of Agriculture (USDA) soil textural classes (the silt class is missing) from Rawls et al. (1982) are shown in Table 5.1. Brooks and Corey (1964) water retention curve parameters (see Chapter 4 for information on water retention functions) are also shown. Saturated hydraulic conductivities were highest in coarse-textured soils (sand, loamy sand, etc.) and declined in fine-textured soils (clay, silty clay, etc.), due to larger pores in coarse-textured soils. This can be corroborated by examining air-entry pressure heads h_a (absolute value of the pressure head at which the largest capillary pore empties), which are higher in fine-textured soils, indicating that the largest capillary pore is relatively small.

Saturated hydraulic conductivity varies widely among soils within a textural class, especially in the more fine-textured classes in Table 5.1. In Figure 5.2, K_s for 324 soil samples from the UNsaturated SOil hydraulic properties Database (UNSODA)

TABLE 5.1 Brooks and Corey (1964) Soil Hydraulic Parameters (K_s, θ_r, θ_s, h_a, and λ) for 11 Textural Classes

Textural Class	K_s (cm day^{-1})	θ_r (cm^3 cm^{-3})	θ_s (cm^3 cm^{-3})	h_a (cm)	λ (−)	λ_c (cm)
Sand	504.0	0.020	0.417	−7.26	0.592	9.88
Loamy sand	146.6	0.035	0.401	−8.69	0.474	12.28
Sandy loam	62.16	0.041	0.412	−14.7	0.322	22.18
Loam	31.68	0.027	0.434	−11.1	0.220	17.79
Silt loam	16.32	0.015	0.486	−20.7	0.211	33.38
Sandy clay loam	10.32	0.068	0.330	−28.1	0.250	44.16
Clay loam	5.52	0.075	0.390	−25.9	0.194	42.27
Silty clay loam	3.60	0.040	0.432	−32.6	0.151	55.03
Sandy clay	2.88	0.109	0.321	−29.2	0.168	48.61
Silty clay	2.16	0.056	0.423	−34.2	0.127	58.96
Clay	1.44	0.090	0.385	−37.3	0.131	64.07

Source: Rawls, W.J., D.L. Brakensiek, and K.E. Saxton. 1982. Estimation of soil water properties. Trans. ASAE 25:1316–1320, 1328.

Macroscopic capillary length (λ_c) calculated using Equations 5.11 and 5.16.
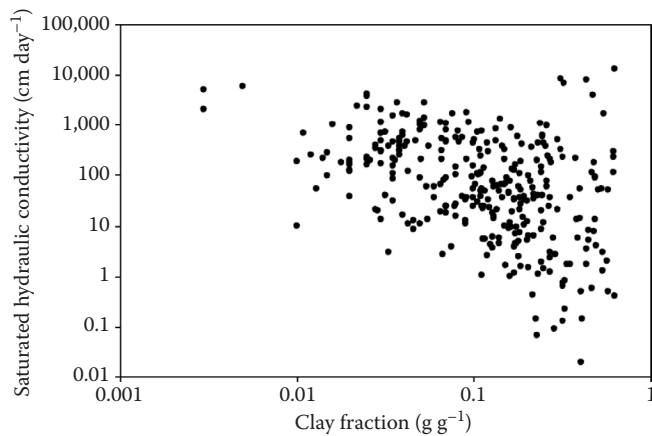
FIGURE 5.2 Saturated hydraulic conductivity as a function of clay fraction for 324 soils from the UNSODA database. (From Nemes, A., M.G. Schaap, F.J. Leij, and J.H.M. Wösten. 2001. Description of the unsaturated soil hydraulic database UNSODA version 2.0. J. Hydrol. 251:151–162.)

database (Nemes et al., 2001) are plotted as a function of clay content. Saturated hydraulic conductivity tends to decline as clay content increases, but there is considerable scatter at higher clay contents. The clayey soils with high K_s are probably soils with strong structure. Hence, soil structure as well as soil texture affects water flow in all but very sandy soils.

Saturated hydraulic conductivity is highly variable in space, not only in the vertical direction as horizons change but also in the horizontal direction within the same horizon (Nielsen et al., 1973; Russo and Bresler, 1980; Russo and Bouton, 1992). Saturated hydraulic conductivity also varies in time at much shorter scales than that of pedogenesis. This is especially true near the soil surface where land use, biology, and weather have more of an effect. Some authors have suggested that soil properties should be divided into static and dynamic properties (Grossman and Reinsch, 2002; Tugel et al., 2005). Hence, the effect of texture and mineralogy on K_s might be considered static, whereas other effects might be considered dynamic.

5.2.3 Flow Perpendicular to Layers

Most soils other than entisols and inceptisols have well-developed horizons or layers, which have different hydrologic characteristics including K_s. Above the water table, water usually flows vertically so that flow is perpendicular to these layers. Below the water table, flow can occur parallel to the layers.

Steady flow through a layered soil can be described by making an analogy with electric current flow through resistors in series (Jury and Horton, 2004). For a soil column with N layers of thickness L_1, …, L_N and saturated hydraulic conductivity K_1, …, K_N (column on the left side of Figure 5.3), the effective saturated hydraulic conductivity K_{eff} (L T^{-1}) is:

$$K_{eff} = \frac{\sum_{j=1}^{N} L_j}{\sum_{j=1}^{N} L_j/K_j}. \tag{5.3}$$

The effective hydraulic conductivity is not a simple arithmetic average of conductivities of individual layers or their limiting value, but a weighted average that considers the layer thicknesses as well as their conductivities. Thus, the thickness of layers is important.
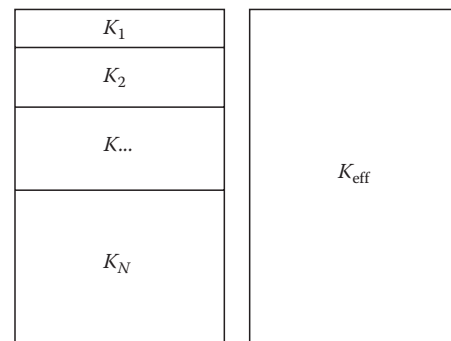
FIGURE 5.3 A soil column with N layers of thickness L_1, …, L_N and saturated hydraulic conductivities K_1, …, K_N on the left. An equivalent uniform column on the right with a saturated hydraulic conductivity K_{eff}.

Because the flux across all layers is the same under steady-flow conditions, Darcy's law can be used to show that when water flows down through a high-conductivity to a low-conductivity layer (e.g., a sand horizon over a clay horizon), a pressure head occurs at the interface that is greater than would be present in a uniform soil. This elevated pressure slows flow through the upper layer and speeds flow through the lower layer until flow through both layers is the same. In the reverse situation, (e.g., a clay over a sand), the interface pressure head is less than in a uniform soil, which speeds flow through the upper layer and slows flow through the lower layer. In a saturated, layered soil under steady flow, the pressure head distribution is linear with depth within each layer but with a different rate of change.

For flow from low- to high-conductivity layers where a large difference in conductivity is present, negative pressure heads can occur at the interface that exceed the air-entry pressure head of soil in the lower layer. This causes unsaturated conditions in the lower layer that greatly reduce flow through the profile. This often occurs when a *surface seal* or *crust* is present. Another way to equalize flow through a layered soil with a low-conductivity layer over a high-conductivity layer is for water to flow through only part of the less restrictive layer. This is a form of *preferential flow* called *fingering*. Both surface seals and fingering are discussed later in this chapter.

5.2.4 Flow Parallel to Layers

For flow parallel to layers (which usually occurs below the water table), the resistances of particular layers can be considered in parallel (more water flows through high- than low-conductivity layers). If the soil layers are of equal thickness, then K_{eff} is the arithmetic average of the hydraulic conductivities. For layers of different thicknesses, K_{eff} is a weighted average where the layer thicknesses are the weights:

$$K_{\text{eff}} = \frac{\sum_{j=1}^{N} L_j K_j}{\sum_{j=1}^{N} L_j}. \tag{5.4}$$

5.2.5 Flow to a Drain

Flow to a drain provides another example of steady saturated flow (Dupuit, 1863; Forchheimer, 1930). An equation can be developed for the maximum height of the water table between two drains or ditches under steady-flow conditions (Warrick, 2003; Figure 5.4). For a steady rainfall or irrigation rate R (L T^{-1}) and a drain separation distance $2s$ (L) with the water level in the ditch or drain maintained at a height z_0 (L), the assumption is that water in the saturated zone will flow only in the horizontal direction and the gradient will be the change in height of the water table:

$$J_x = -K_s \frac{dz}{dx}, \tag{5.5}$$
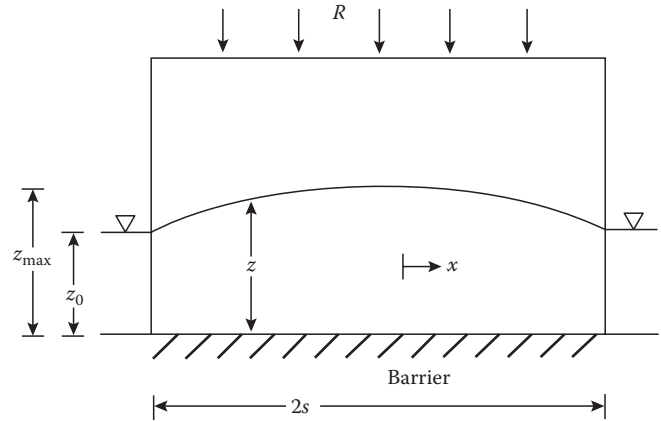
FIGURE 5.4 Geometry considered for defining the Dupuit–Forchheimer drainage equation (5.6). Water height in the ditch or drain is z_0. (From Warrick, A.W. 2003. Soil water dynamics. Oxford University Press, Oxford, U.K.)

where z is the height of the water table (L). Under these conditions, it can be shown that the maximum height of the water table z_{max} (L) between the drains or ditches is

$$z_{\text{max}} = \sqrt{\frac{Rs^2}{K_s} + z_0^2}. \tag{5.6}$$

From this equation, it can be seen that the water table will not rise as high between the drains and ditches when K_s is large. Also, the greater the rainfall or irrigation rate, the higher the water table will rise. The larger the spacing between drains, the higher the water table will rise. Equation 5.6 can be used to determine the distance for separating drains in order to maintain the water table within a given range from the soil surface. This equation is also useful later in the chapter when we consider groundwater recharge and hillslopes.

5.3 Steady Flow in Unsaturated Soil

The unsaturated zone lies between the water table and the soil surface (Figure 5.1). In this region, the water content of the soil is less than saturation ($\theta < \theta_s$), many pores are air filled, and pressure heads are negative.

Rising above the water table is the capillary fringe, a zone where water is under tension but very near saturation. The capillary fringe is important for water flow because only a slight change in water content or pressure head can cause a sharp change in the water table position (Sklash et al., 1986). Also, lateral flow increases in this zone (Liu and Dane, 1996; Abit et al., 2008).

The thickness of the capillary fringe depends on the water retention curve and can be approximated by the air-entry pressure head (h_a). From Table 5.1, it is apparent that sands have thinner capillary fringes than unstructured clays. Capillary fringes are also found around other zones of saturation in the unsaturated zone, such as injection boreholes, surface seepage basins, and losing streams.

5.3.1 Buckingham–Darcy Equation

In the unsaturated zone, larger pores drain more readily than smaller pores, as can be noted using the capillary rise equation (see Chapter 4). Therefore, the hydraulic conductivity is much smaller under unsaturated than saturated conditions due to water moving through smaller pores or as films along the walls of larger pores. At very low water contents, continuous fluid paths may not exist and water may move in the vapor phase. The unsaturated hydraulic conductivity is therefore represented as a function of pressure head $K(h)$ or as a function of water content $K(\theta)$. Buckingham (1907) modified Darcy's equation for unsaturated flow as follows:

$$J_w = -K(h)\frac{\partial H}{\partial z} = -K(h)\frac{\partial (h+z)}{\partial z} = -K(h)\left(\frac{\partial h}{\partial z} + \frac{\partial z}{\partial z}\right)$$

$$= -K(h)\left(\frac{\partial h}{\partial z} + 1\right), \qquad (5.7)$$

where the total head or hydraulic head is the sum of the pressure head and gravitational head, $H = h + z$.

5.3.2 Unsaturated Hydraulic Conductivity

As is the case with the water retention functions, equations have been developed to describe the hydraulic conductivity function. One of the simplest equations is the Gardner (1958) equation:

$$K(h) = K_s\,e^{\alpha h}, \qquad (5.8)$$

where α (L^{-1}) is a positive constant dependent on the soil. Since h is negative, the exponential term causes $K(h)$ to drop rapidly as h becomes more negative. A coarse-textured soil (sand), which has a rapidly decreasing hydraulic conductivity function compared to a fine-textured soil (clay), would be expected to have a larger value of α.

Other simple equations include the Campbell (1974) equation:

$$K(\theta) = K_s\left(\frac{\theta}{\theta_s}\right)^m, \qquad (5.9)$$

where m is a fitting parameter [-], and the Haverkamp equation (Haverkamp et al., 1977)

$$K(h) = \frac{K_s}{1 + (h/a)^N}, \qquad (5.10)$$

where a (L^{-1}) and N [-] are fitting parameters.

Many hydraulic conductivity functions have been derived using the pore-size distribution models of Burdine (1953) or Mualem (1976), in combination with various water retention functions (water retention functions are discussed in Chapter 4). The Brooks and Corey (1964) retention equation is commonly associated with Burdine's pore-size distribution model, leading to the hydraulic conductivity function:

$$K(S_e) = K_s S_e^{1+l+2/\lambda}, \qquad (5.11)$$

where

λ [-] is the pore-size distribution index from the Brooks and Corey (1964) water retention function (see Table 5.1)

l is a pore-connectivity parameter (assumed to be 2.0)

S_e is effective saturation [-]:

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r}, \qquad (5.12)$$

where

θ_s is the saturated volumetric water content (L^3 L^{-3})

θ_r is the residual volumetric water content (L^3 L^{-3})

The van Genuchten (1980) water retention function is similarly coupled most often with the model of Mualem to give

$$K(S_e) = K_s S_e^l[1 - (1 - S_e^{1/m})^m]^2 \qquad (5.13)$$

where m [-] and n [-] are empirical shape parameters from the van Genuchten (1980) water retention function and it is assumed that

$$m = 1 - \frac{1}{n}, \quad n > 1. \qquad (5.14)$$

The pore-connectivity parameter l in Equation 5.13 was estimated by Mualem (1976) to be 0.5 as an approximate average for many soils. However, Schaap and Leij (2000) recommended using $l = -1$ as an appropriate value for most soil textures.

Finally, Durner (1994) associated a *dual-porosity* retention model with Mualem's pore-size distribution model to give

$$K(S_e)$$

$$= K_s\frac{(w_1 S_{e_1} + w_2 S_{e_2})^l\{w_1\alpha_1[1 - (1 - S_{e_1}^{1/m_1})^{m_1}] + w_2\alpha_2[1 - (1 - S_{e_2}^{1/m_2})^{m_2}]\}^2}{(w_1\alpha_1 + w_2\alpha_2)^2},$$

$$(5.15)$$

where

w_i [-] are the weighting factors for the overlapping regions of the water retention functions

α_i [-] and m_i [-] are fitting parameters for the separate functions S_{ei} ($i = 1, 2$)

Values of $K(h)$ vary by orders of magnitude with pressure head, are sensitive to texture, clay mineralogy, and structure, and are highly variable in space and (in some cases) time.

Figure 5.5 presents examples of hydraulic conductivity functions for the sand, loam, and clay textural classes, using the
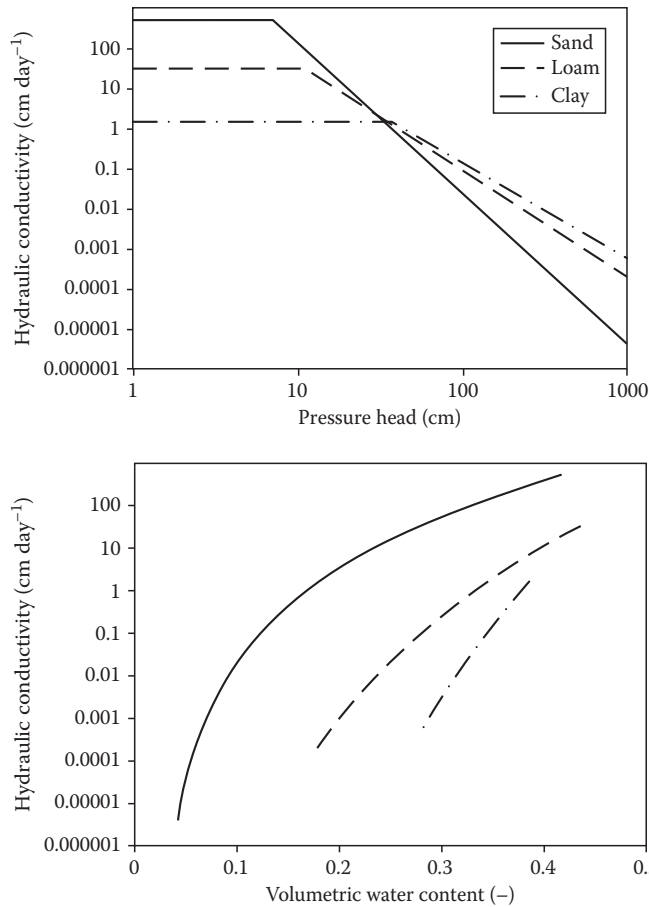
FIGURE 5.5 Examples of unsaturated hydraulic conductivity curves as a function of pressure head (top) and volumetric water content (bottom) for the sand, loam, and clay soil textural classes based on the Brooks and Corey (1964) equation (Equation 5.11) and parameters in Table 5.1.

Brooks and Corey (1964) unsaturated hydraulic conductivity equation (Equation 5.11) and the parameters given in Table 5.1. The hydraulic conductivity curves are presented as functions of both the pressure head (top) and water content (bottom). It is apparent that Equation 5.11 results in a linear relationship between ln $K(h)$ and ln h. Notice that the hydraulic conductivity at saturation is significantly larger for the sand than for the loam or clay. This difference is often several orders of magnitude. Also notice that the hydraulic conductivity decreases several orders of magnitude as the soil becomes unsaturated. This decrease, when expressed as a function of the pressure head (Figure 5.5, top), is much more significant for the sand than for the loam, and even more so than for the clay. The decrease for the sand is so dramatic that eventually the hydraulic conductivity becomes smaller than for the loam or clay. These properties of the hydraulic conductivity function are sometimes used in the design of capillary barriers in landfill covers to divert water from flowing through the underlying waste or in thin sand or gravel layers at the soil surface to prevent or limit evaporation.

5.3.3 Macroscopic Capillary Length

A useful way to quantify the unsaturated hydraulic conductivity function in a single parameter is in terms of the *macroscopic capillary length* λ_c (L) defined as

$$\lambda_c = \frac{\int_{h_i}^{h_0} K(h)\,dh}{K(h_0) - K(h_i)}, \tag{5.16}$$

where

h_0 (L) is a pressure head at or near zero
h_i (L) is a more negative pressure head such that $h_0 > h_i$ (White and Sully, 1987)

There are a number of ways to interpret λ_c. One way is a $K(h)$-weighted average pressure head over the range from h_0 to h_i (except that it is a positive number). If λ_c is small, this indicates that the *effective pressure head* for unsaturated flow is near zero. In other words, $K(h)$ is negligible except for pressure heads near zero (and water contents near saturation). This would be the case for a sand where $K(h)$ rises sharply at the wet end of the curve. An unstructured clay might have a larger λ_c, because the effective pressure head for unsaturated flow is more negative and $K(h)$ does not approach zero as fast at the dry end of the curve. Typical values for λ_c in four broad groups of soils are shown in Table 5.2. The effect of structure in fine-textured soils is apparent in that an unstructured clay has a λ_c of 25 cm, whereas a structured clay has a λ_c of 2.8 cm. Values for λ_c are also shown in Table 5.1 for each of the textural classes. These values were calculated using Equation 5.16 with $h_0 = 0$ cm, $h_i = -15{,}000$ cm, and the Brooks and Corey (1964) equation for $K(S_e)$ (Equation 5.11). In both tables, leaving aside the effect of structure, λ_c increases as the clay content increases.

If the Gardner equation (Equation 5.8) is used to describe $K(h)$ in Equation 5.16, it can be shown that

$$\lambda_c = \frac{1}{\alpha}. \tag{5.17}$$

TABLE 5.2 Soil Texture/Structure Categories for Estimation of Macroscopic Capillary Length (λ_c)

Soil Texture/Structure Category	λ_c (cm)
Coarse and gravelly sands; may also include some highly structured soils with large cracks and/or macropores	2.8
Most structured soils from clays through loams; also includes unstructured medium and fine sands	8.3
Soils that are both fine textured (clayey) and unstructured	25
Compacted, structureless, clayey materials such as landfill caps and liners, lacustrine or marine sediments, etc.	100

Source: Adapted from Elrick, D.E., and W.D. Reynolds. 1992. Infiltration from constant-head well permeameters and infiltrometers, p. 1–24. *In* G.C. Topp, W.D. Reynolds, and R.E. Green (eds.) Advances in measurement of soil physical properties: Bringing theory into practice. SSSA, Madison, WI.

Thus, a small value of λ_c indicates a rapid exponential decrease of the $K(h)$ function. Macroscopic capillary length appears in a number of equations later in this chapter, where it represents the effect of soil texture (and the shape of the unsaturated hydraulic conductivity curve) on unsaturated water flow.

5.3.4 Analytical Solutions for Steady Flow

It is useful to consider several cases of steady unsaturated flow. Although true steady flow may occur only rarely under unsaturated conditions, transient flow (discussed later in this chapter) usually approaches steady flow over long time periods, and the steady flow equations help to show what factors may be important under transient conditions.

As noted before, steady 1D flow in a uniform soil can be described by the Buckingham–Darcy equation (Equation 5.7). If the water content is uniform with depth, then $\partial h/\partial z = 0$, only gravity causes flow and the flux is equal to the hydraulic conductivity:

$$J_w = -K(h). \tag{5.18}$$

In this case, h is the same at all depths. This is called *gravity flow* or *unit-gradient flow*.

Analytical solutions to the Buckingham–Darcy differential equation (Equation 5.7) have been developed for a number of steady-flow conditions. The usual procedure is to rewrite the equation in an integral form (Jury and Horton, 2004):

$$\int_{h_1}^{h_2} \frac{dh}{1 + (J_w/K(h))} = z_1 - z_2. \tag{5.19}$$

If the equation for $K(h)$ is such that the integral on the left side can be evaluated (has an *analytical solution*), then the differential equation can be solved.

An example is steady infiltration ($J_w = -i$) in a uniform soil from the surface to a water table at depth L (Jury and Horton, 2004). This might occur in a soil remediation effort designed to flush contaminants from the unsaturated zone down into the capture zone of a well. In this case, $h_1 = 0$ at the water table, where $z_1 = -L$. The other limit is an arbitrary depth below the surface ($h_2 = h$ and $z_2 = z$) so that integration will result in an equation for the distribution of h with depth under steady flow conditions. The $K(h)$ function is described using the Haverkamp equation (Equation 5.10). To get an integral that can be evaluated, it is assumed that $N = 2$ in the Haverkamp equation. It can be shown (Jury and Horton, 2004) that the solution is

$$h = a\sqrt{\frac{K_s}{i} - 1}\,\tanh\left[-\frac{\sqrt{(1 - (i/K_s))(i/K_s)}}{a}(L + z)\right], \tag{5.20}$$

where $\tanh(x)$ is the hyperbolic tangent function.
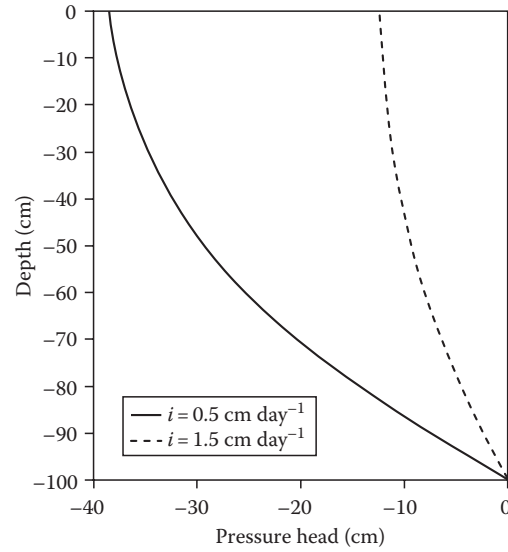
FIGURE 5.6 Distribution of h as a function of z for a steady low infiltration rate ($0.5\,\text{cm day}^{-1}$) and a steady high infiltration rate ($1.5\,\text{cm day}^{-1}$) based on Equation 5.20.

The distribution of h versus z predicted by Equation 5.20 for a Chino clay with $K_s = 1.95\,\text{cm day}^{-1}$, $a = -23.8\,\text{cm}$, and $N = 2$ (Gardner and Fireman, 1958) is shown for a low ($i = 0.5\,\text{cm day}^{-1}$) and a high ($i = 1.5\,\text{cm day}^{-1}$) infiltration rate in Figure 5.6. The distribution of h is curvilinear, contrary to a uniform or a layered saturated soil, especially with the lower infiltration rate, which causes a lower h at the surface. At the water table, the pressure head is zero, of course. Near the surface, especially at the high infiltration rate, the curve is nearly vertical indicating that gravity flow is the dominant form of flow (Equation 5.18).

Another example of steady downward flow is 3D water flow from a point source such as a ring placed at the soil surface (Figure 5.7). A steady infiltration rate will be approached after a period of time. Wooding (1968) developed an equation for these circumstances:

$$i_s = K(h_0)\left(1 + \frac{4\lambda_c}{\pi r}\right), \tag{5.21}$$
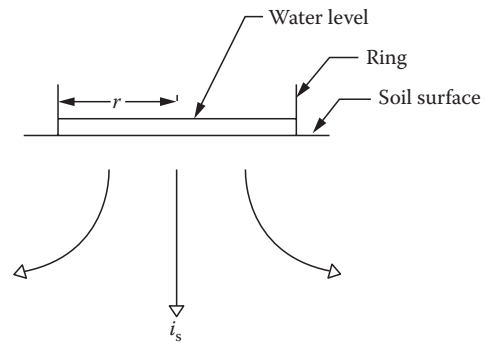
FIGURE 5.7 Steady 3D infiltration of water from a ring of radius r at the soil surface.

where

i_s is the steady infiltration rate (L T^{-1})

r is the radius of the ring (L)

$K(h_0)$ is the hydraulic conductivity (L T^{-1}) corresponding to the pressure head of the water supply at the soil surface ($z = 0$)

It is apparent by comparing Equations 5.21 and 5.18 that the second term in Equation 5.21 accounts for the increased infiltration rate from a ring due to lateral flow into dry soil (Figure 5.7). It is no surprise that this term includes λ_c since this represents the effect of capillarity, which will vary with soil texture and the shape of the $K(h)$ function. If water is ponded at the surface, $h_0 > 0$ and $K(h_0)$ should be what corresponds to K_s in the field. Experiments have shown that the *field-saturated hydraulic conductivity* (K_{fs}) is less than the saturated hydraulic conductivity that would be measured on a carefully saturated, intact core in the lab (K_s). This has been attributed to the effect of entrapped air under field conditions that reduces the cross-sectional area available for water flow and the water potential gradient. Bouwer (1969) recommended using a value for K_{fs} equal to one-half of K_s for predicting the steady-state infiltration rate.

The Buckingham–Darcy differential equation can also be solved for steady flow from a borehole (Figure 5.8):

$$Q_s = K_s \left(\pi r^2 + \frac{H\lambda_c}{G} + \frac{H^2}{G} \right), \qquad (5.22)$$

where

Q_s is the steady volumetric flow rate (L^3 T^{-1})

H is the height of water ponded in the borehole (L)

r is the radius of the hole (L)

G is a dimensionless geometric factor that depends primarily on the ratio of H/r (Elrick and Reynolds, 1992)

If the steady infiltration flux i_s is defined as the volumetric flow rate divided by the cross-sectional area of the borehole $Q_s/\pi r^2$,
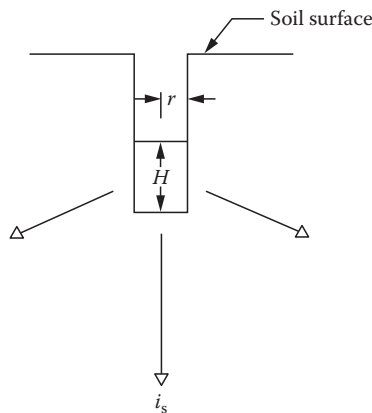
FIGURE 5.8 Steady 3D infiltration of water from a borehole.

TABLE 5.3 Dimensionless Coefficients for the Polynomial (Equation 5.24) Describing the Geometric Factor *G*, Valid for *H/r* < 10

Soil Texture/Structure	A_1	A_2	A_3	A_4
Sand	0.079	0.516	−0.048	0.002
Structured loams and clays	0.083	0.514	−0.053	0.002
Unstructured clays	0.094	0.489	−0.053	0.002

Source: Bosch, D.D., and L.T. West. 1998. Hydraulic conductivity variability for two sandy soils. Soil Sci. Soc. Am. J. 62:90–98.

the equation can be expressed in a form comparable to the Wooding equation (Equation 5.21):

$$i_s = K_s \left(1 + \frac{H\lambda_c}{G\pi r^2} + \frac{H^2}{G\pi r^2} \right). \qquad (5.23)$$

The similarity between Equation 5.23 and the Wooding equation (Equation 5.21) is apparent. In sequence, the terms inside the brackets in Equation 5.23 account for the effect of gravity, the shape of the unsaturated hydraulic conductivity function (note the λ_c term), and hydrostatic pressure in the borehole.

Bosch and West (1998) fitted a polynomial equation to the data of Elrick and Reynolds (1992) to determine the value of *G*:

$$G = \frac{1}{2\pi} \left[A_1 + A_2 \frac{H}{r} + A_3 \left(\frac{H}{r} \right)^2 + A_4 \left(\frac{H}{r} \right)^3 \right]. \qquad (5.24)$$

The values of the coefficients $A_1, \ldots, A_4$ [-] in this polynomial depend on texture and structure (Table 5.3). Equation 5.22 is used to measure K_s using a borehole permeameter, described later in this chapter.

5.4 Transient Water Flow

In this and the following sections, we address transient water movement. Analytical solutions of the differential equations that govern soil water flow for nonsteady, unsaturated flow conditions are very few. Simplifications of the unsaturated flow equations have been developed to describe infiltration because of the importance of this transient flow process. Alternatively, numerical models are commonly used to solve most transient flow problems in the unsaturated zone.

5.4.1 Governing Equations

Here, we discuss the Richards equation and the initial conditions and boundary conditions that apply to transient water flow.

5.4.1.1 The Richards Equation

A conservation equation for nonsteady vertical water flow can be developed by considering a representative elementary volume. The water flux into this volume during a small time interval

Δt must be equal to the water flux out of the volume, plus the change in water stored within the volume, and any sources or sinks. The resulting conservation equation is

$$\frac{\partial \theta}{\partial t} = -\frac{\partial J_w}{\partial z} - S(h), \tag{5.25}$$

where

 θ is the volumetric water content ($L^3 L^{-3}$)
 t is time (T)
 z is the spatial coordinate (L) (positive upward)
 J_w is the water flux ($L T^{-1}$)
 $S(h)$ is a pressure head dependent sink ($L^3 L^{-3} T^{-1}$) (a negative S is a source), usually accounting for root water uptake (transpiration)

Equation 5.25 is often referred to as the *mass conservation equation* or the *continuity equation*. The conservation equation must be combined with one or several equations describing J_w to produce the governing equation for variably saturated water flow. Formulations of governing equations for the different types of flow (uniform and preferential flow) are all based on this continuity equation. If the Buckingham–Darcy equation (Equation 5.7) is substituted for J_w, the result is

$$\frac{\partial \theta(h)}{\partial t} = \frac{\partial}{\partial z}\left(K(h)\frac{\partial h}{\partial z}\right) + \frac{\partial K(h)}{\partial z} - S(h). \tag{5.26}$$

This partial differential equation, first developed by Richards (1931), is the equation governing flow in the unsaturated zone. The first term on the right-hand side accounts for the effect of capillarity on water movement and the second term on the right-hand side accounts for the effect of gravity. It can be expanded to two and three dimensions.

Equation 5.26 is the *mixed form* of the Richards equation in that there are two dependent variables: θ and h. Various other formulations of the Richards equation are possible. For example, when the time derivative of the water content is expanded using the chain rule as follows:

$$\frac{\partial \theta h}{\partial t} = \frac{d\theta}{dh}\frac{\partial h}{\partial t} = C_w(h)\frac{\partial h}{\partial t} \tag{5.27}$$

one obtains the *pressure-head form*

$$C_w(h)\frac{\partial h}{\partial t} = \frac{\partial}{\partial z}\left(K(h)\frac{\partial h}{\partial z}\right) + \frac{\partial K(h)}{\partial z} - S(h), \tag{5.28}$$

where $C_w(h)$ is the soil water (or hydraulic) capacity function (L^{-1}) that characterizes the slope of the retention curve. The only independent variable is h:

The Richards equation (in any form) is a nonlinear partial differential equation. It is similar to the heat-flow equation that includes convection and conduction but differs in one very important way. The coefficient $K(h)$ that multiplies the gradient term is a function of the dependent variable h. This makes it a *nonlinear* partial differential equation. Because of its strongly nonlinear makeup, only a relatively few simplified analytical solutions can be derived. Most practical applications of the Richards equation require a numerical solution (see Section 5.6). Alternatively, the equation is simplified in some manner as happens with infiltration equations (see Section 5.7.3).

5.4.1.2 Initial Conditions

The solution to the Richards equation will depend on the initial and boundary conditions specified. Initial conditions characterizing the initial state of the system can be specified either in terms of the water content

$$\theta(z,t) = \theta_i(z,0) \tag{5.29}$$

or the pressure head

$$h(z,t) = h_i(z,0), \tag{5.30}$$

where θ_i ($L^3 L^{-3}$) and h_i (L) are initial values of water content and pressure head, respectively. For numerical models, it is generally best to specify initial conditions in terms of the pressure head, since this variable is the driving force for water flow. Specifying initial conditions in terms of the water content often leads to unrealistically large pressure-head gradients, and consequently water fluxes, across textural boundaries.

5.4.1.3 Boundary Conditions

Two types of conditions can be specified on boundaries of the transport domain: system-dependent and system-independent boundary conditions. System-independent boundary conditions are boundary conditions for which the specified boundary value (i.e., pressure head, water content, water flux, or gradient) does not depend on the status of the soil system. System-dependent boundary conditions are boundary conditions for which the actual boundary condition depends on the status of the system and is calculated by the model itself.

Several system-independent boundary conditions may be applied to the transport domain boundaries. When the pressure head at the boundary is known, one can use the *Dirichlet* or *type-1* boundary condition:

$$h(z, t) = h_0(z,t), \tag{5.31}$$

where h_0 is a prescribed pressure head (L). This boundary condition is often referred to as a *pressure-head boundary condition*. This boundary condition must be used when simulating ponded infiltration; for describing the hydrostatic pressure at

the boundary between the soil and standing or flowing water in a furrow, lake, or river; to specify the water level in a well; or to define the position of the groundwater table. The water flux across a Dirichlet boundary is not known a priori but must be calculated from the mathematical solution (either analytical or numerical) of the governing flow problem.

When the water flux across the boundary is known, one can use the *Neumann* or *type-2* boundary condition:

$$-K(h)\left(\frac{\partial h}{\partial z}+1\right)=J_0(z,t),\qquad(5.32)$$

where J_0 is a prescribed water flux (L T^{-1}). This boundary condition is often referred to as a *flux boundary condition*. A Neumann boundary condition must be used only along boundaries where the flux is known, provided the flux does not depend on the soil system. Hence, this boundary condition cannot be used to model precipitation or irrigation since the precipitation or irrigation rate may exceed the infiltration capacity of the soil, in which case ponding will occur and the actual boundary flux will decrease.

A particular form of the type-2 boundary condition is the *gradient boundary condition*:

$$\frac{\partial h}{\partial z}+1=g_0(z,t),\qquad(5.33)$$

where g_0 is a prescribed total gradient (L L^{-1}). Equation 5.33 is commonly used to specify a unit vertical hydraulic gradient simulating free drainage from the bottom of a soil profile when the water table is situated far below the domain of interest. The condition is consistent with unit-gradient conditions (see Equation 5.18) often observed in field studies of water flow, especially during redistribution (Sisson, 1987; McCord, 1991).

In many applications, neither the flux across nor the pressure head or gradient at a boundary is known a priori but follows from interactions between the soil and its surroundings (e.g., the atmosphere or deeper subsurface). The boundary representing the soil–air interface, which is exposed to atmospheric conditions, is one example of the system-dependent boundary. The potential water flux across this interface is controlled exclusively by external conditions (precipitation or evaporation). However, the actual flux depends also on the (transient) moisture conditions in the soil. Soil surface boundary conditions may change from prescribed flux to prescribed head conditions (and vice versa). This occurs, for example, when the precipitation rate exceeds the infiltration capacity of the soil, resulting in either surface runoff or accumulation of excess water on top of the soil, depending upon the soil conditions. The infiltration rate in that case is not controlled anymore by the precipitation rate but instead by the infiltration capacity of the soil. A system-dependent boundary condition also occurs when the potential evaporation rate, as calculated from meteorological conditions (the evaporative demand of the atmosphere), exceeds the capability of the soil to deliver enough water toward the soil surface. In this case, the potential

evaporation rate can be significantly reduced to an actual evaporation rate that is again controlled by the soil.

System-dependent atmospheric boundary conditions can be implemented mathematically using an approach of Neuman et al. (1974), which limits the absolute value of the flux such that the following two conditions are satisfied:

$$\left|K(h)\left(\frac{\partial h}{\partial z}+1\right)\right|\le E\qquad(5.34)$$

and

$$h_A\ ''\ h\ ''\ h_S,\qquad(5.35)$$

where

E is the maximum potential rate of infiltration or evaporation under the current atmospheric conditions (L T^{-1})

h is the pressure head at the soil surface (L)

h_A and h_S are, respectively, the minimum and maximum pressure heads allowed under the prevailing soil conditions (L)

The value for h_A is determined from the equilibrium conditions between soil, water, and atmospheric water vapor, whereas h_S is usually set equal to zero (which would initiate instantaneous surface runoff) or results from the accumulation of excess water in the surface ponding layer, in which case its value must be calculated from the difference between the infiltration and precipitation (or irrigation) rates. When one of the limits of Equation 5.35 is reached, a prescribed head boundary condition will be used to calculate the actual surface flux. Methods for calculating E and h_A on the basis of atmospheric data have been discussed by Feddes et al. (1974).

Another example of a system-dependent boundary condition is a seepage face through which water leaves the saturated part of the flow domain. The boundary condition in this case can be formulated mathematically as follows:

$$\begin{aligned}J_0(z,t)&=0\quad\text{for }h(z,t)<0,\\ h_0(z,t)&=0\quad\text{for }h(z,t)\ge0.\end{aligned}\qquad(5.36)$$

This boundary condition states that there is no flux across the boundary as long as the boundary is unsaturated and that the pressure head changes to zero once saturation is reached. The flux across the boundary is then calculated from the flow field by solving the governing flow equations. This boundary condition can be used at the bottom of certain types of lysimeters or along tile drains.

A special system-dependent boundary condition is sometimes implemented in 1D numerical models (e.g., van Dam et al., 1997; Šimůnek et al., 2006) to account for flow to horizontal subsurface tile drains. An equivalent drainage flux from the bottom of the simulated soil profile is then calculated using an appropriate analytical solution for the tile drainage system (e.g., Hooghoudt, 1940; Ernst, 1962; van Hoorn, 1997). The drainage equations involved generally hold for steady-state flow into the drain and depend

upon the geometry of the system (e.g., depth of tile drain, depth to impermeable layer, location of water table midway between two drains, and possibly information about soil layering).

5.5 Measurements of Hydraulic Properties

Accurate measurement of hydraulic parameters by both laboratory and field methods is essential for the prediction of water movement in soils. However, uncertainties in parameter estimates arise because of spatial variation, measurement accuracy, and scale effects. Many of these methods use steady flow conditions but recent advances have been made in using transient conditions. The important laboratory and field methods commonly employed in estimating hydraulic properties using steady-flow conditions are discussed below with additional procedures available in Dane and Topp (2002). Transient flow methods using inverse solutions and parameter optimization are also discussed. Lastly, pedotransfer functions (PTFs) as an alternative to direct measurement of hydraulic properties are described.

5.5.1 Laboratory Methods

The constant-head and falling-head methods are common for measuring K_s. In the former method, a constant head of water is maintained at the top and bottom of a soil core and the steady water flux J_w is recorded (Reynolds and Elrick, 2002a). The gradient is known, given the heads at the top and bottom of the core and the length of the core, so Darcy's equation (Equation 5.1) can be used to determine K_s.

In the falling-head method, a standpipe is attached to the top of a soil core and the head at the top is allowed to drop over time (Reynolds and Elrick, 2002c). Under these conditions, it can be shown that

$$K_s = \frac{L}{t_1} \ln\left(\frac{b_0 + L}{b_1 + L}\right), \qquad (5.37)$$

where

b_0 is the height of water ponded (L) at the top of the core at time zero

b_1 is the height of water ponded (L) at a later time t_1 (T)

The falling-head method is useful in soils with very low K_s, where it is difficult to collect sufficient drainage from a core to measure J_w accurately within an interval of several hours. The initial reading b_0 can be recorded and the core left overnight before recording b_1 without concern about the effect of evaporation on the collected drainage.

Values of K_s measured on cores in the laboratory will depend on what measures are used to remove entrapped air from the soil before making the measurement. Flushing the core with CO_2, wetting the core slowly from the bottom, and measuring K_s under conditions of an upward flux are commonly used.

A number of methods have been developed for measuring unsaturated hydraulic conductivity on cores in the laboratory. Unsaturated hydraulic conductivity can be measured in a manner similar to the constant-head method for K_s by adding water at the top of the core at a rate that does not cause ponding (Booltink and Bouma, 2002a). The total potential within the core is measured at two heights using miniature tensiometers and the steady flux is recorded. The Buckingham–Darcy equation (Equation 5.7) is solved for $K(h)$, where h in this case is the average pressure head within the core. Steady evaporation from the top of a core can also be used to measure $K(h)$ by placing tensiometers at several heights, recording the steady evaporation rate, and calculating the gradient between each pair of tensiometers (Arya, 2002).

Steady unsaturated flow can be imposed on a long soil core by supplying water continuously at a rate less than K_s. This can be done by applying a crust made of a mixture of gypsum and sand, or cement and sand, to the soil surface (Booltink and Bouma, 2002b). The steady water flux is measured with a constant positive head above the crust. Because of the low K_s of the crust, flow in the core is unsaturated (see Section 5.2.3). A tensiometer in the core records h, and if a unit gradient is assumed, then the measured flux is $K(h)$. Successive flux measurements with different crusts, each with a different conductivity, provide measurements of $K(h)$ over the range of pressures achieved. Bouma and Denning (1972) used this method in the field, as well.

In all laboratory methods of measuring K_s and $K(h)$, the effect of the chemistry of the added water on clay dispersion must be considered. The goal is to prevent dispersion so that pore-size distribution and connectivity do not change during the measurement, which can be achieved by using a dilute solution with a divalent cation (Chiang et al., 1987).

5.5.2 Field Methods

Common field methods for measuring K_s are use of the ring infiltrometer and borehole permeameter. Ring infiltrometers can be used to pond water on the soil surface and measure the soil infiltration rate (Reynolds et al. 2002). It is usually assumed that the Wooding equation (Equation 5.21) describes the 3D steady infiltration rate. Typical values of λ_c (Table 5.1 or 5.2) can be used to convert steady flow rates to K_s using this equation. Concentric double rings have been used to force 1D flow in the interior ring. In this case, the steady infiltration rate is an estimate of K_s in a uniform soil. However, Bouwer (1986) found that, for the typical dimensions used in a double ring, flow was not 1D.

Borehole permeameters are used to measure saturated hydraulic conductivity below the soil surface in the unsaturated zone (Reynolds and Elrick, 2002b). They consist of a Mariotte siphon (McCarthy, 1934) that maintains water at a constant level in a borehole and they allow measurement of the flow rate into the soil. The steady flow rate into the borehole Q_s is measured and then Equation 5.22 applies, but there are two unknowns in this equation: K_s and λ_c. One approach is to measure Q_s at two values of H in the same borehole and solve simultaneous equations for K_s and λ_c. Since changing the level of H in the borehole necessarily changes the region of soil that is being sampled, soil heterogeneity in the form of layering or macropores can result in unrealistic and invalid (i.e., negative) K_s and λ_c. As many as

30%–80% of measurements of K_s and λ_c in structured soils may be invalid (both negative) according to Elrick and Reynolds (1992). Alternatively, a single measurement of Q_s may be used and λ_c estimated using Table 5.1 or 5.2. Studies suggest that this method yields values of K_s that are usually accurate to within a factor of 2 (Reynolds et al., 1992).

Tension infiltrometers (also called disc permeameters) are a popular method for determining unsaturated hydraulic conductivity and other hydraulic parameters in field soils. They consist of a circular porous plate or membrane, which is placed on the soil surface or an excavated surface. Water is supplied to the plate under tension using a Mariotte bottle arrangement and the rate of water entry into the soil can be measured on a graduated cylinder or with a pressure transducer (Clothier and Scotter, 2002). In some cases, a ring is attached to the tension infiltrometer to allow ponded infiltration measurements at the same location (Perroux and White, 1988). Two approaches are common, one in which the infiltration rate is measured after it attains a steady rate and the other in which the early nonsteady infiltration rate is measured (Warrick, 1992).

5.5.3 Inverse Solutions and Parameter Optimization

Modeling transient water flow requires an accurate estimation of a number of hydraulic parameters including the soil water retention curve, saturated hydraulic conductivity, and the unsaturated hydraulic conductivity function. As an alternative to direct measurement of these parameters, they can be estimated indirectly using *inverse modeling*. This differs from the direct methods where steady-flow conditions are required due to the analytical solutions that are used. Whereas steady-state methods invert Darcy's equation, transient methods invert the Richards equation. Inverse methods for these conditions have become possible because advances in computer science have made obtaining numerical solutions much quicker, so solutions can be run repeatedly to find the optimum parameter values.

Inverse modeling is a form of model calibration, which is frequently used in hydrology (Gupta et al., 2003). It requires a set of observed data such as measured water contents or pressure heads. In model calibration, the objective is usually to obtain better model predictions. Parameter optimization differs in that the objective is to determine the best estimate of the parameters as an alternative to directly measuring them.

The HYDRUS models have an inverse modeling capability. The objective function Φ that is minimized in HYDRUS-1D is defined as

$$\Phi(b,q,p) = \sum_{j=1}^{m_q} v_j \sum_{i=1}^{n_{qj}} w_{i,q}[q_j^\star(z,t_i) - q_j(z,t_i,b)]^2$$

$$+ \sum_{j=1}^{m_p} \bar{v}_j \sum_{i=1}^{n_{pj}} \bar{w}_{i,q}[p_j^\star(\theta_i) - p_j(\theta_i,b)]^2 + \sum_{j=1}^{n_b} \hat{v}_j(b_j^\star - b_j)^2.$$

$$(5.38)$$

The first term on the right-hand side of Equation 5.38 represents the deviations between measured and predicted variables

such as pressure heads or water contents at different depths and times. Predicted and measured fluxes or cumulative fluxes across boundaries can also be used as variables. In the first term, m_q is the number of different sets of measurements, n_{qj} is the number of measurements in a particular measurement set, $q_j^\star(z,t_i)$ represents specific measurements at time t_i for the jth measurement set at depths z (a vector of locations), $q_j^\star(z,t_i,b)$ are the corresponding model predictions for the vector of optimized parameters (e.g., θ_r, θ_s, α, n, K_s, etc.) and v_j and w_{ij} are weights associated with a particular measurement set or point, respectively. The second term on the right-hand side of Equation 5.38 represents differences between independently measured and predicted soil hydraulic properties (e.g., $\theta(h)$ or $K(h)$ data), while the terms m_p, n_{pj}, $p_j^\star(\theta_i)$, $p_j(\theta_i,b)$, $\bar{v}_j$, and $\bar{w}_{ij}$ have similar meanings as for the first term but now apply to hydraulic properties. The last term on the right-hand side of Equation 5.38 represents a penalty function for deviations between prior knowledge of the soil hydraulic parameters $b_j^\star$ and their final estimates b_j, with n_b being the number of parameters with prior knowledge and $\hat{v}_j$ representing preassigned weights. Minimization of the objective function is accomplished in HYDRUS using the Levenberg–Marquardt nonlinear minimization (Marquardt, 1963).

Transient flow parameter optimization methods are described by Hopmans et al. (2002). With these methods, both the soil retention and unsaturated hydraulic conductivity functions can be estimated from a single transient experiment. Inverse methods can be used in laboratory experiments and in field experiments. A common laboratory experiment is the multistep outflow method. In this method, a pressure cell similar to what is used for the determination of a water retention curve is employed. The air pressure on the cell is increased once (the *one-step method*) or a number of times to successively higher pressures (the *multistep method*) much like what is done when measuring a water retention curve. However, instead of waiting until the cell reaches equilibrium at a given pressure to measure the water content, pressure head in the cell and water flow out of the cell are measured continuously during the experiment. The observed and predicted pressure heads in a multistep method experiment analyzed using HYDRUS-1D are shown in Figure 5.9. Four pressure changes were used and the experiment ran for nearly 200 h. After eight iterations, HYDRUS-1D reduced the objective function sufficiently for the final parameters to be considered optimized. Other transient flow experiments that can be used for parameter optimization include evaporation, use of a tension infiltrometer, and field drainage (Hopmans et al., 2002).

5.5.4 Pedotransfer Functions

Considering that direct measurements of the unsaturated soil hydraulic properties are relatively difficult and time consuming, many have attempted to predict soil hydraulic properties from more easily measurable surrogate properties, such as soil texture. Relationships between the soil hydraulic and other properties are commonly called PTFs. The functions are usually obtained using
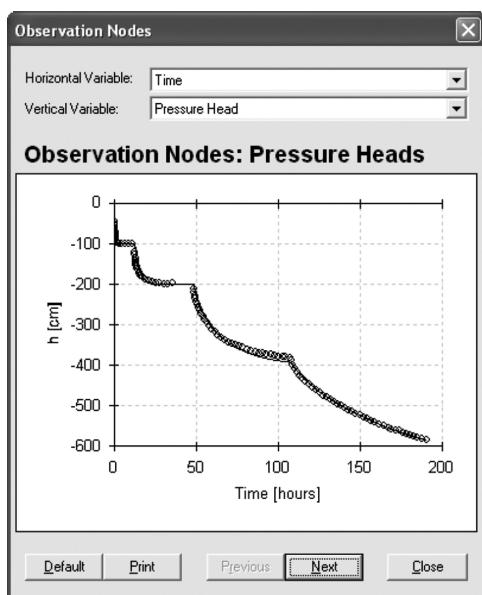
FIGURE 5.9 Observed pressure head (circles) and predicted pressure head (curve) in a multistep outflow experiment analyzed using HYDRUS-1D.

various mathematical and statistical approaches, such as regression (Rawls and Brakensiek, 1985; Vereecken et al., 1989, 1990; Rawls et al., 1991) or neural network analysis (Schaap et al., 1998; Schaap and Leij, 2000). PTFs can be used to predict either the soil hydraulic properties directly, such as the water content at specified pressure heads or the saturated hydraulic conductivity, or parameters in the analytical models used for the soil hydraulic properties (Rawls and Brakensiek, 1985; Vereecken et al., 1989, 1990). Recent reviews of PTFs are given by Leij et al. (2002) and Pachepsky et al. (2006).

The HYDRUS models implement the Rosetta Lite module of Schaap et al. (2001) to predict soil hydraulic parameters using five different levels of input data. The simplest model (Model 1) uses the average of fitted hydraulic parameters within a textural class in the USDA textural triangle. The four other models use progressively more detailed input data, starting with the sand, silt, and clay

fractions (Model 2), then adding a measured bulk density (Model 3), and water contents at 33 (Model 4) and 1500 kPa (Model 5).

Although very convenient to use, PTFs can produce relatively large errors in the predicted soil hydraulic properties. Schaap et al. (1998) and Schaap and Leij (1998) reported that the root mean square errors (RMSE) for water content and $\log(K_s)$ predictions were 0.108 and 0.741, respectively, using the simplest Rosetta model (Model 1). They still obtained relatively large RMSE values of 0.063 and 0.610 for water contents and $\log(K_s)$, respectively, with Model 5 that requires the most detailed information. Other studies found that the accuracy for predicting water content using various PTFs is relatively good (RMSE of $0.02–0.11\,cm^3\ cm^{-3}$) compared to predicting $\log(K_s)$ (RMSE no better than 0.5) (Jaynes and Tyler, 1984; Ahuja et al., 1989; Tietje and Hennings, 1996; Donatelli et al., 2004).

5.6 Numerical Methods

Numerical methods can be used to solve the Richards equation under conditions where analytical solutions do not exist. These methods were first applied to soil water movement in the early 1960s (Ashcroft et al., 1962; Hanks and Bowers, 1962). The two most common methods are *finite differences* and *finite elements* (Huyakorn and Pinder, 1983; Smith, 1985). For 1D flow, the first step is to discretize the soil profile into a number of depth intervals Δz. The point at which intervals join is called a *node*. Time is also discretized by intervals Δt. The depth and time intervals can vary. A set of algebraic equations is developed that include the pressure head or water content at each node at a given time level. This leads to a system of *simultaneous equations* that can be solved to find the pressure head or water content using matrix algebra at each time level. The method can be applied to 2D and 3D problems in which case, spatial discretization leads to a grid.

Commonly used numerical models that solve the Richards equation to simulate water flow in variably saturated soils are shown in Table 5.4. In the following sections, HYDRUS-1D and HYDRUS (2D/3D) are used to illustrate important transient flow processes.

TABLE 5.4 Widely Used Numerical Models for Simulating Variably Saturated Water Flow in Soils

Model Name	Internet Address	Reference
COUP	http://www.lwr.kth.se/vara%20datorprogram/CoupModel	Jansson and Karlberg (2001)
DAISY	http://code.google.com/p/daisy-model/	Hansen et al. (1990)
HYDRUS-1D	http://pc-progress.com/en/Default.aspx?hydrus-1d	Šimůnek et al. (2008)
HYDRUS-2D	http://pc-progress.com/en/Default.aspx?hydrus-2d	Šimůnek et al. (1999)
HYDRUS (2D/3D)	http://pc-progress.com/en/Default.aspx?hydrus-3d	Šimůnek et al. (2006)
MACRO	http://www.mv.slu.se/BGF/Macrohtm/macro.htm	Jarvis (1994)
RZWQM	http://www.ars.usda.gov/Main/docs.htm?docid=17740	Ahuja and Hebson (1992)
SHAW	http://www.ars.usda.gov/Services/docs.htm?docid=16931	Flerchinger et al. (1996)
SWAP	http://www.swap.alterra.nl/	van Dam et al. (1997)
SWIM	http://www.clw.csiro.au/products/swim/	Verburg et al. (1996)
TOUGH2	http://www-esd.lbl.gov/TOUGH2/	Pruess (1991)
UNSAT-H	http://hydrology.pnl.gov/resources/unsath/unsath.asp	Fayer (2000)
VS2DI	http://wwwbrr.cr.usgs.gov/projects/GW_Unsat/vs2di1.2/	Healy (1990)

5.7 Infiltration

Infiltration is a key process because it determines how much water from rainfall, irrigation, or a contaminant spill enters the soil and how much becomes runoff. It is also important in erosion since there can be no erosion without runoff to transport and scour sediment. In this section, we describe infiltration into a uniform soil, the effect of soil crusts and subsurface layers, infiltration equations, infiltration from a borehole, and subsurface irrigation.

5.7.1 Infiltration into a Uniform Soil

In this example, HYDRUS-1D is used to simulate infiltration into a uniform soil with the hydraulic properties of the Ap horizon of a Cecil sandy loam soil in Georgia (Table 5.5). The water retention parameters were obtained by fitting the van Genuchten (1980) equation to the data from Bruce et al. (1983) using the RETention Curve (RETC) optimization software (van Genuchten et al., 1991). See Radcliffe and Šimůnek (2010) for more details.

Infiltration once ponding starts at the surface is simulated. Therefore, the surface boundary condition for infiltration after ponding occurs (with no surface storage) is a pressure head at the soil surface equal to zero (Dirichlet or type-1 boundary condition):

$$h(0, t) = 0. \tag{5.39}$$

It is assumed that there is no flux (Neumann or type-2 boundary condition) at the bottom boundary ($z = -250\,\text{cm}$):

$$J_w(-250, t) = 0. \tag{5.40}$$

For initial conditions, the soil profile has a pressure head of $-100\,\text{cm}$ at all depths:

$$h(z, 0) = -100. \tag{5.41}$$

The numerical solution is used to plot profile distributions of pressure head for different times after water began to pond at the surface of a soil during an infiltration event. These are shown in Figure 5.10 for various times. During infiltration, a wetting front
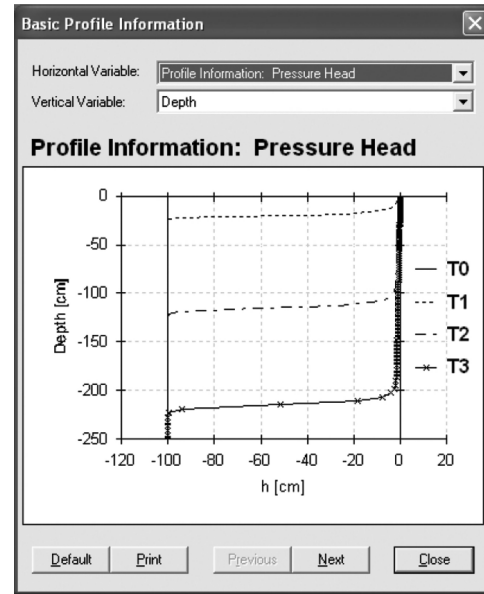
FIGURE 5.10 Pressure heads as a function of depth for various times during infiltration into a uniform loamy sand; $T0 = 0.0\,\text{h}$, $T1 = 0.1\,\text{h}$, $T2 = 1.0\,\text{h}$, and $T3 = 2.0\,\text{h}$.

of higher pressure heads moves down through the soil over time. For a given soil, the abruptness of the wetting front will depend on the shape of the $\theta(h)$ and $K(h)$ functions. Both of these depend on pore-size distribution. For coarse-textured soils with a narrow pore-size distribution, the wetting front will be more abrupt; in a fine-textured soil, the wetting front will be more diffuse. In this case, the wetting front is very sharp as one might expect in a loamy sand. The wetting front is a combination of new water added by the rain and old water displaced to lower depths.

The numerical solution can also be used to estimate the infiltration rate $i(t)$ as a function of time. The infiltration rate is the water flux at the soil surface, so it is described by the Buckingham–Darcy equation for the flux at the surface:

$$i(t) = -J_w(t)\big|_{z=0} = \left[K(h)\frac{\partial h}{\partial z} + K(h) \right]\bigg|_{z=0}, \tag{5.42}$$

where the vertical bar and subscript indicates that the flux is evaluated at the soil surface.

TABLE 5.5 Soil Properties and van Genuchten (1980) Water Retention and Hydraulic Conductivity Parameters for the Cecil Loamy Sand, plot 4, of Bruce et al. (1983)

Horizon	Texture	Depth (cm)	θ_r (cm³ cm⁻³)	θ_s (cm³ cm⁻³)	α (cm⁻¹)	n	K_s (cm h⁻¹)
Ap	Loamy sand	0–21	0.032	0.399	0.0495	1.46	19.19
BA	Clay loam	21–26	0.137	0.346	0.1396	1.25	7.69
Bt1	Clay	26–102	0.000	0.433	0.0382	1.08	10.73
Bt2	Clay	102–131	0.000	0.424	0.0058	1.09	0.206
BC	Clay loam	131–160	0.000	0.423	0.0042	1.14	0.035
C	Sandy clay loam	160–250+	0.000	0.449	0.0078	1.22	0.467
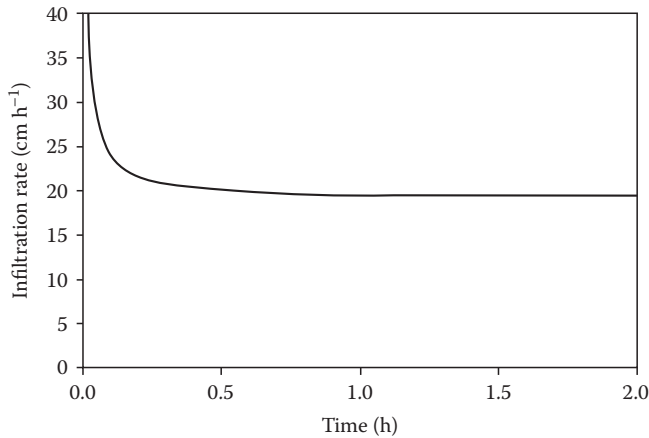
FIGURE 5.11 Infiltration rate as a function of time in a uniform loamy sand.

The infiltration rate is shown in Figure 5.11. It decreases with time and approaches an asymptotic minimum infiltration rate. Experimental observations confirm that $i(t)$ follows the pattern in this figure. This is for a uniform soil with no layers or crusting at the surface and relatively wet antecedent water contents. Equation 5.42 can be used to understand this pattern. When infiltration first starts, the wetting front is steep and very close to the surface (Figure 5.10). As a result, $\partial h/\partial z$ at the surface is large and the first term in Equation 5.42 produces a large value for $i(t)$. Under these conditions, the gradient in pressure head (due mainly to capillarity) is responsible for the rapid movement of water into a moist soil. The second term in Equation 5.42 represents the effect of gravity and may have little effect on water movement during the initial stages of infiltration. Later in the infiltration event, the wetting front has moved deeper into the soil (Figure 5.10). As a result, $\partial h/\partial z$ at the surface is much smaller (maybe even zero) and the first term in Equation 5.42 (and capillarity) has little effect. When the first term approaches zero, $i(t) \approx K_s$ and only gravity causes flow. Under these circumstances, a minimum infiltration rate is reached and it is approximately the saturated hydraulic conductivity of the Cecil Ap horizon, 19.19 cm h^{-1} (Table 5.5), as demonstrated in Figure 5.11.

Because the initial high infiltration rate in a dry soil is due to the pressure head gradient term, if the soil is wetter before the rainfall event, the initial infiltration rate will be lower. However, the final infiltration rate in both a wet and a dry soil will be the same, approximately K_s.

Cumulative infiltration $I(t)$ (L) is the integral of the infiltration rate:

$$I(t) = \int_{0}^{t} i(t)\,dt. \qquad (5.43)$$

It is the area under the rate curve in Figure 5.11. Once the infiltration rate becomes constant, cumulative infiltration increases linearly.

To get a more realistic simulation, one can specify slightly different initial and boundary conditions. For the surface boundary condition, a system-dependent boundary condition (see Equations 5.34 and 5.35) is used that starts with a flux equal to a rainfall rate of 25 cm h^{-1}. When the soil near the surface wets up sufficiently that the pressure head at the first node reaches 0 (h_S in Equation 5.35), the boundary condition changes to a constant-pressure-head boundary condition equal to 0 (ponding with no surface storage). For the bottom boundary condition, a constant-pressure-head boundary condition of zero is used to represent a water table at a depth of 250 cm. For the initial conditions, a pressure-head distribution starting with −250 cm at the soil surface and increasing linearly to the pressure head of 0 cm at the bottom of the model space is specified. This represents a soil profile where the pressure heads are in equilibrium with a water table at a depth of 250 cm.

The simulated distribution of pressure heads as a function of depth for various times are shown in Figure 5.12. The initial equilibrium distribution of pressure heads is apparent. The movement of a wetting front down the profile is also apparent with the front reaching a depth of about 200 cm at the end of the simulation. Compare this figure with the earlier example of infiltration into the same soil with an initial distribution of pressure heads equal to a constant −100 cm (Figure 5.10). The results are similar except for the initial distribution of pressure heads.

The simulated infiltration rate as a function of time is shown in Figure 5.13. Infiltration is a constant flux equal to 25 cm h^{-1} until about 0.2 h and then decreases to a steady rate of 19.19 cm h^{-1}, which is K_s for the soil. The effect of the system-dependent boundary condition at the surface is apparent when this
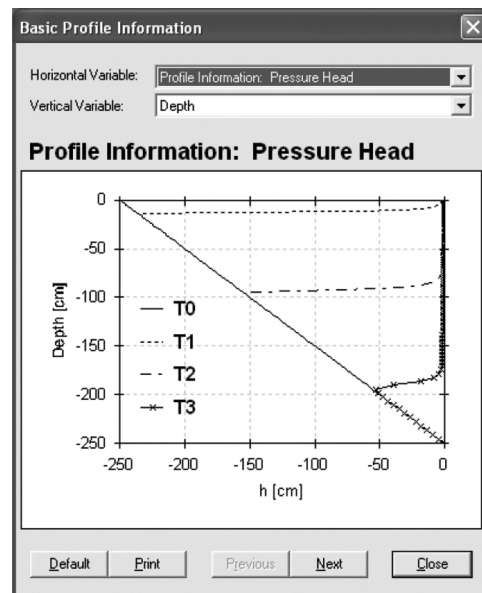
FIGURE 5.12 Resulting pressure head distributions in the infiltration example with an equilibrium initial distribution of pressure heads: $T0 = 0.0\,h$, $T1 = 0.1\,h$, $T2 = 1.0\,h$, and $T3 = 2.0\,h$.
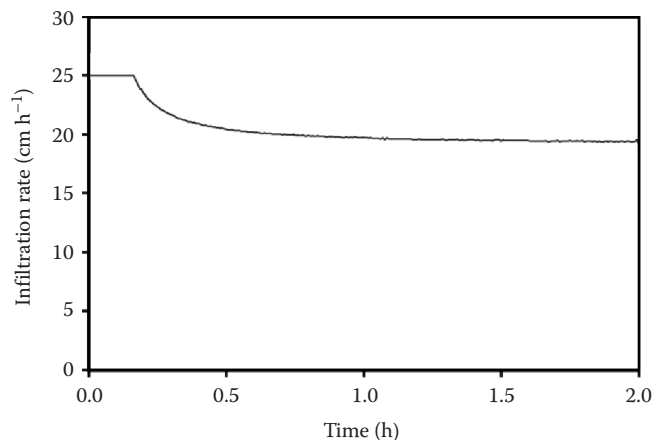
FIGURE 5.13 Infiltration rate as a function of time in the infiltration example with a system-dependent boundary condition.

figure is compared to the earlier infiltration example (Figure 5.11). With the rainfall rate of 25 cm h^{-1}, once the infiltration rate for the soil drops below this rate, water starts to run off (assuming that surface storage is negligible). In the earlier example, we assumed that water was ponded at the surface from the beginning of the infiltration event. It would take a very high-intensity rainfall event to cause ponding early in the event (more than 40 cm h^{-1}, see Figure 5.11). The system-dependent boundary condition used in this example is more realistic.

5.7.2 Soil Crusts and Subsurface Layers

Another factor that can cause the infiltration rate to decrease is the formation over time of a *surface seal* or *crust* at the soil surface. A surface seal is a very thin layer (1–5 mm) at or just below the soil surface that forms due to the breakdown of soil aggregates and chemical dispersion of clay particles under raindrop impact. The clay particles fill the soil pores and create a layer with a saturated hydraulic conductivity several orders of magnitude less than the undisturbed soil (Miller and Radcliffe, 1992). This low-conductivity layer can prevent saturation of the soil just beneath the seal due to the suction that occurs at the interface (see Section 5.2.3), further reducing the infiltration rate.

Subsurface clay layers near the surface can also reduce infiltration rates. An unstructured clay layer will usually have a lower K_s than an overlying sand layer and reduce K_{eff} (Equation 5.3) and $i(t)$ once the wetting front enters the clay layer. Again, HYDRUS-1D can be used to plot distributions of pressure heads and water contents in a soil profile that consists of the various horizons of a Cecil sandy loam soil in Georgia (Bruce et al., 1983; Table 5.5). The boundary and initial conditions are the same as in the first infiltration example for a uniform soil (Equations 5.39 through 5.41). The pressure heads at various times are shown in Figure 5.14. The curves are remarkably different from Figure 5.10 for a uniform soil, due primarily to the presence of positive pressures that have peak values at a depth of 100 cm after 1 h. This depth corresponds to the top of the Bt2 horizon. This horizon and the horizons below it have sharply lower K_s than the horizons above (Table 5.5).
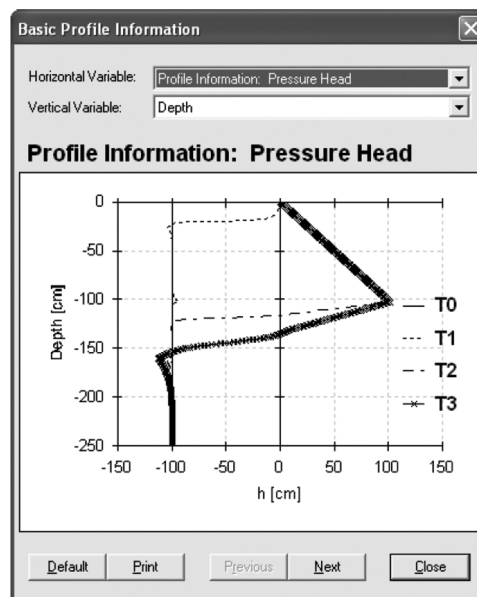
FIGURE 5.14 Pressure heads as a function of depth for various times during infiltration into a layered soil: $T0 = 0.0$ h, $T1 = 0.1$ h, $T2 = 1.0$ h, and $T3 = 4.0$ h.

When the wetting front reaches this horizon, water flow is slowed causing pressure to build above the Bt2 horizon. This produces a larger gradient in pressure head across the lower horizons and compensates to a degree for the low conductivities. At a depth of approximately 160 cm, pressure heads have become more negative than the initial conditions. This is due to the increase in K_s that occurs in the C horizon (Table 5.5).

As noted earlier in Section 5.2.3, under steady saturated flow, pressure heads are distributed linearly with depth within a horizon. In Figure 5.14, the distribution is linear above the wetting front, indicating that θ is constant (saturated) in this region. At the wetting front, the lines curve, indicating that θ is changing over time in this region (unsaturated).

The effect of soil layers on the infiltration rate in the layered soil is shown in Figure 5.15. The infiltration rate for a uniform
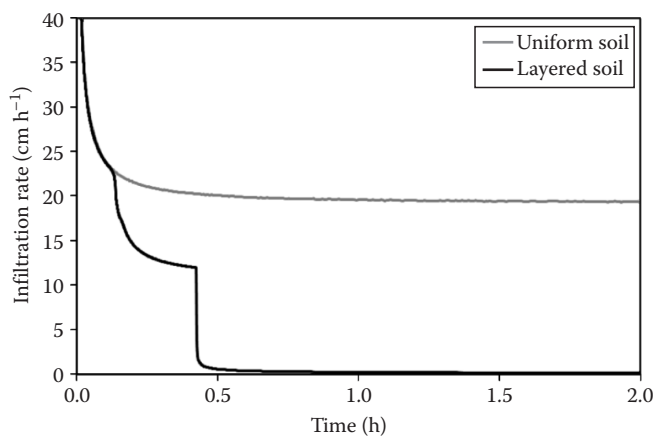
FIGURE 5.15 Infiltration rate as a function of time in a uniform soil and a layered soil.

soil with the same properties of the first horizon in the layered soil (the curve in Figure 5.11) is included for comparison purposes. Initially, the infiltration rates are the same, but once the wetting front reaches the second (BA) horizon after about 0.2 h, the infiltration rate in the layered soil decreases. This is due to a pressure head gradient at the soil surface that opposes infiltration (Figure 5.14) caused by the lower K_s in the BA horizon. A more severe drop in the infiltration rate occurs after about 0.4 h when the wetting front reaches the Bt2 horizon. After entering this horizon, the infiltration rate in the layered soil approaches a constant value. Thus, layers have an effect on the infiltration rate at the surface through their effect on movement of the wetting front.

A dry coarse-textured layer under a fine-textured layer will also impede deeper movement of the wetting front and reduce the infiltration rate but through a different mechanism. To illustrate this problem, a 2D water flow model, HYDRUS (2D/3D), is used for a soil profile consisting of a clay loam layer to a depth of 20 cm, underlain by a sand layer that extends to a depth of 50 cm. The van Genuchten (1980) hydraulic parameters are taken from the textural class averages of Carsel and Parrish (1988) (clay loam: $\theta_r = 0.095$, $\theta_s = 0.41$, $\alpha = 0.019\,cm^{-1}$, $n = 1.31$, $K_s = 0.26\,cm\,h^{-1}$, and $l = 0.5$; sand: $\theta_r = 0.045$, $\theta_s = 0.43$, $\alpha = 0.145\,cm^{-1}$, $n = 2.68$, $K_s = 29.7\,cm\,h^{-1}$, and $l = 0.5$). The model region is 100 cm in width and a point source of water (as might occur with drip or furrow irrigation) is placed at a distance of 50 cm from the left side ($h_0 = 0$). The initial pressure heads are −500 cm everywhere.

The distribution of water contents 10 h after the point source of water is started is shown in Figure 5.16. The location of the sand layer is apparent in that it has a lower water content ($\theta = 0.05\,cm^3\,cm^{-3}$) under the initial pressure head than the clay loam layer ($\theta = 0.25\,cm^3\,cm^{-3}$). At this early time, the wetting front has not reached the sand layer and the wetting front is semicircular in shape.

The distribution of water contents after 24 h is shown in Figure 5.17. The wetting front has become distorted because water cannot enter the dry sand layer. The reason for this is that pressure heads at the leading edge of the wetting front are negative and water cannot enter the smallest air-filled pores in the sand layer until pressure heads increase to the point where capillarity will draw water into these pores (the water entry pressure head for the sand). This stalls the wetting front until pressure heads rise to the critical level for entry. Since the pore-size distribution is narrower in the sand, it is not long after water first enters the sand that the pressure is high enough at the wetting front to fill the largest capillary pores in the sand. Once the sand is saturated, it no longer impedes flow because K_s is high in the sand compared to the clay layer above. Baver et al. (1972) referred to the action of a buried dry sand layer in temporarily impeding water flow and infiltration as a *check valve*. This principle is used in the design of golf greens, which have a sand surface layer over a coarse gravel layer at about 50 cm. The gravel layer prevents drainage from the root zone under frequent light irrigations, but if there is a heavy rain, the gravel layer will fill and drain the root zone, so that the green does not become waterlogged.

5.7.3 Infiltration Equations

Because of the importance of the infiltration process, simplified solutions to the Richards equation have been developed to predict infiltration. Most of the infiltration equations have been developed for conditions when rainfall does not limit infiltration. In this case, the infiltration rate is less than the rainfall or irrigation rate and runoff occurs. Soil hydraulic properties control the infiltration rate so it is *profile controlled* (Hillel, 2004), also called ponded infiltration, although the depth of ponding may be negligible if surface storage is small. Two examples are shown that represent very different approaches: the Green and

0.045 0.078 0.112 0.145 0.178 0.211 0.244 0.277 0.311 0.344 0.377 0.410

Water content - th[-], Min = 0.045, Max = 0.410
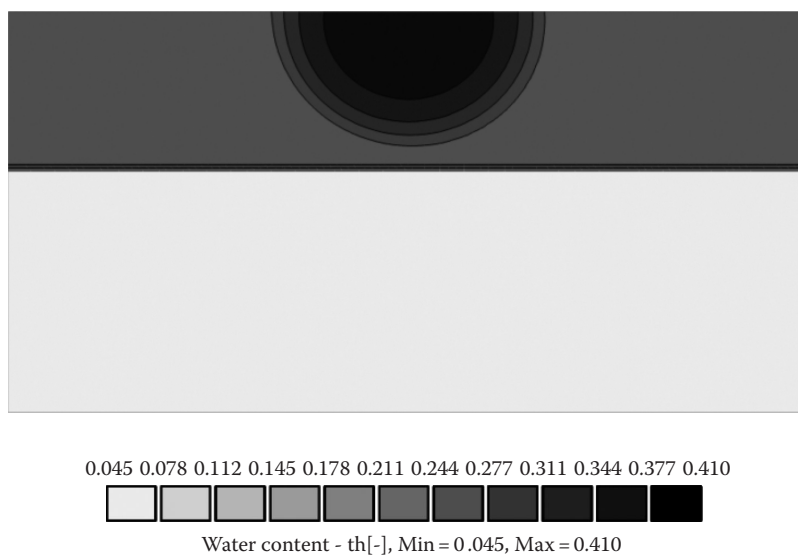
FIGURE 5.16 Distribution of water contents after 10 h in a soil profile consisting of a clay loam layer from 0 to 20 cm and a sand layer from 20 to 50 cm.

0.045 0.078 0.112 0.145 0.178 0.211 0.244 0.277 0.311 0.344 0.377 0.410

Water content - th[-], Min = 0.045, Max = 0.410
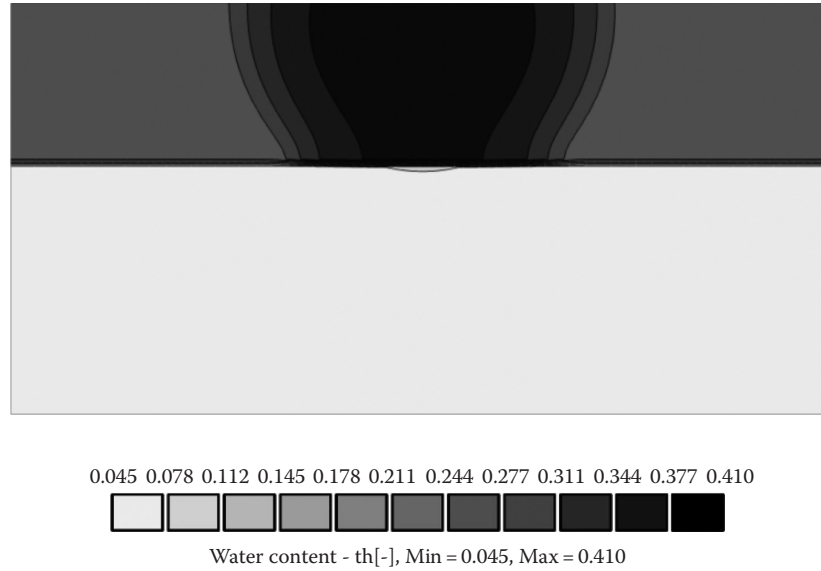
FIGURE 5.17 Distribution of water contents after 24 h in a soil profile consisting of a clay loam layer from 0 to 20 cm and a sand layer from 20 to 50 cm.

Ampt (1911) equations and the curve number (CN) method. Many infiltration equations have been developed including those by Kostiakov (1932), Horton (1940), Philip (1957), and Parlange et al. (1985).

5.7.3.1 Green–Ampt Equations

Green and Ampt (1911) developed a simplified mechanistic equation for infiltration by assuming that the wetting front in a soil was a square wave or sharp front. Although this is approximately true only in coarse-textured soils, there is no error in predicting the infiltration rate as long the amount of water behind the predicted square front is equal to the amount of new water behind the true wetting front.

The Green–Ampt equation for cumulative infiltration in a uniform soil, including the effect of gravity, is:

$$I(t) = K(h_0)t + \Delta h \Delta \theta \ln \left(1 + \frac{I(t)}{\Delta h \Delta \theta} \right), \qquad (5.44)$$

where

$$\begin{aligned} \Delta \theta &= \theta_0 - \theta_i, \\ \Delta h &= h_0 - h_f, \end{aligned} \qquad (5.45)$$

where

θ_i is the initial water content
θ_0 is the water content at the surface

The pressure head h_0 corresponds to the water content θ_0. The pressure head h_f is the pressure head at the wetting front. Under ponded conditions, θ_0 and $K(h_0)$ can be approximated by θ_s and K_s, respectively, and h_0 can be assumed to be zero (for negligible ponding) or the depth of ponded water. The unknowns in

Equation 5.44 are then $I(t)$ and h_f. White and Sully (1987) showed that the wetting front pressure head can be approximated using the macroscopic capillary length λ_c (Tables 5.1 and 5.2):

$$h_f = -\frac{\lambda_c}{2b}, \qquad (5.46)$$

where b is a dimensionless factor that has a theoretical range of $1/2$ to $\pi/4$ but can be assumed to be equal to 0.55 in most cases (Warrick and Broadbridge, 1992). Hence, the only unknown in Equation 5.44 is $I(t)$, but the equation cannot be solved directly for this variable because it appears both inside and outside the natural log function. Therefore, it must be solved iteratively. Alternatively, Equation 5.44 can be solved for t and the times corresponding to a range of values for cumulative infiltration can be calculated. In contrast to White and Sully (1987), Haverkamp et al. (1985) found that it was best to consider h_f a fitting parameter.

Once the cumulative infiltration curve as a function of time is known, the infiltration rate can be calculated as the instantaneous slope of this curve:

$$i(t) = \frac{dI(t)}{dt}. \qquad (5.47)$$

5.7.3.2 CN Approach

The CN method is an empirically determined rainfall–runoff relationship that provides an indirect estimate of the total depth of water that infiltrates during a storm (McCuen, 1982; NRCS, 2004). It does not give the infiltration rate $i(t)$ but simply the cumulative runoff R at the end of the storm. The method is based on numerous measurements of runoff for many soil types and considers soil texture, soil drainage class, antecedent moisture

conditions, and vegetative cover. The total runoff (*R* in cm) during an event is as follows:

$$R = \frac{(P - 0.2S)^2}{P + 0.8S}, \quad P > I_a, \quad R = 0, \quad P \leq I_a, \qquad (5.48)$$

where

> *P* is the total precipitation in cm
> *S* is the *water storage capacity* of the soil in cm
> I_a is the *initial abstraction* in cm

The water storage capacity is calculated using the CN:

$$S = \frac{2540}{\text{CN}} - 25.4. \qquad (5.49)$$

CNs are tabulated in Chapter 9 of *The National Engineering Handbook* (NRCS, 2004) and vary with land use, ranging from near 0 for a dry, fully vegetated, highly permeable surface to near 100 for an impervious surface. The CN method assumes that an initial amount of rain must occur before runoff starts and this is called the initial abstraction I_a, which is assumed to be equal to 0.2*S*. Runoff is zero until precipitation exceeds I_a (see Equation 5.48).

The CN also varies with the soil hydrologic group, which ranges from Group A (soils with low runoff potential when thoroughly wet) to Group D (soils with high runoff potential when thoroughly wet). Hydrologic groups are described in Chapter 7 of the *National Engineering Handbook* (NRCS, 2004) and they are also given for all U.S. soil series in the National Soil Information System (NASIS; NRCS, 2009). This information is also available through the Web Soil Survey at http://websoilsurvey.nrcs.usda.gov/app. The CN can be adjusted for antecedent runoff conditions I through III, representing dry, normal and wet conditions but some studies have found that this does not improve the predictions (see Chapters 9 and 10, NRCS, 2004). With the CN approach, there is an emphasis on management and cover. This is a recognition of the importance of crusting (and impervious surface percentage) on infiltration. The use of soil hydrologic groups is also a recognition of the importance of the soil profile (e.g., subsurface impeding horizons), not just the hydraulic properties of the surface layer.

A distinction can be made between *infiltration-excess runoff* and *saturation-excess runoff* (Beven, 2001). Infiltration-excess runoff occurs when downward movement of water into the soil is less than the rainfall rate. This is also called Hortonian flow after Horton (1933). This type of process is described by the Green–Ampt (1911) equations. Saturation-excess runoff occurs when water moves laterally within the soil during a storm and causes lower hillslope positions to be saturated, thereby decreasing the infiltration rate to less than the infiltration capacity. This is also called the *variable-source-area* concept (Hewlett and Troendle, 1975) because the area where runoff occurs expands with the size of the storm as the soil landscape becomes wetter. The CN approach can be considered a saturation-excess approach (Steenhuis et al., 1995). Another way to look at the difference between these processes is to consider the gradients and conductivities (components of Darcy's law, see

Equation 5.42). In infiltration-excess runoff, there is a large gradient favoring infiltration, but the soil conductivity is not sufficiently high to prevent runoff. In saturation-excess runoff, the conductivity may be very high, but the gradient is low or even favors exfiltration (a seep). A final distinction is that the infiltration-excess approach is essentially a 1D (vertical) view of the infiltration process. The saturation-excess approach is a 3D view.

5.7.4 Borehole Infiltration

Infiltration from a borehole is a 3D flow problem (see Figure 5.8). It can be modeled in two dimensions, however, using cylindrical coordinates available in HYDRUS (2D/3D).

Infiltration is simulated from an unlined borehole with a constant head of water into a loamy sand. The van Genuchten (1980) hydraulic properties are taken from the Rosetta Lite database in HYDRUS: $\theta_r = 0.049$, $\theta_s = 0.39$, $\alpha = 0.0347 \, \text{cm}^{-1}$, $n = 1.75$, and $K_s = 4.38 \, \text{cm h}^{-1}$. Since infiltration from a borehole will be symmetrical around the vertical centerline of the borehole (in the absence of any slope or heterogeneities in the soil surrounding the borehole), a 2D slice through the soil from the centerline of the borehole on the left radiating out into the soil to the right is simulated. The horizontal dimension of the model space is 100 cm and the vertical dimension is also 100 cm with the soil surface at the top (Figure 5.18). A borehole with a radius of 5 cm penetrates to a depth of 30 cm below the soil surface. The boundary condition at the bottom of the model space (at a depth of 100 cm below the soil surface) is a free-drainage boundary condition simulating a deep water table (see Section 5.4.1.3). To simulate water ponded in the borehole to a height 10 cm above the bottom of the borehole, a constant-head boundary condition is used. Along the boundary representing the bottom of the borehole and up the side wall to a height of 10 cm, the boundary condition is a constant pressure head of 10 cm in equilibrium with the lowest nodal point on this boundary. This ensures that the pressure at the bottom of the borehole is 10 cm and pressure decreases linearly up the side wall to a value of 0 at 10 cm above the borehole bottom. All other points along the model-space boundary are a no-flux boundary condition. The distribution of pressure heads after 3 h of simulation are shown in Figure 5.18. Capillarity has drawn the wetting front into the soil in all directions including upward toward the soil surface. The influence of gravity is most apparent in the saturated region immediately surrounding the ponded portion of the borehole where a teardrop shape to the contour is apparent.

The infiltration rate from the borehole as a function of time is shown in Figure 5.19. The volumetric infiltration rate (*Q* in cm^3 h^{-1}) has been converted to the infiltration flux (*i* in cm h^{-1}) by dividing by the bottom area of the borehole (πr^2). The infiltration rate declines to a near-steady value within one-half hour, much like 1D infiltration into a uniform soil with water ponded at the surface (see Figure 5.11). Unlike 1D infiltration, however, the steady value in the borehole infiltration example is much greater than the soil K_s (4.38 cm h^{-1}). This is because flow is occurring in three dimensions (Darcy's law for 1D flow does not apply) and because the gradient is much more than a unit gradient. This is why the borehole equation (Equation 5.22) must be used to calculate K_s.
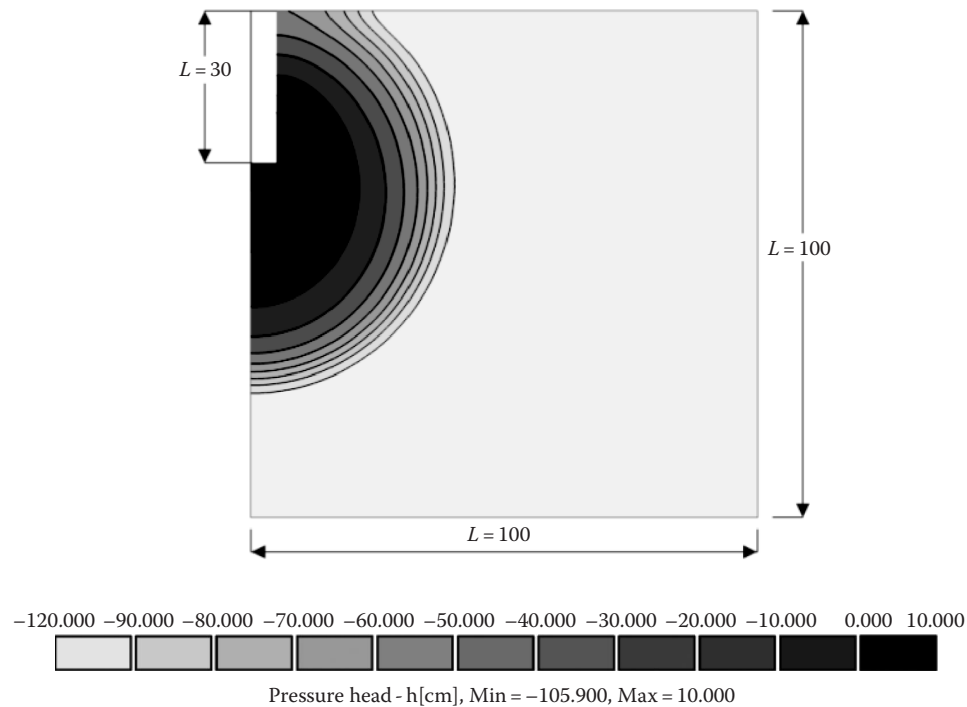
FIGURE 5.18   Distribution of pressure heads after 3 h in the borehole infiltration example.
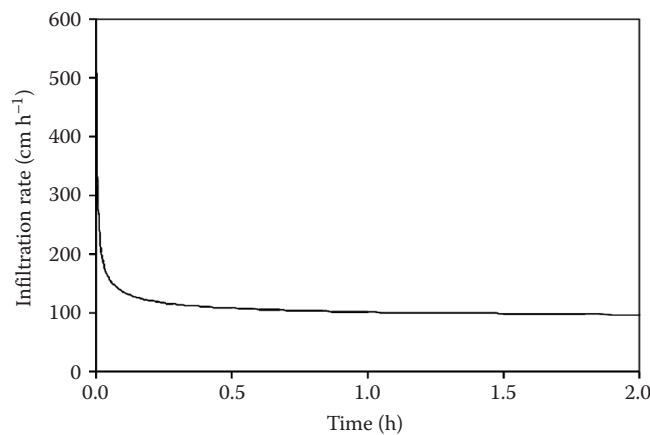
FIGURE 5.19   Infiltration rate as a function of time in the borehole infiltration example.

5.7.5 Subsurface Irrigation

Irrigation using a buried irrigation line results in flow of water that can be modeled as 2D infiltration. Skaggs et al. (2004) compared measured water contents in a Hanford sandy loam soil with those simulated using HYDRUS-2D. Soil properties were estimated using the Rosetta database for van Genuchten (1980) parameters (θ_r = 0.021, θ_s = 0.34, α = 0.023 cm^{-1}, n = 1.4, K_s = 1.6 cm h^{-1}, and l = −0.92). The irrigation line was buried at a depth of 5 cm. The model space consisted of a 2D soil block, 50 cm wide and 60 cm deep with the soil surface at the top (Figure 5.20). It was assumed that water flow would be
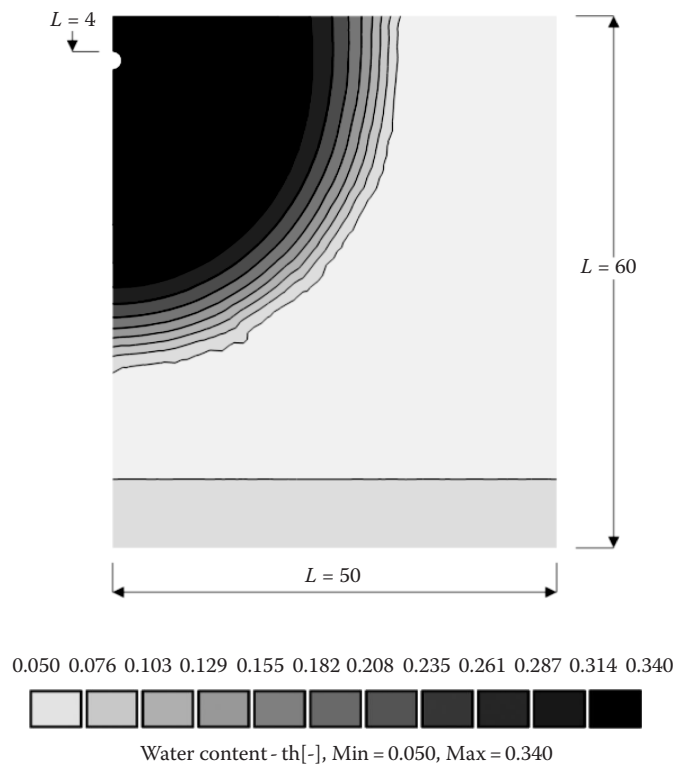
FIGURE 5.20   Distribution of water contents around a subsurface irrigation line after 10 h. (From Skaggs, T.H., T.J. Trout, J. Šimůnek, and P.J. Shouse. 2004. Comparison of HYDRUS-2D simulations of drip irrigation with experimental observations. J. Irrigat. Drain. Eng. 130:304–310.)

symmetrical about the vertical centerline of the irrigation pipe so only half the area around the pipe was simulated. All of the boundaries were considered to be zero-flux except for a 2-cm diameter semicircle on the left boundary, which represented the irrigation line. The boundary condition on this semicircle was specified as a flux of 6.37 cm h^{-1} (the irrigation rate in cm^3 h^{-1} for a given segment of irrigation line divided by one-half the surface area of the irrigation line segment). Water contents are shown after 10 h of simulated irrigation in Figure 5.20. The wetting front has moved out in a nearly cylindrical manner due to capillarity but some distortion of the wettest contour due to gravity is apparent. The HYDRUS simulations agreed quite well with the measured distributions of water contents (Skaggs et al., 2004).

5.8 Redistribution

Once a rainfall event is over, water movement does not cease. For example, the final distribution of pressure heads in Figure 5.12 will continue to change after 2 h even if there is no further precipitation. This process is known as *redistribution* or drainage (Jury and Horton, 2004). In theory, with a water table at a depth of 250 cm, pressure heads will return to the initial distribution of an equilibrium profile with pressure heads increasing linearly from −250 cm at the surface to 0 cm at the water table (Figure 5.12). In reality, it takes so long for redistribution to occur that an equilibrium profile rarely occurs unless the water table is at a very shallow depth. To see this, the previous infiltration simulation shown in Figure 5.12 is extended to 1 week (168 h) with no precipitation (or evapotranspiration) beyond 2 h using HYDRUS-1D.

The resulting pressure head distribution is shown in Figure 5.21. The pressure head at time $T1 = 2$ h shows the depth to
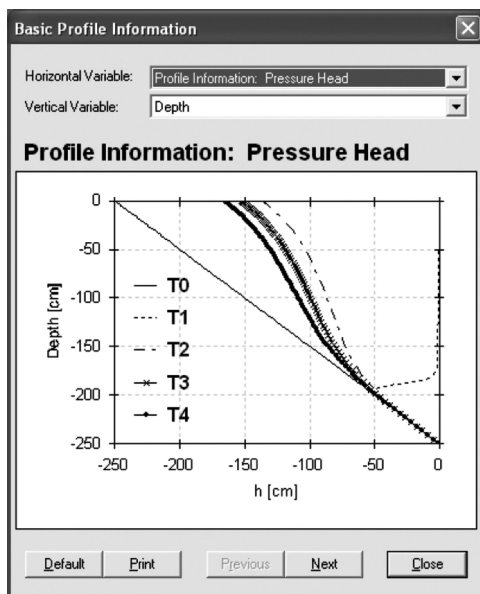
which the wetting front moved (about 200 cm) during the rainfall event and is the last time in the earlier infiltration example (Figure 5.12). It is clear that pressure heads continued to change beyond 2 h with pressure heads decreasing at more shallow depths. The curves are moving in the direction of the equilibrium distribution (the initial conditions in Figure 5.21), but even after a week of redistribution they are far from a true equilibrium distribution. This is because as pressure heads decrease at shallow depths, unsaturated hydraulic conductivity decreases and water movement is slower. The average pressure head in the profile after about a week is close to −100 cm, which is the pressure head often used to estimate field capacity in coarse-textured soils (Romano and Santini, 2002).

5.9 Evaporation

Evaporation is the reverse process of infiltration: a flux *out of* the soil at the surface. HYDRUS-1D is used to illustrate this process with the same loamy sand employed for the borehole example ($\theta_r = 0.049$, $\theta_s = 0.39$, $\alpha = 0.0347$ cm^{-1}, $n = 1.75$, and $K_s = 4.38$ cm h^{-1}, $l = 0.5$). The surface boundary condition is a system-dependent boundary condition for evaporation: an evaporative flux of 1 cm day^{-1} (0.042 cm h^{-1}) until the surface dries to a pressure head of −15,000 cm ($E = 0.042$ cm h^{-1}, $h_A = -15,000$ cm, see Equations 5.34 and 5.35). At that time, the boundary condition switches to a constant pressure head of −15,000 cm. The resulting water content profiles for different times after evaporation started are shown in Figure 5.22.

The numerical solution shows that water contents decrease with depth as time progresses due to the loss of water through evaporation at the surface. Instead of a wetting front seen in

FIGURE 5.21 Simulated pressure head distributions in the redistribution example: $T0 = 0$ h, $T1 = 2$ h, $T2 = 84$ h, $T3 = 126$ h, and $T4 = 168$ h.
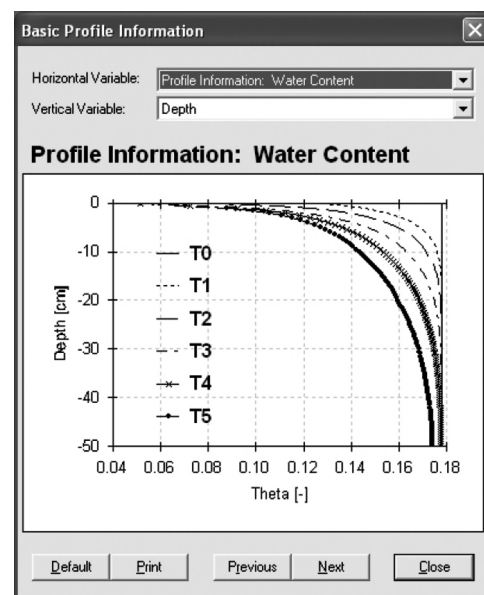
FIGURE 5.22 Water contents as a function of depth for various times during evaporation from a uniform loamy sand: $T0 = 0$ h, $T1 = 3$ h, $T2 = 6$ h, $T3 = 12$ h, $T4 = 24$ h, and $T5 = 48$ h.
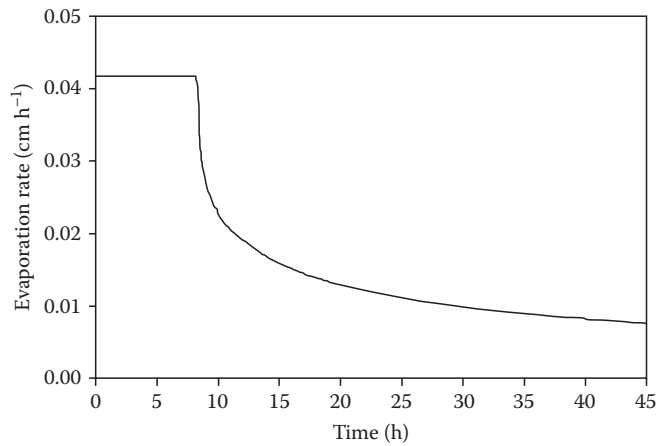
FIGURE 5.23 Evaporation rate as a function of time in a uniform loamy sand.

the infiltration example, in this case, a drying front progresses deeper into the soil. The evaporation rate for the same soil is shown in Figure 5.23. Evaporation is high initially and constant. During this period, evaporation from the soil is able to keep up with the potential evaporation rate of 0.042 cm h^{-1}. The reason for the high initial evaporation rate is the steep gradient in water potential between the wet soil and the dry air above the soil and the high hydraulic conductivity of the wet soil surface layer (Figure 5.22). During this period, the decrease in the hydraulic conductivity of the surface layer is fully compensated by the increase in the hydraulic gradient. This period is referred to as *stage one* evaporation (Hillel, 2004). After about 10 h, the evaporation rate starts to drop. At this point, the soil surface has dried completely (reached the minimum pressure head of −15,000 cm) and the actual evaporation rate becomes less than the potential rate. The evaporation rate decreases with time because the hydraulic gradient at the surface gradually decreases. This period of decreasing evaporation rate is referred to as *stage two* evaporation. Eventually, the evaporation rate approaches a constant residual rate where most of the flux at the surface is in the vapor phase. This final period is referred to as *stage three* evaporation.

The heat and radiant energy balance at the soil surface determines the potential evaporation rate ET_0. This varies from day-to-day and even minute-to-minute as wind speed, radiation, and relative humidity change. The actual evapotranspiration rate ET is usually less than ET_0. Part of the evapotranspiration demand is satisfied by soil evaporation. As can be seen in Figure 5.23, evaporation can be high when the soil is wet (e.g., shortly after a rainfall event) but rapidly drops to a very low rate. Thus, evaporation alone usually will not meet the potential evapotranspiration demand ET_0.

5.10 Transpiration

5.10.1 Soil–Plant–Atmosphere Continuum

Where nondormant plants are present, most of the evaporative demand is satisfied by transpiration. Photosynthesis uses only about 1% of the water transpired by plants (Hillel, 2004).

For photosynthesis to occur, however, CO_2 must enter the plant leaves and reach chloroplasts. As a result, plants must maintain open stomata, which allow CO_2 to enter the interior of leaves (Figure 5.1). Plant cells must be moist to function so the open stomata result in a continuous loss of water from the interior of the leaf through the stomata due to evaporation. This water must be replenished by plant uptake of water through the root system or the plant leaf cells will desiccate and the plant will wilt.

Water moves through the plant to the atmosphere along a continuous path. This is called the *soil–plant–atmosphere continuum* (Philip, 1966). Water moves from the soil across the plant root to the xylem vessels. It then moves up the xylem vessels to the plant leaves. In the leaves, water exits the xylem vessels and moves through leaf cells to the stomatal cavities where it evaporates. A steady gradient of decreasing water potential pulls water from the soil through the plant to the atmosphere. Water is drawn first from the surface horizons where roots are more numerous. As this water is depleted, a plant draws water from deeper in the soil, but this requires more of an energy gradient due to the effect of gravity. Also root densities (length of root per volume of soil) decrease with depth.

There is a limiting leaf-water potential, below which stomatal guard cells begin to lose turgor and close. As soon as this happens, the actual evapotranspiration rate drops below the potential rate: $ET < ET_0$. Stomata can be partially closed, which reduces transpiration but does not completely stop it. If stomata close completely then $ET = 0$.

5.10.2 Root Water Uptake

Root water uptake can be considered a sink term in the Richards equation (Equation 5.26). Various terms have been developed including those by Molz (1981), Jarvis (1989), and Vrugt et al. (2001). Feddes et al. (1978) modeled root water uptake as a function of soil water pressure head h:

$$S(h) = \alpha(h)S_p, \qquad (5.50)$$

where

$S(h)$ is the root water uptake or volume of water removed from a unit volume of soil per unit time (T^{-1})

S_p is the potential water uptake rate (T^{-1})

$\alpha(h)$ is a dimensionless stress response function of the pressure head ($0 \leq \alpha \leq 1$)

The stress response function is shown in Figure 5.24a. Water uptake is assumed to be 0 close to saturation due to a lack of oxygen in the root zone (pressure head greater than h_1). For a pressure head less than h_4 (the wilting-point pressure head), water uptake is also assumed to be zero. Water uptake is optimal between pressure heads of h_2 and h_3. The pressure head h_3 may be adjusted depending on the transpiration rate so that it is more negative when transpiration rates are low (optimal root water uptake occurs over a wider range in pressure head at lower transpiration rates).
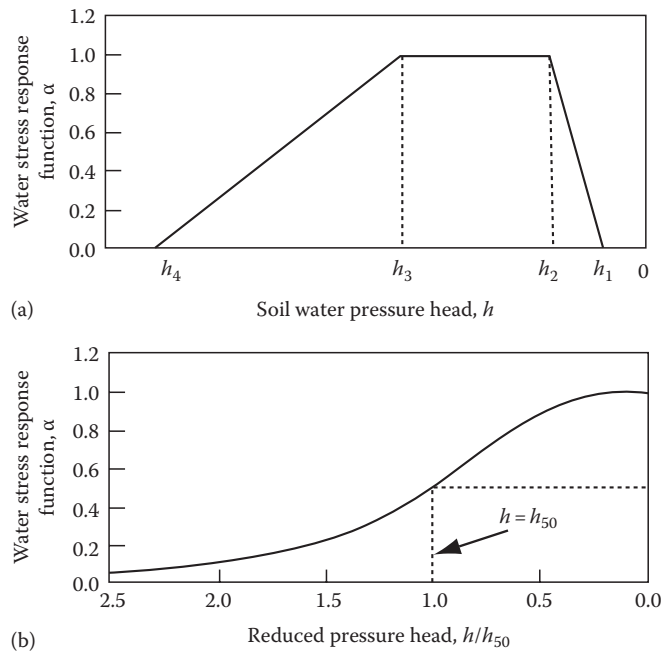
FIGURE 5.24 Plant-water stress response function $\alpha(h)$ as used by (a) Feddes et al. (1978) and (b) van Genuchten (1987).

FIGURE 5.25 Precipitation (a negative surface flux) as a function of time in the transpiration example.

Van Genuchten (1987) suggested an *S*-shaped root water stress response function (Figure 5.24b). The parameter h_{50} represents the pressure head at which root water uptake is reduced by 50% during conditions of negligible osmotic stress. Note that in contrast to the response function of Feddes et al. (1978), this response function does not consider the reduction near saturation due to low oxygen levels. Van Genuchten (1987) also expanded the root water stress response function to include osmotic effects.

HYDRUS-1D can be used to simulate evapotranspiration. The same loamy sand as in the previous examples ($\theta_r = 0.049$, $\theta_s = 0.39$, $\alpha = 0.0347\,cm^{-1}$, $n = 1.75$, and $K_s = 105.12\,cm\,day^{-1}$) is used to simulate plant transpiration. The simulation is run for 1 year using precipitation and potential evapotranspiration from Athens, Georgia, for 1995, when total precipitation was 141 cm (slightly greater than the long-term average of 127 cm). The simulation is for a grass lawn that is not dormant in winter.

A system-dependent atmospheric boundary condition with surface runoff is used for the upper boundary condition to simulate infiltration and evaporation. The free-drainage boundary condition is used to simulate a deep water table for the lower boundary condition (see Section 5.4.1.3). The Feddes et al. (1978) water uptake reduction model (Equation 5.50) is used with threshold pressure heads appropriate for grass: $h_1 = -10\,cm$, $h_2 = -25\,cm$, $h_3 = -300\,cm$ for high transpiration rates and $-1000\,cm$ for low transpiration rates, and $h_4 = -15,000\,cm$ (see Figure 5.24a). A full canopy is assumed and consequently there is no evaporation. A root distribution that decreases linearly from the soil surface to a depth of 100 cm is specified.

The potential surface flux as a function of time is shown in Figure 5.25. Since there is no evaporation, the only potential surface flux is precipitation, so this is a plot of daily rainfall
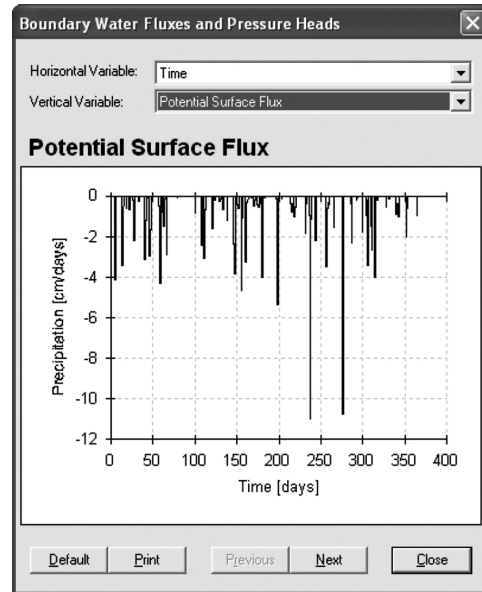
(as negative fluxes) for 1995 in Athens, Georgia. There is a dry period in early spring and there are large rainfall events in the early fall.

Potential and actual evapotranspiration (root water uptake in this case) are shown in Figure 5.26. It is clear that potential evapotranspiration is low during the winter months and reaches peak values of nearly 1 cm day^{-1} in summer. Actual evapotranspiration is identical to potential evapotranspiration until about day 80 when the first dry spell starts. At this point, actual transpiration drops below the potential rate due to plant-water stress and remains below the potential rate for much of the summer. In the fall, when potential evapotranspiration rates decrease and the soil profile water contents start to increase, actual and potential evapotranspiration rates again coincide.

The pressure head at a depth of 5 cm is shown in Figure 5.27. When the first dry period occurs in the early spring, the pressure head at this depth drops quickly to the minimum value of $-15,000\,cm$ (wilting point) due to root uptake. The pressure head at a depth of 50 cm (Figure 5.28) also drops to the wilting point but the drop occurs later. With the rainfall that starts shortly after 100 days, pressure head near the surface recovers, but the pressure head at 50 cm remains at wilting point because this rainfall does not penetrate to deeper depths in the soil profile. At about 120 days, there has been sufficient rainfall that pressure heads at both depths return to near zero. However, the pressure head at the deeper depth soon returns to wilting point because water from small rainfall events that keep the shallow depths relatively wet do not penetrate to deeper depths. A higher rainfall period from about 140 to 180 days results in pressure heads near zero at both depths. After about 180 days, the pressure head at the deeper depth returns to wilting point and stays there until the large rainfall event on about day 230.
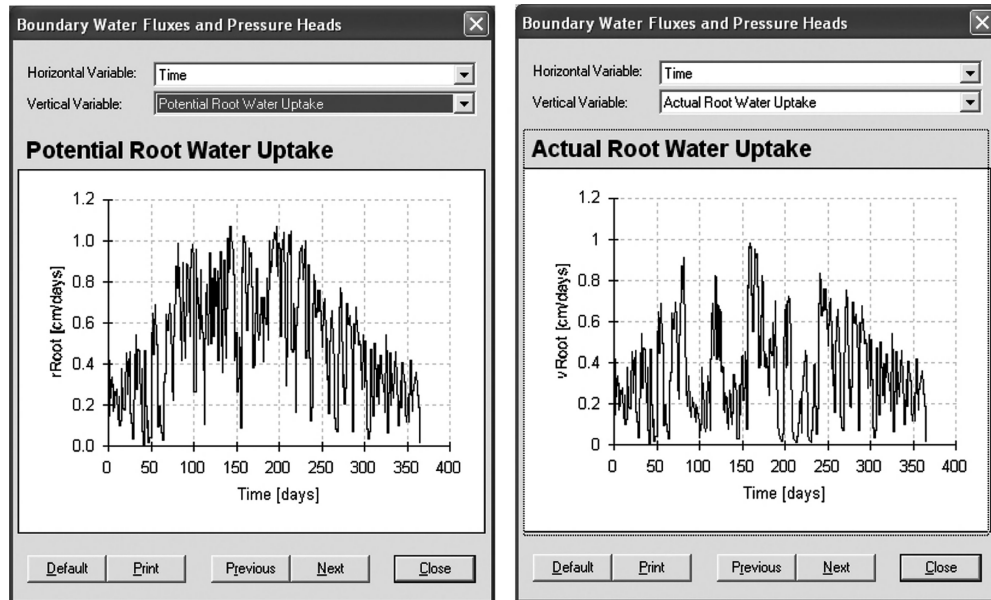
FIGURE 5.26   Potential root water uptake (left) and actual root water uptake (right) of water as a function of time in the transpiration example.
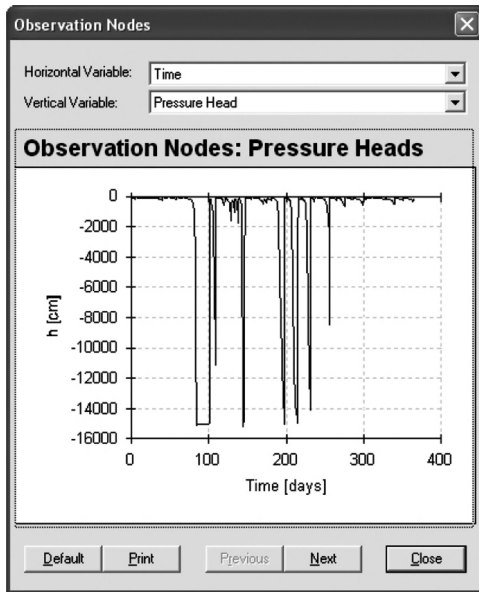
FIGURE 5.27   Pressure head as a function of time at a depth of 5 cm in the transpiration example.

FIGURE 5.28   Pressure head as a function of time at a depth of 50 cm in the transpiration example.

In contrast, the pressure head near the surface responds to the small rainfall events during this interval. Late in the season, a large rainfall event on about day 280 and lower potential evapotranspiration rates result in pressure heads at both depths returning to near zero.

Overall, these results show that even in a slightly wetter than normal year in a region of the United States with relatively high rainfall, this lawn experienced considerable water stress during the summer. This is due in large part to the low plant-available water in such a coarse-textured soil.

5.11 Preferential Flow

Increasing evidence exists that variably saturated flow in many field soils is not consistent with the uniform flow pattern typically predicted with the Richards equations (Quisenberry and Phillips, 1976; Luxmoore, 1981; Flury et al., 1994; Hendrickx and Flury, 2001). This is due to the presence of macropores, fractures, or other structural voids or biological channels through which water and solutes may move preferentially while bypassing a large

part of the matrix pore space. *Preferential flow* and transport processes are probably the most frustrating processes hampering accurate predictions of contaminant transport in soils and fractured rocks. Contrary to uniform flow, preferential flow results in irregular wetting of the soil profile as a direct consequence of water moving faster in certain parts of the soil profile than in others. Hendrickx and Flury (2001) defined preferential flow as constituting all phenomena where water and solutes move along certain pathways while bypassing a fraction of the porous matrix. Water and solutes for these reasons can propagate quickly to far greater depths than would be predicted with the Richards equation describing uniform flow.

The most important causes of preferential flow are the presence of *macropores* and other structural features, development of flow instabilities (i.e., *fingering*) caused by profile heterogeneities or water repellency (Hendrickx et al., 1993), and *funneling* of flow due to the presence of sloping soil layers that redirect downward water flow. While the latter two processes (i.e., flow instability and funneling) are usually caused by textural differences and other factors at scales significantly larger than the pore scale, macropore flow and transport are usually generated at the pore or slightly larger scales, including scales where soil structure first manifests itself (i.e., the pedon scale; Šimůnek et al., 2003).

5.11.1 Macropores

Uniform flow in granular soils and preferential flow in structured media (both macroporous soils and fractured rocks) can be described using a variety of multipermeability models. While *single-permeability* models assume that a single-pore system exists that is fully accessible to water, *dual-permeability* models assume that the porous medium consists of two interacting pore regions, one associated with the interaggregate, macropore, or fracture system, and one comprising the micropores (or intraaggregate pores) inside soil aggregates or the rock-soil matrix.

The simplest formulation for water flow is a single-permeability (equivalent porous medium) model applicable to uniform flow in soils. Other models apply in some form or another to preferential flow or transport. Dual-permeability models are those in which water can move in both the inter- and intraaggregate pore regions (or matrix and fracture domains). These models in various forms are now also becoming increasingly popular (Pruess and Wang, 1987; Gerke and van Genuchten, 1993). They differ mainly in how they implement water flow in and between the two pore regions. Common approaches to calculating water flow in macropores or interaggregate pores use Poiseuille's equation (Ahuja and Hebson, 1992), the Green and Ampt or Philip infiltration models (Chen and Wagenet, 1992), the kinematic wave equation (Germann, 1985), or the Richards equation (Gerke and van Genuchten, 1993). Multipermeability models are based on the same concept as dual-permeability models but include additional interacting pore regions. These models can be readily simplified to the dual-permeability approaches. Recent reviews of preferential flow processes and available mathematical models are provided by Hendrickx and Flury (2001), Šimůnek et al. (2003), and Jarvis (2007).

Another way to simulate the effect of macropores on water flow is through the use of *scaling*. This is a procedure designed to simplify the description of the spatial variability in the soil hydraulic properties. Here, we describe scaling as implemented in the HYDRUS models. It is assumed that variability in the hydraulic properties of a given soil profile can be approximated by means of a set of linear scaling transformations, which relate the soil hydraulic characteristics $\theta(h)$ and $K(h)$ at a point in the soil to reference characteristics $\theta^\star(h^\star)$ and $K^\star(h^\star)$. The technique is based on the similar media concept introduced by Miller and Miller (1956) for porous media, which differ only in the scale of their internal geometry. The concept was extended by Simmons et al. (1979) to materials that differ in morphological properties, but that exhibit scale-similar soil hydraulic functions. Three independent scaling factors are employed in HYDRUS. These three scaling parameters may be used to define a linear model of the actual spatial variability in the soil hydraulic properties as follows (Vogel et al., 1991):

$$K(h) = \alpha_K K^\star(h^\star),$$
$$\theta(h) = \theta_r + \alpha_\theta [\theta^\star(h^\star) - \theta_r^\star],$$
$$h = \alpha_h h^\star,$$

(5.51)

in which, for the most general case, α_K, α_θ, and α_h are mutually independent scaling factors for the hydraulic conductivity, water content, and the pressure head, respectively. Less general scaling methods arise by invoking certain relationships between α_K, α_θ, and α_h. In HYDRUS, scaling factors for each node can be generated stochastically by specifying the mean and standard deviation of the factors. Spatial correlation of the scaling factors can also be included.

As an example of scaling, HYDRUS (2D/3D) is used to simulate infiltration under conditions of incipient ponding of water at the surface into a 2D block of soil 300 cm in width and 100 cm deep. The soil is a loam with hydraulic properties taken from the Rosetta Lite database ($\theta_r = 0.061$, $\theta_s = 0.399$, $\alpha = 0.011$ cm^{-1}, $n = 1.474$, and $K_s = 0.50$ cm h^{-1}). A standard deviation of 1.0 is specified for log α_K with a correlation length 10 times larger in the vertical direction than the horizontal direction. The spatial distribution of K_s for a given realization is shown in Figure 5.29. In this case, the stochastic assignment of scaling factors has produced two zones of high K_s and the larger correlation length in the vertical direction has stretched these zones along the *z*-axis. The distribution of water contents in this soil profile 0.3 h after infiltration began is shown in Figure 5.30. Preferential flow in the two zones causes the wetting front to advance further in these areas.

5.11.2 Fingering and Funnel Flow

Fingering and funnel flow occur in layered soil profiles where a fine-textured layer overlies a coarse-textured layer. Fingering can occur when there is no slope to the coarse-textured layer. In an experiment to examine fingering, Baker and Hillel (1990) used a plexiglass chamber 75 cm wide, 52 cm high, and 1.3 cm thick. The top layer was fine sand with particle sizes between 45 and

| 0.068 | 11.001 | 21.935 | 32.868 | 43.801 | 54.735 | 65.668 | 76.601 | 87.534 | 98.468 | 109.401 | 120.334 |

Min = 0.068, Max = 120.334
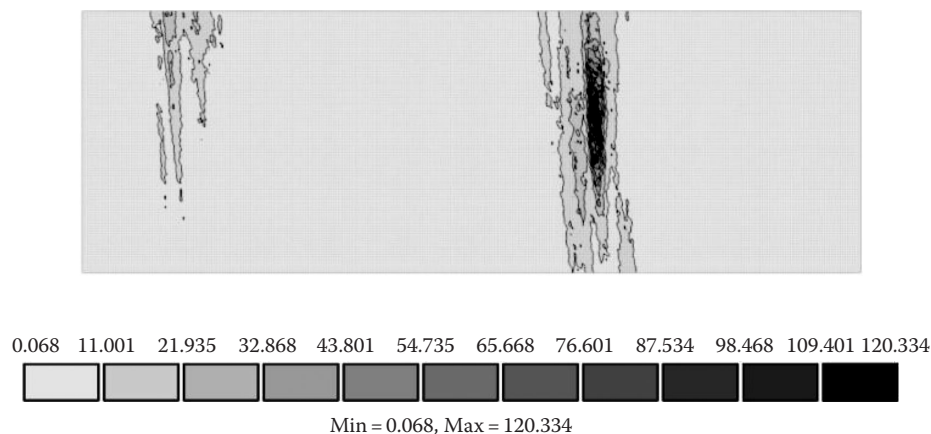
FIGURE 5.29   Spatial distribution of K_s (cm h^{-1}) generated by HYDRUS (2D/3D) using scaling to simulate preferential flow in a loam soil block that is 300 cm in width and 100 cm in depth.

| 0.324 | 0.331 | 0.338 | 0.345 | 0.351 | 0.358 | 0.365 | 0.372 | 0.379 | 0.385 | 0.392 | 0.399 |

Water content - th[-], Min = 0.324, Max = 0.399
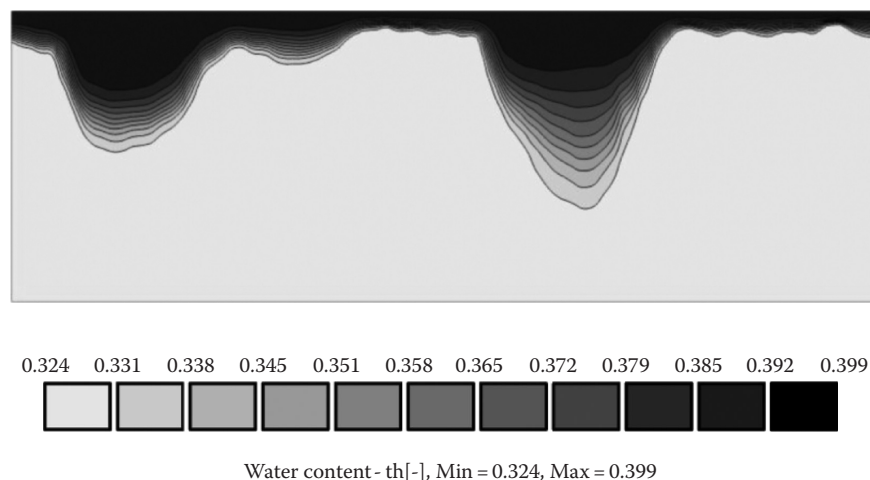
FIGURE 5.30   The effect of preferential flow on water contents 0.3 h after infiltration began simulated by HYDRUS (2D/3D) in a loam soil block that is 300 cm in width and 100 cm in depth.

106 μm. The sublayer consisted of a coarse-textured sand with a particle-size range that varied from experiment to experiment.

When the wetting front reached the coarse layer, water could not enter because the wetting front pressure head was less than the water entry value for the coarse sublayer (as shown in the HYDRUS simulation in Figure 5.17). The wetting front would pause at the interface and over time the wetting front pressure head would increase. When the wetting front pressure head increased to the value required to enter the smallest pore in the coarse sublayer, water started to infiltrate. Eventually, the wetting front pressure head increased to a value where the flux through the coarse sublayer was equal to or greater than the flux through the top layer. Baker and Hillel (1990) called this the effective water entry value h_e. If the flux was greater through the coarse-textured sublayer than through the fine-textured top layer, then the only way for flow in the two layers to be equal was for water to flow through a fraction of the sublayer in discreet *fingers* (Figure 5.31). The effective water entry value corresponded with the inflection point on the water release curve. This represented the dominant pore size. The fraction of wetted
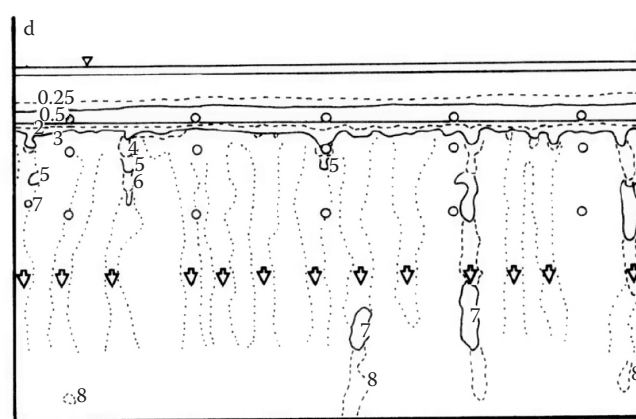
FIGURE 5.31   Tracings (alternating solid and dashed lines) of wetting front position with time during infiltration in the experiment by Baker and Hillel (1990). The numbers in the figure show the time (minutes) of the wetting fronts. The sublayer in this case consisted of a coarse sand with particle sizes between 500 and 710 μm. (From Baker, R.S., and D. Hillel. 1990. Laboratory tests of a theory of fingering during infiltration into layered soils. Soil Sci. Soc. Am. J. 54:20–30.)
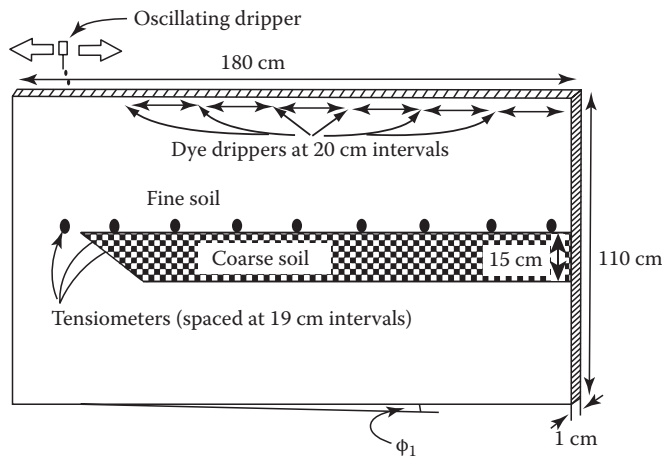
FIGURE 5.32 Schematic of the experimental setup of a plexiglass chamber 180 cm wide, 110 cm tall, and 1 cm thick, filled with fine soil and a discontinuous 15 cm thick coarse layer used by Walter et al. (2000). (From Walter, M.T., J.-S. Kim, T.S. Steenhuis, J.-Y. Parlange, A. Heilig, R.D. Braddock, J.S. Selker, and J. Boll. 2000. Funneled flow mechanisms in a sloping layered soil: Laboratory investigations. Water Resour. Res. 36:841–849.)

soil in the coarse sublayer at steady state was equal to the ratio of the flux through the top layer and $K(h_e)$ in the sublayer.

When the coarse sublayer has a slope, *funneling* occurs. In another experiment, Walter et al. (2000) packed fine sand into a plexiglass chamber 1 cm in thickness with a discontinuous, embedded, layer of coarse sand (Figure 5.32). They investigated the effect of slope and infiltration rate. Different color dyes were applied to intervals at the infiltration surface once steady flow was achieved to show stream lines.

During infiltration when the wetting front reached the coarse layer, it could not enter this layer because the wetting front pressure heads were less than the water entry value, just as shown in the HYDRUS (2D/3D) simulation (Figure 5.17). Since there was a slope to the layer, water funneled laterally to the end of the layer before penetrating deeper into the chamber (Figure 5.33).
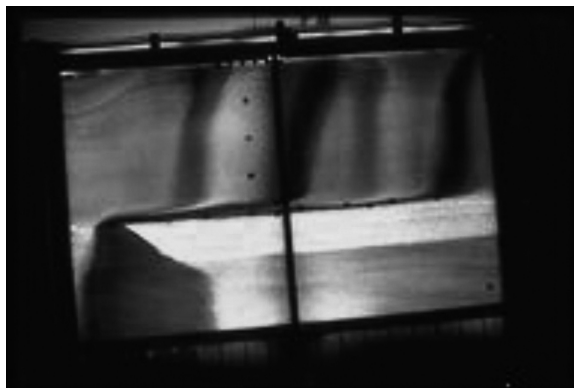
FIGURE 5.33 Photograph of an experimental run with an angle of 7.1° and an infiltration rate of 120 mm day⁻¹ by Walter et al. (2000). (From Walter, M.T., J.-S. Kim, T.S. Steenhuis, J.-Y. Parlange, A. Heilig, R.D. Braddock, J.S. Selker, and J. Boll. 2000. Funneled flow mechanisms in a sloping layered soil: Laboratory investigations. Water Resour. Res. 36:841–849.)

Above the coarse layer, pressure head increased at the interface in the downslope direction because more and more water was being diverted. If the infiltration rate was high enough, pressure head reached the water entry value and infiltration occurred before the end of the embedded layer was reached. At the tip of the embedded layer was a toe area where water never infiltrated the coarse layer. This was because the pressure was relieved by flow around the toe. Thus, funneling above sloping coarse layers results in water (and solutes) penetrating only a portion of a soil profile and is therefore a type of preferential flow (Kung, 1990a, 1990b).

5.12 Groundwater Recharge and Discharge

The processes of groundwater recharge, groundwater discharge, the position of the water table, and streamflow are all interrelated in a humid environment (Fetter, 1988). This can be seen by simulating the wetting of an initially dry hillslope using HYDRUS (2D/3D). In this case, the 2D model space is much larger than any of the earlier simulations. The hillslope is 100 m long in the horizontal direction and extends from a ridgetop on the left to a stream channel on the right (Figure 5.34). At the ridgetop, the soil surface is 50 m above the bottom of the model space. At the stream channel, the soil surface is 20 m above the bottom of the model space. The stream channel is 5 m deep (indicated by the nickpoint in Figure 5.34). The soil surface is given a convex shape. The boundary condition at the soil surface is a constant flux of 2 cm day⁻¹. The boundary condition on the left is a no-flux condition appropriate for a ridgetop, where no lateral flow (only recharge) is expected. The bottom boundary condition is also a no-flux condition indicating impermeable rock. The boundary condition beneath the stream channel where no lateral flow (only discharge) is expected is a no-flux boundary condition. The boundary condition at the stream channel is a seepage condition, a type of system-dependent condition discussed previously in Section 5.4.1.3. The initial pressure heads are −100 m everywhere in the hillslope. The hydraulic properties of the hillslope are that of a loam soil textural class in the Carsel and Parrish (1988) database ($\theta_r = 0.078$, $\theta_s = 0.43$, $\alpha = 0.036$ cm⁻¹, $n = 1.56$, and $K_s = 24.96$ cm day⁻¹). The simulation is run for 1000 days.

To do this simulation accurately, so that the short-term response of the hillslope to a rainfall event would produce timely streamflow, would require many more nodes than we have used here (we used 3921). Fiori and Russo (2007) found that a 3D numerical model that included preferential flow was required to simulate the short-term response. Our example used a low-constant flux at the soil surface instead of daily rainfall. We did not include evaporation or root water uptake driven by daily potential evapotranspiration. However, our simulation will be adequate to show the long-term response to wetting.

The distribution of hillslope pressure heads over time are shown in Figures 5.34 through 5.36. After 100 days, a wetting
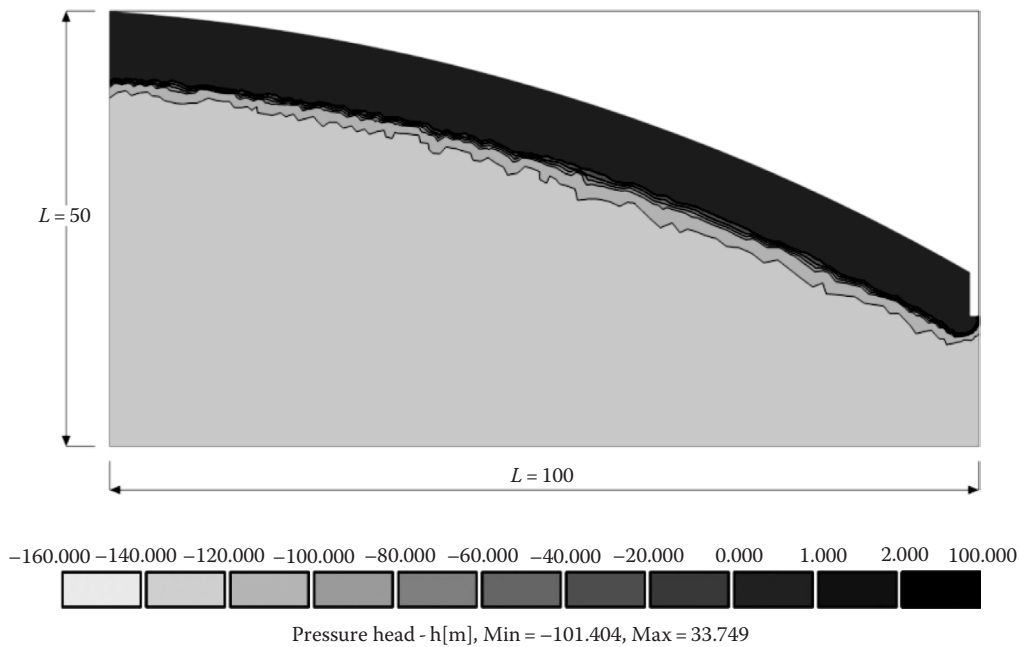
FIGURE 5.34 Pressure heads after 100 days in the hillslope example.
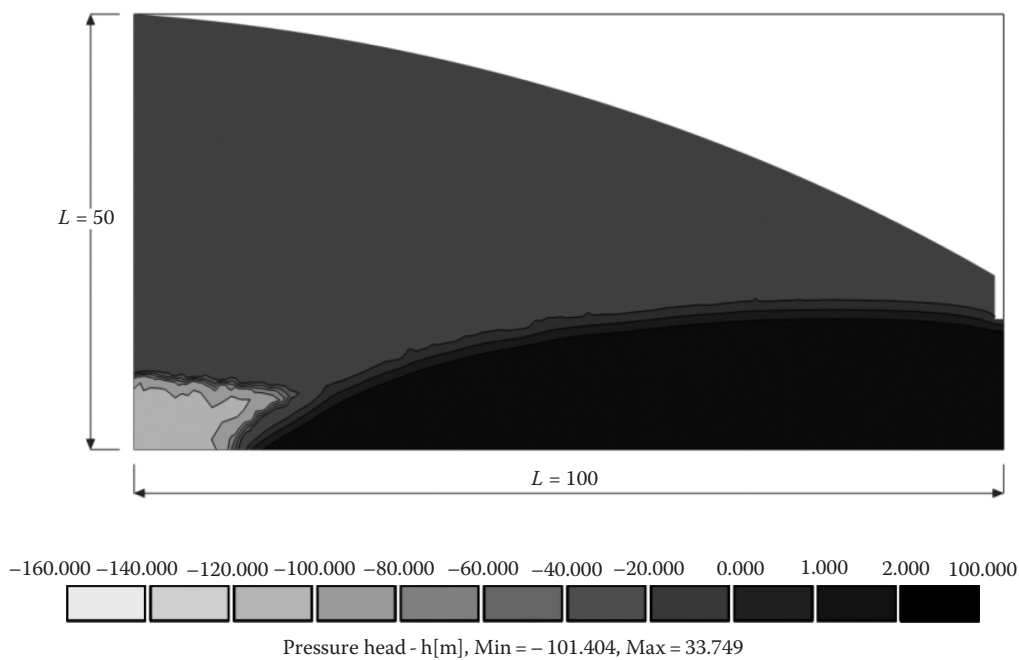
FIGURE 5.35 Pressure heads after 570 days in the hillslope example.

front of unsaturated pressure heads has moved uniformly downward in the hillslope to a depth of about 5 m (Figure 5.34). After 570 days, the wetting front has reached the bottom of the model space on the right (where the distance from the soil surface is the shortest) and a zone of saturated soil has expanded from the right corner to cover nearly the entire lower region of the hillslope (Figure 5.35). The top of the 0–1 cm pressure head contour represents the water table. The last portion of the

hillslope to wet up is the lower left corner because this region is farthest from the soil surface and the extending wetting zone coming in from the right. On the right boundary, the water table has reached the bottom of the channel so water will start exiting the hillslope through the seep into the stream channel. Further rise in the water table next to the stream channel is limited due to the loss of water to the stream. After 800 days, the water table beneath the ridgetop has risen to a height of
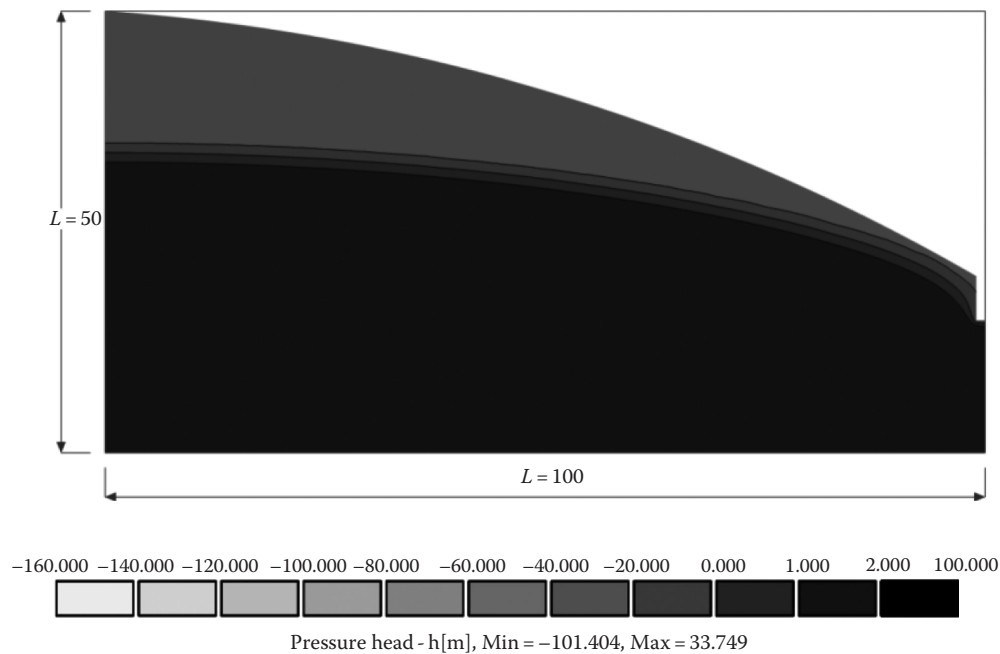
FIGURE 5.36   Pressure heads after 800 days in the hillslope example.

about 35 m above the bottom of the model space (Figure 5.36). On the right, the water table comes very close to the soil surface at about 5 m upslope from the stream channel. The water table at the stream channel has risen to the point that it is about 3 m above the bottom of the channel indicating that the height of water in the stream should be 3 m. There was no further change in the water table from the positions seen for day 800 through the remainder of the simulation (up to 1000 days), indicating that the system has reached a steady state of dynamic equilibrium. The shape of the water table is a subdued replica of the hillslope topography.

In Figure 5.37, the flow through the seepage boundary is shown as a function of time. There is no seepage to the channel

FIGURE 5.37   Seepage flow in the hillslope example.

until about day 500 when the water table intersects the stream channel. The seepage flux increases until about 800 days when it attains a steady value of $2\,m^2\,day^{-1}$, which is the steady infiltration rate through the soil surface (input equals output). At that point, the slope of the water table is such that a sufficient gradient has been generated, given the hydraulic conductivity of the soil that the hillslope can transmit the incoming water at the rate at which it enters the hillslope. Hence, the water table can never be completely flat or there would be no gradient and no flow through the hillslope (unless there is no input). The steepness of the water table will depend on the input and the hydraulic conductivity and length of the hillslope.

The effect of hydraulic conductivity on the slope of the water table can be seen by rerunning the simulation with a saturated hydraulic conductivity that is 10 times the value for the loam in the example so far (249.6 instead of $24.96\,cm\,day^{-1}$). The distribution of pressure heads after 1000 days is shown in Figure 5.38. The water table is nearly flat because only a very small gradient is required to produce the equilibrium flow through the hillslope, given the high K_s in this example. In this case, the shape of the water table is not as reflective of the hillslope topography.

Overall, these results are similar to what was shown earlier in Figure 5.4 in regard to flow to drains. In that simplified analysis of steady flow, it was found that the water table had an elliptical shape and the maximum height of the water table was a function of the infiltration rate at the soil surface, the saturated hydraulic conductivity of the soil, and the distance between the drains (which would correspond to the hillslope length in these examples).

These simulations imply that during wet periods (such as winter months in the eastern United States), the water table will rise to an average height characteristic of the climate and
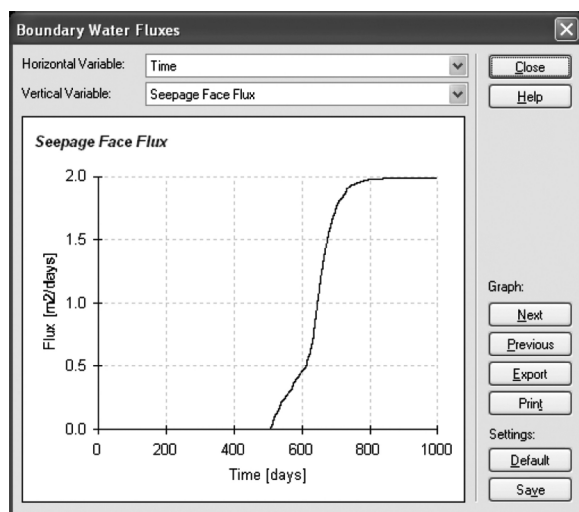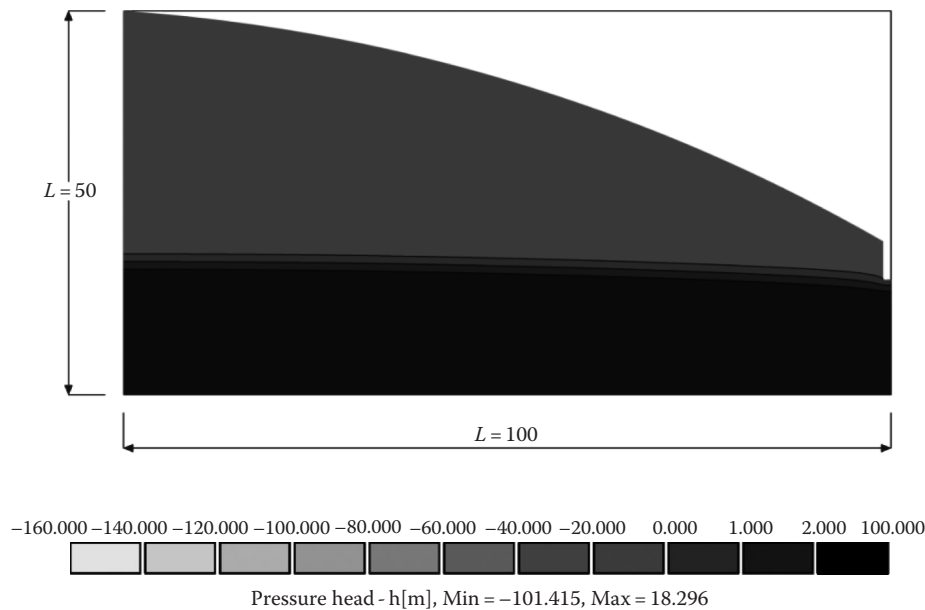
FIGURE 5.38 Pressure heads after 1000 days in the hillslope example using a saturated hydraulic conductivity 10 times the value for a loam soil.

hillslope hydraulic properties. After a long drought, it may take several winter seasons to reestablish the water table at its equilibrium height, even though water contents at the soil surface are near normal. Stream base flow would not return to normal levels until the equilibrium winter water table was fully established.

The shape of the hillslope can also have important effects. If the surface is changed from a convex to a concave hillslope surface, the water table may rise to the surface at the bottom of the footslope in the concave area. This area can be expected to produce saturation-excess runoff as described earlier in Section 5.7.3. In contrast to the previous examples, the water table will be deeper higher up the hillslope at the *shoulder* position. Daniels and Gamble (1967) noted that the shoulder position in hillslopes with broad ridges and narrow valleys in the coastal plain region of the southeastern United States were drier than other areas in the landscape and produced red soils indicative of oxidized iron (aerobic conditions). Gray soils indicative of reduced iron formed at the base of the hillslope and near the center of the broad ridge. They attributed these soil formation processes, not to differences in parent material but rather to the position of the water table.

5.13 Summary

Soil water movement is a key process that affects both water quantity and quality in our environment. Water movement in soils occurs primarily under unsaturated, transient conditions, but saturated flow and steady-state conditions can occur. Although true steady flow may occur only rarely, transient flow usually approaches steady flow at long times. The governing equations for steady and transient flow are the Darcy and

Richards equations, respectively. The most important hydraulic parameters are the soil saturated hydraulic conductivity and unsaturated hydraulic conductivity function. Laboratory and field methods have been developed to measure these parameters and most of these methods require steady flow conditions. PTFs can be used to predict soil hydraulic properties from more easily measurable surrogate properties, such as soil texture.

Analytical solutions of the Richards equation for soil water flow under unsaturated, transient conditions are very few. Numerical methods such as those implemented in the HYDRUS models are commonly used to solve the equation. Numerical methods can also be used to determine soil hydraulic properties under transient flow conditions in an inverse mode. Simplifications of the unsaturated flow equations have been developed to describe infiltration because of the importance of this transient flow process. However, most of the infiltration equations assume 1D vertical flow and do not account for saturation-excess runoff. Numerical solutions can be used to understand complex soil water processes such as infiltration, evaporation, drainage, transpiration, groundwater recharge, and groundwater discharge. Preferential flow is one of the most challenging aspects of predicting soil water movement but various methods have been developed to account for the effects of macropores, fingering, funnel flow, and spatial variability.

References

Abit, S.M., A. Amoozegar, M.J. Vepraskas, and C.P. Niewoehner. 2008. Solute transport and the capillary fringe and shallow groundwater: Field evaluation. Vadose Zone J. 7:890–898.

Ahuja, L.R., and C. Hebson. 1992. Root zone water quality model. GPSR Tech. Rep. No. 2. USDA-ARS, Fort Collins, CO.

Ahuja, L.R., D.L. Nofziger, D. Swartzendruber, and J.D. Ross. 1989. Relationship between Green and Ampt parameters based on scaling concepts and field-measured hydraulic data. Water Resour. Res. 25:1766–1770.

Arya, L.M. 2002. Wind and hot-air methods, p. 916–926. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Ashcroft, G., D.D. Marsh, D.D. Evans, and L. Boersma. 1962. Numerical methods for solving the diffusion equation: I. Horizontal flow in semi-infinite media. Soil Sci. Soc. Proc. 26:522–525.

Baker, R.S., and D. Hillel. 1990. Laboratory tests of a theory of fingering during infiltration into layered soils. Soil Sci. Soc. Am. J. 54:20–30.

Baver, L.D., W.H. Gardner, and W.R. Gardner. 1972. Soil physics. John Wiley & Sons, New York.

Beven, K.J. 2001. Rainfall-runoff modeling: The primer. John Wiley & Sons, Chichester, U.K.

Booltink, H.W.G., and J. Bouma. 2002a. Bypass flow, p. 930–933. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Booltink, H.W.G., and J. Bouma. 2002b. Suction crust infiltrometer, p. 926–930. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Bosch, D.D., and L.T. West. 1998. Hydraulic conductivity variability for two sandy soils. Soil Sci. Soc. Am. J. 62:90–98.

Bouma, J., and J.H. Denning. 1972. Field measurements of unsaturated hydraulic conductivity by infiltration through gypsum crusts. Soil Sci. Soc. Am. Proc. 36:846–847.

Bouwer, H. 1969. Infiltration of water into nonuniform soil. J. Irrigat. Drain. Div. Proc. ASCE IR 4:451–462.

Bouwer, H. 1986. Intake rate: Cylinder infiltrometer, p. 825–844. *In* A. Klute (ed.) Methods of soil analysis. Part 1. Physical and mineralogical methods. 2nd edn. ASA, Madison, WI.

Brooks, R.H., and A.T. Corey. 1964. Hydraulic properties of porous media. Colorado State University, Fort Collins, CO.

Bruce, R.R., J.H. Dane, V.L. Quisenberry, N.L. Powell, and A.W. Thomas. 1983. Physical characteristics of soils in the southern region: Cecil. Southern Cooperative Series. Bulletin 267. Georgia Agricultural Experiment Stations, Athens, GA.

Buckingham, E. 1907. Studies on the movement of soil moisture. Bulletin 38. Bureau of Soils, USDA, Washington, DC.

Burdine, N.T. 1953. Relative permeability calculations from pore-size distribution data. Trans. AIME 198:71–77.

Campbell, G.S. 1974. A simple method for determining unsaturated conductivity from moisture retention data. Soil Sci. 117:311–314.

Carsel, R.F., and R.S. Parrish. 1988. Developing joint probability distributions of soil water retention characteristics. Water Resour. Res. 24:755–769.

Chen, C., and R.J. Wagenet. 1992. Simulation of water and chemicals in macropore soils. Part 1. Representation of the equivalent macropore influence and its effect on soil water flow. J. Hydrol. 130:105–126.

Chiang, S.C., D.E. Radcliffe, W.P. Miller, and K.D. Newman. 1987. Hydraulic conductivities of three southeastern soils as affected by sodium, electrolyte concentration, and pH. Soil Sci. Soc. Am. J. 51:1293–1299.

Clothier, B., and D. Scotter. 2002. Unsaturated water transmission parameters obtained through infiltration, p. 878–898. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Dane, J.H., and G.C. Topp. 2002. Methods of soil analysis. 3rd edn. SSSA, Madison, WI.

Daniels, R.B., and E.E. Gamble. 1967. The edge effect in some ultisols in the North Carolina coastal plain. Geoderma 1:117–124.

Darcy, H. 1856. Les fontaines publiques de la ville de Dijon. Dalmont, Paris, France.

Donatelli, M., J.H.M. Wösten, and G. Belocchi. 2004. Methods to evaluate pedotransfer functions, p. 357–411. *In* Y. Pachepsky and W.J. Rawls (eds.) Development of pedotransfer functions in soil hydrology. Elsevier, Amsterdam, the Netherlands.

Dupuit, J.A. 1863. Etudes theoriques et pratiques sur le movement des eaux dans les canaux decouverts e a travers les terrains permeables. 2nd edn. Dunod, Paris, France.

Durner, W. 1994. Hydraulic conductivity estimation for soils with heterogeneous pore structure. Water Resour. Res. 32:211–223.

Elrick, D.E., and W.D. Reynolds. 1992. Infiltration from constant-head well permeameters and infiltrometers, p. 1–24. *In* G.C. Topp, W.D. Reynolds, and R.E. Green (eds.) Advances in measurement of soil physical properties: Bringing theory into practice. SSSA, Madison, WI.

Ernst, L.F. 1962. Groundwaterstromingen in de verzadigde zone en hun berekening bij aanwezigheid van horizontale evenwijdige open leidengen (Groundwater flow in the saturated zone and its calculation when horizontal parallel open conduits are present). Versl. Landbouwk. Onderz. 67.1. Pudoc, Wageningen, the Netherlands.

Fayer, M.J. 2000. UNSAT-H Version 3.0: Unsaturated soil water and heat flow model. Theory, user manual, and examples. Pacific Northwest National Laboratory-13249, Richland, WA.

Feddes, R.A., E. Bresler, and S.P. Neuman. 1974. Field test of a modified numerical model for water uptake by root systems. Water Resour. Res. 10:1199–1206.

Feddes, R.A., P.J. Kowalik, and H. Zaradny. 1978. Simulation of field water use and crop yield. John Wiley & Sons, New York.

Fetter, C.W. 1988. Applied hydrogeology. 2nd edn. Merrill Publishing Co., New York.

Fiori, A., and D. Russo. 2007. Numerical analysis of subsurface flow and a steep hill slope under rainfall: The role of the spatial heterogeneity of the formation and hydraulic properties. Water Resour. Res. 43:W07445.

Flerchinger, G.N., C.L. Hanson, and J.R. Wight. 1996. Modeling evapotranspiration and surface energy budgets across a watershed. Water Resour. Res. 32:2539–2548.

Flury, M., H. Fluhler, W. Jury, and J. Leuenberger. 1994. Susceptibility of soils to preferential flow of water: A field study. Water Resour. Res. 30:1945–1954.

Forchheimer, P. 1930. Hydraulik. 3rd edn. B.G. Teubner, Ed., Leipzig und Berlin, Germany, p. 596.

Gardner, W.R. 1958. Some steady state solutions of the unsaturated moisture flow equation with application to evaporation from a water table. Soil Sci. 85:228–232.

Gardner, W.R., and M. Fireman. 1958. Laboratory studies of evaporation from soil columns in the presence of a water table. Soil Sci. 85:244–249.

Gerke, H.H., and M.Th. van Genuchten. 1993. A dual-porosity model for simulating the preferential movement of water and solutes in structured porous media. Water Resour. Res. 29:305–319.

Germann, P.F. 1985. Kinematic wave approach to infiltration and drainage into and from soil macropores. Trans. ASAE 28:745–749.

Green, W.H., and G.A. Ampt. 1911. Studies in soil physics: I. The flow of air and water through soils. J. Agric. Sci. 4:1–24.

Grossman, R.B., and T.G. Reinsch. 2002. Bulk density and linear extensibility, p. 201–228. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Gupta, H.V., S. Sorooshian, T.S. Hogue, and D.P. Boyle. 2003. Advances in automatic calibration of watershed models, p. 113–124. *In* Q. Duan (ed.) Calibration of watershed models. American Geophysical Union, Washington, DC.

Hanks, R.J., and S.A. Bowers. 1962. Numerical solution of the moisture flow equation for infiltration into layered soils. Soil Sci. Soc. Am. Proc. 26:530–534.

Hansen, S., H.E. Jensen, N.E. Nielsen, and H. Svendsen. 1990. DAISY: Soil plant atmosphere system model, NPO, Report No. A 10. The National Agency for Environmental Protection, Copenhagen, Denmark.

Haverkamp, R., M. Kutilek, J.-Y. Parlange, L. Rendon, and M. Krejca. 1985. Infiltration under ponded conditions: 2. Infiltration equations tested for parameter time-dependence and predictive use. Soil Sci. 145:305–311.

Haverkamp, R., M. Vauclin, J. Tovina, P.J. Wierenga, and G. Vachaud. 1977. A comparison of numerical simulation models for one-dimensional infiltration. Soil Sci. Soc. Am. Proc. 41:285–294.

Healy, R.W. 1990. Simulation of solute transport in variably saturated porous media with supplemental information on modifications to the U.S. Geological Survey's computer program VS2DI. Water-Resources Investigation Report 90-4025, U.S. Geological Survey.

Hendrickx, J.M.H., L.W. Dekker, and O.H. Boersma. 1993. Unstable wetting fronts in water repellent field soils. J. Environ. Qual. 22:109–118.

Hendrickx, J.M.H., and M. Flury. 2001. Uniform and preferential flow, mechanisms in the vadose zone, p. 149–187. *In* National Research Council (ed.) Conceptual models of flow and transport in the fractured vadose zone. National Academy Press, Washington, DC.

Hewlett, J.D., and C.A. Troendle. 1975. Non-point and diffused water sources: A variable source area problem. Utah State University Symposium. American Society for Civil Engineering, New York.

Hillel, D. 2004. Introduction to environmental soil physics. Elsevier, Amsterdam, the Netherlands.

Hooghoudt, S.B. 1940. Bijdrage tot de kennis van enige natuutkundige grootheden van de grond (Contribution to the knowledge of several physical soil parameters). Versl. Landbouwk. Onderz. 46 B:515–707.

Hopmans, J.W., J. Šimůnek, N. Romano, and W. Durner. 2002. Inverse modeling of transient water flow, p. 936–1008. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Horton, R.E. 1933. The role of infiltration in the hydrologic cycle. Trans. Am. Geophys. Union 14th Ann. Mtg. 446–460.

Horton, R.E. 1940. An approach towards a physical interpretation of infiltration capacity. Soil Sci. Soc. Am. Proc. 5:399–417.

Hubbert, M.K. 1956. Darcy's law and the field equations of the flow of underground fluids. Trans. AIME 207:222–239.

Huyakorn, P.S., and G.F. Pinder. 1983. Computational methods in subsurface flow. Academic Press, New York.

Jansson P.-E., and L. Karlberg. 2001. Coupled heat and mass transfer model for soil-plant-atmosphere systems. Royal Institute of Technology, Department of Civil and Environmental Engineering, Stockholm, Sweden.

Jarvis, N.J. 1989. A simple empirical model of root water uptake. J. Hydrol. 107:57–72.

Jarvis, N.J. 1994. The MACRO model (Version 3.1): Technical description and sample simulations, reports and dissertations 19. Department of Soil Science, Swedish University of Agricultural Sciences, Uppsala, Sweden.

Jarvis, N.J. 2007. Review of non-equilibrium water flow and solute transport in soil macropores: Principles, controlling factors and consequences for water quality. Eur. J. Soil Sci. 58:523–546.

Jaynes, D.B., and E.J. Tyler. 1984. Using soil physical properties to estimate hydraulic conductivity. Soil Sci. 138:298–305.

Jury, W., and R. Horton. 2004. Soil physics. John Wiley & Sons, Hoboken, NJ.

Kostiakov, A.N. 1932. On the dynamics of the coefficient of water percolation in soils and on the necessity of studying it from a dynamic point of view for purposes of amelioration, p. 17–21. Trans. 6th Comm. Int. Soc. Soil Sci. Part A.Moscow, Russia.

Kung, K.-J.S. 1990a. Preferential flow in a sandy vadose zone. 2. Mechanism and implications. Geoderma 46:59–71.

Kung, K.-J.S. 1990b. Preferential flow in a sandy vadose zone: 1. Field observation. Geoderma 46:51–58.

Leij, F.J., M.G. Schaap, and L.M. Arya. 2002. Indirect methods, p. 1009–1045. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Liu, H.H., and J.H. Dane. 1996. Two approaches to modeling unstable flow and mixing of variable density fluids in porous media. Transport Porous Media 23:219–236.

Luxmoore, R.J. 1981. Micro-, meso-, and macroporosity of soil. Soil Sci. Soc. Am. J. 45:671.

Marquardt, D.W. 1963. An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math. 11:431–441.

McCarthy, E.L. 1934. Mariotte's bottle. Science 80:100.

McCord, J.T. 1991. Application of second-type boundaries in unsaturated flow modeling. Water Resour. Res. 27:3257–3260.

McCuen, R.H. 1982. A guide to hydrologic analysis using SCS methods. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Miller, E.E., and R.D. Miller. 1956. Physical theory for capillary flow phenomena. J. Appl. Phys. 27:324–332.

Miller, W.P., and D.E. Radcliffe. 1992. Soil crusting in the southeastern U.S., p. 233–266. *In* M.E. Sumner and B.A. Stewart (eds.) Soil crusting: Chemical and physical processes. S. Lewis Publishers, Boca Raton, FL.

Molz, F.J. 1981. Models of water transport in the soil-plant system: A review. Water Resour. Res. 17:1245–1260.

Mualem, Y. 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resour. Res. 12:593–622.

Nemes, A., M.G. Schaap, F.J. Leij, and J.H.M. Wösten. 2001. Description of the unsaturated soil hydraulic database UNSODA version 2.0. J. Hydrol. 251:151–162.

Neuman, S.P., R.A. Feddes, and E. Bresler. 1974. Finite element simulation of flow in saturated–unsaturated soils considering water uptake by plants. 3rd Annual Report, Project No. A10-SWC-77. Hydraulic Engineering Laboratory, Technion, Haifa, Israel.

Nielson, D.R., J.W. Biggar, and K.T. Erh. 1973. Spatial variability of field measured soil-water properties. Hilgardia 42:215–259.

NRCS. 2004. National engineering handbook. Part 630. USDA-NRCS, Washington, DC. Available online at: http://policy.nrcs.usda.gov/viewerFS.aspx?hid=21422 (accessed July 6, 2009).

NRCS. 2009. National soil information service. USDA-NRCS, Washington, DC. Available online at: http://soils.usda.gov/technical/nasis/documents/metadata/nmdoview.html (accessed July 6, 2009).

Pachepsky, Y.A., W.J. Rawls, and H.S. Lin. 2006. Hydropedology and pedotransfer functions. Geoderma 131:308–316.

Parlange, J.-Y., R. Haverkamp, and J. Touma. 1985. Infiltration under ponded conditions: 1. Optimal analytical solution and comparison with experimental observations. Soil Sci. 139:305–311.

Perroux, K.M., and I. White. 1988. Designs for disc permeameters. Soil Sci. Soc. Am. J. 52:1205–1215.

Philip, J.R. 1957. The theory of infiltration. 1. The infiltration equation and its solution. Soil Sci. 83:345–357.

Philip, J.R. 1966. Plan water relations: Some physical aspects. Ann. Rev. Plant Physiol. 17:245–268.

Pruess, K. 1991. Tough2—A general-purpose numerical simulator for multiphase fluid and heat flow. Report LBL-29400. Lawrence Berkeley Laboratory, Berkeley, CA.

Pruess, K., and J.S.Y. Wang. 1987. Numerical modeling of isothermal and non-isothermal flow in unsaturated fractured rock—A review, p. 11–22. *In* D.D. Evans and T.J. Nicholson (eds.) Flow and transport through unsaturated fractured rock. Geophysics monograph 42. American Geophysical Union, Washington, DC.

Quisenberry, V.L., and R.E. Phillips. 1976. Percolation of surface-applied water in the field. Soil Sci. Soc. Am. J. 40:484–489.

Radcliffe, D.E., and J. Šimůnek. 2010. Soil physics with HYDRUS: Modeling and applications. CRC Press, Boca Raton, FL, p. 373.

Rawls, W.J., and D.L. Brakensiek. 1985. Prediction of soil water properties for hydrologic modeling, p. 293–299. *In* E.B. Jones and T.J. Ward (eds.) Watershed management in the eighties. Proc. Irrigat. Drain. Div. April 30–May 1, 1985. Reston ASCE, Denver, CO.

Rawls, W.J., D.L. Brakensiek, and K.E. Saxton. 1982. Estimation of soil water properties. Trans. ASAE 25:1316–1320, 1328.

Rawls, W.J., T.J. Gish, and D.L. Brakensiek. 1991. Estimating soil water retention from soil physical properties and characteristics. Adv. Soil Sci. 9:213–234.

Reynolds, W.D., and D.E. Elrick. 2002a. Constant head soil core (tank) method, p. 804–808. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Reynolds, W.D., and D.E. Elrick. 2002b. Constant head well permeameter (vadose zone), p. 844–858. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Reynolds, W.D., and D.E. Elrick. 2002c. Falling head soil core (tank) method, p. 809–812. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Reynolds, W.D., D.E. Elrick, and E.G. Youngs. 2002. Ring or cylinder infiltrometers (vadose zone), p. 818–843. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Reynolds, W.D., S.R. Vieira, and G.C. Topp. 1992. An assessment of the single-head analysis for the constant head well permeameter. Can. J. Soil Sci. 72:489–501.

Richards, L.A. 1931. Capillary conduction of fluid through porous mediums. Physics 1:318–333.

Romano, N., and A. Santini. 2002. Field, p. 721–738. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Russo, D., and M. Bouton. 1992. Statistical analysis of spatial variability in unsaturated flow parameters. Water Resour. Res. 28:1911–1925.

Russo, D., and E. Bresler. 1980. Field determinations of soil hydraulic properties for statistical analyses. Soil. Sci. Soc. Am. J. 44:697–702.

Schaap, M.G., and F.J. Leij. 1998. Database-related accuracy and uncertainty of pedotransfer functions. Soil Sci. 163:765–779.

Schaap, M.G., and F.J. Leij. 2000. Improved prediction of unsaturated hydraulic conductivity with the Mualem–van Genuchten model. Soil Sci. Soc. Am. J. 64:843–851.

Schaap, M.G., F.J. Leij, and M.Th. van Genuchten. 1998. Neural network analysis for hierarchical prediction of soil water retention and saturated hydraulic conductivity. Soil Sci. Soc. Am. J. 62:847–855.

Schaap, M.G., F.J. Leij, and M.Th. van Genuchten. 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. 251:163–176.

Simmons, C.S., D.R. Nielsen, and J.W. Biggar. 1979. Scaling of field-measured soil water properties. Hilgardia 47:101–177.

Šimůnek, J., N.J. Jarvis, M.Th. van Genuchten, and A. Gärdenäs. 2003. Review and comparison of models for describing non-equilibrium and preferential flow and transport in the vadose zone. J. Hydrol. 272:14–35.

Šimůnek, J., M. Šejna, H. Saito, M. Sakai, and M.Th. van Genuchten. 2008. The HYDRUS-1D software package for simulating the movement of water, heat, and multiple solutes in variably saturated media, Version 4.0, HYDRUS Software Series 3. Department of Environmental Sciences, University of California Riverside, Riverside, CA.

Šimůnek, J., M. Šejna, and M.Th. van Genuchten. 1999. The HYDRUS-2D software package for simulating two-dimensional movement of water, heat, and multiple solutes in variably saturated media. Version 2.0. IGWMC-TPS-53. International Ground Water Modeling Center, Colorado School of Mines, Golden, CO.

Šimůnek, J., and M.Th. van Genuchten. 2006. Contaminant transport in the unsaturated zone: Theory and modeling. Chap. 22 in The Handbook of Groundwater Engineering, Ed. Jacques Delleur, Second Edition, CRC Press, pp. 22-1–22-46.

Šimůnek, J., M.Th. van Genuchten, and M. Šejna. 2006. The HYDRUS software package for simulating two- and three-dimensional movement of water, heat, and multiple solutes in variably-saturated media. Technical Manual, Version 1.0. PC Progress, Prague, Czech Republic.

Sisson, J.B. 1987. Drainage from layered field soils: Fixed gradient models. Water Resour. Res. 23:2071–2075.

Skaggs, T.H., T.J. Trout, J. Šimůnek, and P.J. Shouse. 2004. Comparison of HYDRUS-2D simulations of drip irrigation with experimental observations. J. Irrigat. Drain. Eng. 130:304–310.

Sklash, M.G., M.K. Stewart, and A.J. Pearce. 1986. Storm runoff generation in humid headwater catchments 2. A case study of hillslope and low-order stream response. Water Resour. Res. 22:1273–1282.

Smith, G.D. 1985. Numerical solution of partial differential equations: Finite difference methods. Clarendon Press, Oxford, U.K.

Steenhuis, T.S., M. Winchell, J. Rossing, J.A. Zollweg, and M.F. Walter. 1995. SCS runoff equation revisited for variable-source runoff areas. J. Irrigat. Drain. Eng. 121:234–238.

Tietje, O., and V. Hennings. 1996. Accuracy of the saturated hydraulic conductivity prediction by pedo-transfer functions compared to the variability with FAO textural classes. Geoderma 69:71–84.

Tugel, A.J., J.E. Herrick, J.R. Brown, M.J. Mausbach, W. Puckett, and K. Hipple. 2005. Soil change, soil survey, and natural resources decision making: A blueprint for action. Soil Sci. Soc. Am. J. 69:738–747.

van Dam, J.C., J. Huygen, J.G. Wesseling, R.A. Feddes, P. Kabat, P.E.V. van Valsum, P. Groenendijk, and C.A. van Diepen. 1997. Theory of SWAP, Version 2.0. Simulation of water flow, solute transport and plant growth in the soil-water-atmosphere-plant environment. Technical Document 45. Department of Water Resources, WAU, Report 71. DLO Winand Staring Centre, Wageningen, the Netherlands.

van Genuchten, M.Th. 1980. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44:892–898.

van Genuchten, M.Th. 1987. A numerical model for water and solute movement in and below the rootzone. Research Report No. 121. U.S. Salinity laboratory, USDA-ARS, Riverside, CA.

van Genuchten, M.Th., F.J. Leij, and S.R. Yates. 1991. The RETC code for quantifying the hydraulic functions of unsaturated soils. EPA/600/2-91/065. USEPA, Washington, DC. Available online at: www.ars.usda.gov/Main/docs.htm?docid=15992 (verified March 3, 2010).

van Hoorn, J.W. 1997. Drainage for salinity control. Department of Water Resources, Wageningen Agricultural University, Wageningen, the Netherlands.

Verburg, K., P.J. Ross, and K.L. Bristow. 1996. SWIM v2.1 User Manual, Divisional Report 130. CSIRO, Canberra, Australia.

Vereecken, H., J. Maes, and J. Feyen. 1990. Estimating unsaturated hydraulic conductivity from easily measured soil properties. Soil Sci. 149:1–12.

Vereecken, H., J. Maes, J. Feyen, and P. Darius. 1989. Estimating the soil moisture retention characteristic from texture, bulk density, and carbon content. Soil Sci. 148:389–403.

Vogel, T., M. Císlerová, and J.W. Hopmans. 1991. Porous media with linearly variable hydraulic properties. Water Resour. Res. 27:2735–2741.

Vrugt, J.A., M.T. van Wijk, J.W. Hopmans, and J. Šimůnek. 2001. One-, two-, and three-dimensional root water uptake functions for transient modeling. Water Resour. Res. 37:2457–2470.

Walter, M.T., J.-S. Kim, T.S. Steenhuis, J.-Y. Parlange, A. Heilig, R.D. Braddock, J.S. Selker, and J. Boll. 2000. Funneled flow mechanisms in a sloping layered soil: Laboratory investigations. Water Resour. Res. 36:841–849.

Warrick, A.W. 1992. Models for disc infiltrometers. Water Resour. Res. 28:1319–1327.

Warrick, A.W. 2003. Soil water dynamics. Oxford University Press, Oxford, U.K.

Warrick, A.W., and P. Broadbridge. 1992. Sorptivity and macroscopic capillary length relationships. Water Resour. Res. 28:427–431.

White, I., and M.J. Sully. 1987. Macroscopic and microscopic capillary length and time scales from field infiltration. Water Resour. Res. 23:1514–1522.

Wooding, R.A. 1968. Steady infiltration from a shallow circular pond. Water Resour. Res. 4:1259–1273.

6

Water and Energy Balances in the Soil–Plant–Atmosphere Continuum

Steven R. Evett
United States Department of Agriculture

John H. Prueger
United States Department of Agriculture

Judy A. Tolk
United States Department of Agriculture

6.1 Introduction

Energy fluxes at soil–atmosphere and plant–atmosphere interfaces can be summed to zero when the surfaces, including plants and plant residues, have no or negligible capacity for energy storage. The resulting energy balance equations may be written in terms of physical descriptions of these fluxes and have been the basis for problem casting and solving in diverse fields of environmental and agricultural science such as estimation of evapotranspiration (ET) from vegetated surfaces, estimation of evaporation from bare soil, rate of soil heating in spring (important for timing of seed germination), rate of residue decomposition (dependent on temperature and water content at the soil surface), and many other problems. The water balances at these surfaces are implicit in the energy balance equations. So, the soil water balance equation, though different from the surface energy balances, is linked to them; a fact that has often been ignored in practical problem solving. In this chapter the energy balances will be discussed first, followed by the water balance in Section 6.3.

Computer simulation has become an important tool for theoretical investigation of energy and water balances at the earth's surface, and for prediction of important results of the mechanisms involved. This chapter will focus more on the underlying principles of energy and water balance processes, and will mention computer models only briefly. More information on computer models that include surface energy and water

balance components can be found in ASAE (1988), Campbell (1985), Richter (1987), Anlauf et al. (1990), Hanks and Ritchie (1991), Pereira et al. (1995), and Peart and Curry (1998) to mention only a few.

6.2 Energy Balance Equation

A surface energy balance is

$$0 = R_n + G + LE + H, \qquad (6.1)$$

where
    R_n is net radiation
    G is soil heat flux
    LE is the latent heat flux (evaporation to the atmosphere) and is the product of the evaporative flux, E, and the latent heat of vaporization, λ
    H is sensible heat flux (all terms taken as positive when flux is toward the surface and in W m^{-2})

Each term may be expressed more completely as the sum of sub-terms that describe specific physical processes, some of which are shown in Figure 6.1. Note that Equation 6.1 may be written individually for soil, plant, and plant residue surfaces. Net radiation includes the absorption and reflection of shortwave radiation
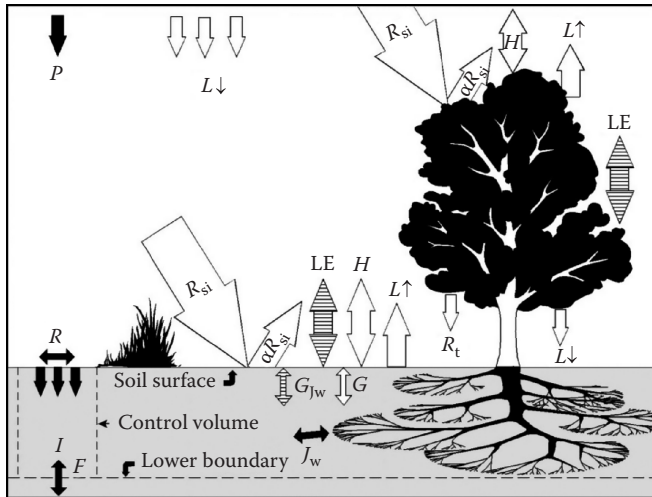
FIGURE 6.1 Water and energy balance components. Water balance components are in black, energy balance components in white. The shared term LE is shaded. Water balance is discussed in Section 6.3.

(sunlight, R_{si}, and the reflected portion, αR_{si}), as well as the emission and reception of longwave radiation ($L\uparrow$ and $L\downarrow$, respectively, Figure 6.1). Soil heat flux involves not only diffusion of heat, G, as expressed by Fourier's law (see Chapter 9) but also convective heat flux, G_{Jw}, as water at temperature T flows at rate J_w into soil at another temperature T'. Evaporation from both soil and plants are examples of latent heat flux; but so also is dew formation, whether it wets the soil surface or plant canopy. Finally, sensible heat flux may occur between soil and atmosphere or between plant and atmosphere, and may be short-circuited between soil and plant, for example, when sensible heat flux from the soil warms the plant. In the next few paragraphs, these fluxes and values they may assume will be illustrated with examples from some contrasting surfaces under variable weather conditions.

Not shown in Figure 6.1 are energy storage due to photosynthesis and that related to temperature and heat capacity of plants. Heat storage in field crops is often ignored as a minor part of the energy balance, but it may not be trivial for forests where heat flux into and out of vegetation can be important, for example, >35 W m^{-2} with heat storage up to 0.3 MJ m^{-2} (Meesters and Vugts, 1996) and >60 W m^{-2} (Haverd et al., 2007). The latter authors included calculation of vegetative heat fluxes and storage in a soil–vegetation–atmosphere transfer model. Photosynthesis is often ignored as a negligible part of the overall energy balance. Further discussion of heat storage and fluxes in vegetation is beyond the scope of this chapter.

Values of the energy fluxes change diurnally (Figures 6.2 through 6.4) and seasonally (Figures 6.5 and 6.6). Regional advection is the large-scale transport of energy in the atmosphere from place to place on the earth's surface. Regional advection events can change the energy balance greatly as illustrated with measurements taken over irrigated wheat at Bushland, Texas (35°11′N Lat; 102°06′W Long) for the 48 h period beginning on day 119, 1992 (Figure 6.2). Total R_{si} was 26.1 and 26.7 MJ m^{-2} on days 119 and 120, respectively; close to the expected maximum
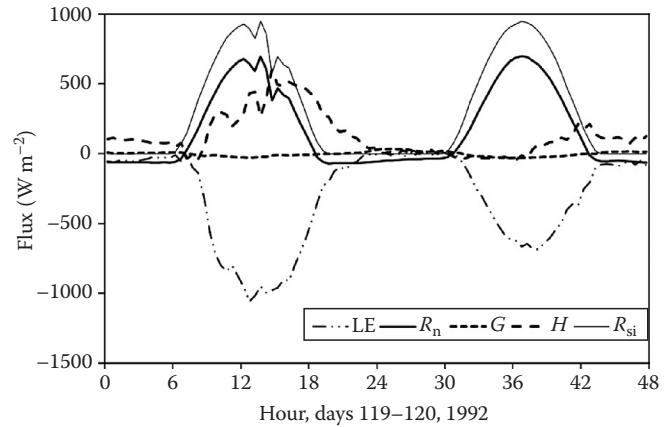
FIGURE 6.2 Energy balance components over irrigated winter wheat at Bushland, Texas.
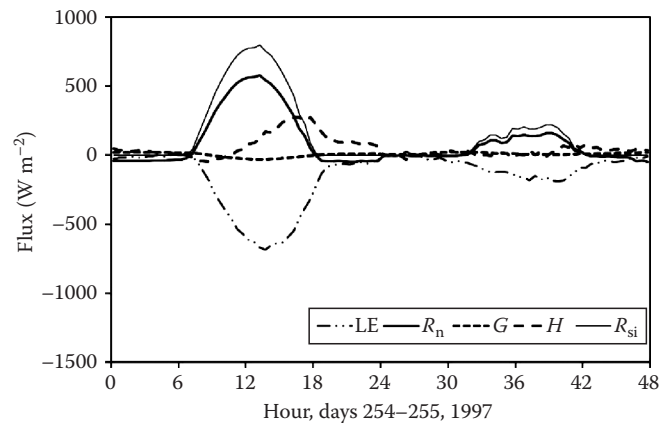
FIGURE 6.3 Energy balance components over irrigated alfalfa at Bushland, Texas.
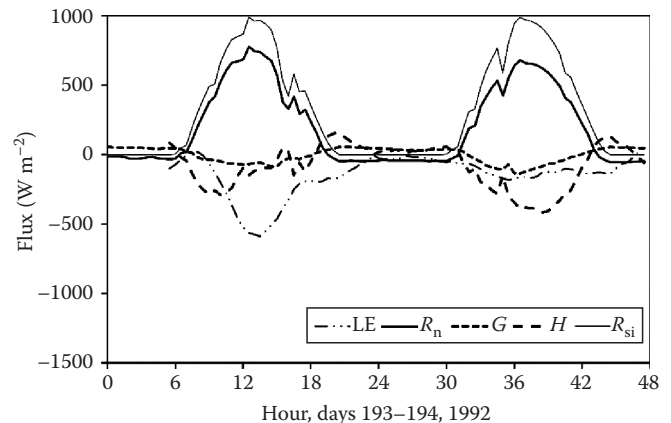
FIGURE 6.4 Energy balance components for bare Pullman clay loam soil after 35 mm of rain at Bushland, Texas.

clear sky value of 28.6 MJ m^{-2} day^{-1} for this latitude and time of year. However, on day 119, strong, dry, adiabatic southwesterly winds (mean 5 m s^{-1}, mean dew point 4.1°C, mean 2 m air temperature 20.1°C) caused H to be strongly positive, providing the extra energy needed to drive total LE to −32.8 MJ m^{-2} day^{-1}, even
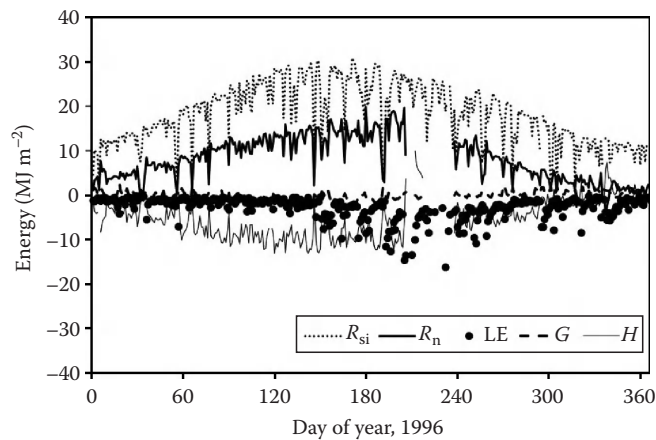
FIGURE 6.5 Daily totals of energy balance terms for a fallow field (mostly bare Pullman clay loam) at Bushland, Texas. Note some missing R_n and H data in late summer.
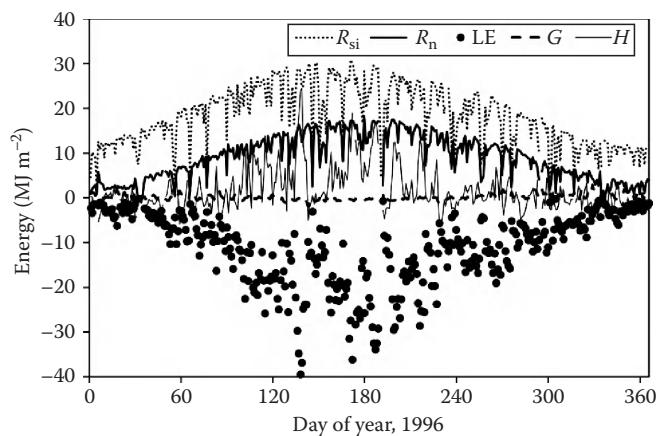
FIGURE 6.6 Daily totals of energy balance terms for irrigated alfalfa at Bushland, Texas.

though both R_{si} and R_n levels were reduced in the afternoon due to cloudiness. Total LE was much larger in absolute magnitude than R_{si} and R_n totals. The next day, the total LE was 39% smaller due to the absence of regional advection, even though total R_{si} and R_n values were slightly larger. Values of G were near zero during this period of full canopy cover when leaf area index (LAI) was 7 (LAI is defined as the single-sided surface area of leaves per unit land area). Note that net radiation was negative at night. This is indicative of strong radiational cooling of the surface, which radiates heat into the clear, low-humidity nighttime skies common to this semiarid location at 1170 m above mean sea level.

Over alfalfa in late summer, R_{si} totals were lower (20.1 and 5.4 MJ m^{-2} day^{-1}, respectively, for days 254 and 255, 1997, Figure 6.3). On the very clear day 254, peak R_{si} was 798 W m^{-2}; and with regional advection occurring, LE flux was large. The 3 h period of negative H just after sunrise was due to the sun-warmed crop canopy being at greater temperature than the air. The arrival of a cool front bringing cloudy skies near midnight caused all fluxes to be much less on day 255, with R_{si} reaching only 220 W m^{-2}, and H near zero for much of the day. The arrival of the cloud

cover and moist air was signaled near midnight by the abrupt change from negative values of R_n and LE to near zero values. In the case of net radiation, this was due to the increased longwave radiation from the clouds, which were warmer and had larger emissivity than the clear sky that preceded them. Also near midnight, latent heat flux neared zero because the strong vapor pressure gradient from moist crop and soil to dry air was reduced by the arrival of moist air. Note that after sunset, but before midnight, latent heat flux was strong, due to continuing strong sensible heat flux, even though net radiation was negative. Again, due to full crop cover (LAI = 3), G values were small, indicating that very little energy was penetrating the soil surface.

For bare soil, G is often larger, becoming an important part of the energy balance (Figure 6.4). After rain and irrigation totaling 35 mm over the previous 2 days, the bare soil was wet on day 193, 1992, at Bushland, Texas. Latent heat flux totaled −14.4 MJ m^{-2} or 6 mm of evaporation; 77% of R_n. Sensible heat flux was negative for the first few hours after sunrise because the soil was warmer than the air, which had been cooled by a nighttime thunderstorm. Later in the day, H and G both approached zero, and near sunset, they became positive, supplying the energy consumed in evaporation that continued well into the night hours. Strong radiational cooling occurred on the nights of days 193 and 194 as indicated by negative values of R_n. Evaporation was probably energy limited on day 193, becoming soil limited on day 194. Latent heat flux on the second day was reduced to −7.4 MJ m^{-2}, and peak daytime values were not much larger in magnitude than those for G. The drying soil became warmer and contributed sensible heat to the atmosphere during almost all daylight hours.

Seasonal variations in daily total energy flux values occur due to changes of sun angle and day length, of distance from the earth to sun (about 3% yearly variation), of seasonal weather, and of surface albedo as plant and residue cover changes (Figures 6.5 and 6.6). A curve describing clear sky solar radiation at Bushland, Texas, would match the high points of R_{si} in Figures 6.5 and 6.6 well. Net radiation was similar for alfalfa and bare soil except for a rainy period beginning about day 190 when the soil was wet and dark and R_n for the fallow field was markedly larger. The big differences were in LE and H. Latent heat flux from the alfalfa was large, reaching nearly −40 MJ m^{-2} (16 mm) on day 136 during a regional advection event that allowed LE to be larger than R_{si}. Sensible heat flux was positive during much of the year. Soil heat flux was small during the growing season, becoming larger as the soil cooled during the fall and winter. For the bare soil, LE values were small during the first 150 days, the latter end of a drought (Figure 6.5). Sensible heat flux was negative during this period and remained negative after rains began until day 203. Evaporative fluxes were fairly small, rarely reaching 6 mm day^{-1} even after rains began. In contrast to alfalfa, soil heat flux for bare soil was larger and more variable throughout the year.

Methods of measurement and estimation of the energy fluxes are needed to characterize the energy balance. Examples of the instrumentation needed to measure components and subcomponents of the energy balance are given in Table 6.1. These will be discussed in the following sections.

TABLE 6.1   Example of Instruments and Deployment Information for Bare Soil Radiation
and Energy Balance Experiments at Bushland, Texas, 1992

Parameter	Instrument	Manufacturer[a] (Model)	Elevation	Description
R_{si}	Pyranometer	Eppley (PSP)[b]	1 m	Solar irradiance
αR_{si}	Pyranometer	Eppley (8–48)	1 m (I[c])	Reflected solar irradiance
$L\downarrow$	Pyrgeometer	Eppley (PIR)	1 m	Incoming longwave radiation
$L\uparrow$	Pyrgeometer	Eppley (PIR)	1 m (I)	Outgoing longwave radiation
R_n	Net radiometer	REBS (Q*6)	1 m	Net radiation
T_s	IR thermometer	Everest (4000; 60° fov)	1 m nadir view angle	Soil surface temperature
T_a	Thermistor	Rotronics		Air temperature and relative humidity
RH	Foil capacitor	(HT225R)	2 m	
U_2	DC generator cups	R.M. Young (12102)	2 m	Wind speed
U_d	Potentiometer vane	R.M. Young (12302)	2 m	Wind direction
T_t	Cu–Co thermocouple	Omega (304SS)	−10 mm and −40 mm	Soil temperature (4)[d]
G_{50}	Plates thermopile	REBS (TH-1)	−50 mm	Soil heat flux (4)
$\theta_{v\text{-}20}$	3-wire	Dynamax	−20 and −40 mm	Soil water content (2)
$\theta_{v\text{-}40}$	TDR probe	TR-100/20 cm	horizontal	
E_m	Lever-scale load cell	Interface (SM-50)	Below lysimeter box	Lysimeter mass change

Source: Adapted from Howell, T.A., J.L. Steiner, S.R. Evett, A.D. Schneider, K.S. Copeland, D.A. Dusek, and A. Tunick. 1993. Radiation balance and soil water evaporation of bare Pullman clay loam soil, p. 922–929. *In* R.G. Allen and C.M.U. Neale (eds.) Management of irrigation and drainage systems, integrated perspectives. Proc. Nat. Conf. Irrigat. Drain. Eng. Park City, UT, July 21–23, 1993. Am. Soc. Civil Eng., New York.

Parameters not shown in Figure 6.1 will be presented later.

[a] Manufacturers and locations are The Eppley Laboratory, Inc., Newport, RI; Radiation and Energy Balance Systems (REBS), Seattle, WA; Everest Interscience, Inc., Fullerton, CA; Rotronic Instrument Corp., Huntington, NY; R.M. Young Co., Traverse City, RI; Omega Engineering, Inc., Stamford, CT; Dynamax, Inc., Houston, TX; Interface, Inc., Scottsdale, AZ.

[b] The mention of trade names of commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

[c] "I" designates instruments that were inverted and facing the ground.

[d] Numbers in parentheses indicate numbers of replicate sensors.

6.2.1 Net Radiation

Net radiation is the sum of incoming and outgoing radiation

$$R_n = R_{si}(1-\alpha) - L\uparrow + L\downarrow, \tag{6.2}$$

where

R_{si} is solar irradiance at the surface (W m^{-2})
α is the albedo or surface reflectance (0–1)
$L\uparrow$ is the longwave radiance of the earth's surface (W m^{-2})
$L\downarrow$ is longwave irradiance from the sky (W m^{-2})

The sun radiates energy like a black body at about 6000 K, while the earth radiates at about 285 K. The theoretical maximum emission power spectra for these two bodies overlap very little (Figure 6.7), a fact that leads to description of radiation from the earth (including clouds and the atmosphere) as longwave, and radiation from the sun as shortwave. Note that the radiance of the earth is about 4 million times less than that of the sun (Figure 6.7). Net radiation may be measured by a net radiometer (Figure 6.8) or its components may be measured separately using pyranometers to measure incoming and reflected shortwave radiation, and pyrgeometers to measure incoming and outgoing longwave radiation (first four instruments
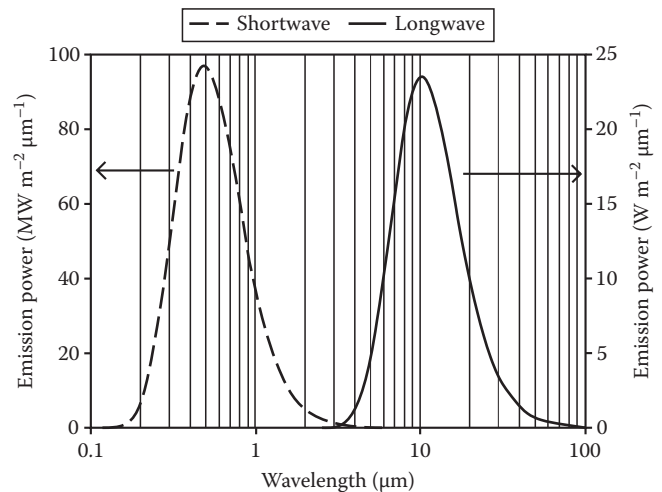
FIGURE 6.7   Emission power spectra for ideal black bodies at 6000 K (left, shortwave range) and 285 K (right, longwave range). Note that the units of the left-hand ordinate are 10^6 larger than those of the right-hand ordinate.

in Table 6.1). Pyranometers and pyrgeometers are thermopile devices that are sensitive equally across the spectrum. While net radiometers are usually much less expensive than four-component suites of pyrgeometers and pyranometers, there have

FIGURE 6.8   REBS Q*7 net radiometer.

FIGURE 6.9   Kipp & Zonen model CM-14 albedometer.

been consistent problems with calibration within a model series and biases between models from different companies leading to inconsistencies of up to 20% (Halldin, S., and A. Lindroth. 1992. Errors in net radiometry: Comparison and evaluation of six radiometer designs. J. Atmos. Ocean. Tech. 9:762–783; Kustas, W.P., J.H. Prueger, L.E. Hipps, J.L. Hatfield, and D. Meek. 1998. Inconsistencies in net radiation estimates from use of several models of instruments in a desert environment. Agr. Forest Meteorol. 90:257–263). Since no international standard exists for longwave radiation measurement and calibration, users are left with the options of cross-calibration of multiple instruments and comparisons with four-component pyrgeometer and pyranometer measurements. Continuing problems with net radiation measurement have made estimation of R_n from R_{si} a common practice that is discussed in Section 6.2.1.6.

6.2.1.1 Outgoing Longwave Radiation

The longwave radiance of the earth's surface, $L\uparrow$, is given by the Stefan–Boltzmann law for radiance from a surface at temperature T (K) and with emissivity ε (0–1).

$$L\uparrow = \varepsilon\sigma T^4, \tag{6.3}$$

where σ is the Stefan–Boltzmann constant (5.67×10^{-8} W m^{-2} K^{-4}). An inverted pyrgeometer (Table 6.1) may be used to measure $L\uparrow$ and, if accompanied by suitable surface temperature measurements, may allow estimation of surface emissivity, ε, by inversion of Equation 6.3. Surface temperature is often measured by suitably placed and shielded thermocouples, or remotely by infrared thermometer (IRT). There are problems with either type of measurement (radiational heating of the thermocouples, and uncertainty of the emissivity needed for accurate IRT measurements).

Values of α and ε for soil and plant surfaces may be estimated from published values relating them to surface properties (see Section 6.2.1.3 and Table 6.4). For soil, the dependence of α on water content is strong, but nearly linear, and amenable to estimation.

6.2.1.2 Solar Irradiance

Solar irradiance, R_{si}, includes both direct beam and diffuse shortwave radiation reaching the earth's surface and is defined

as the radiant energy reaching a horizontal plane at the earth's surface. It may be easily measured by pyranometer with calibration to international standards (Table 6.1) or by solar cells. Silicon photodetector solar radiation sensors such as the LI-COR model LI-200SA are sensitive in only part of the spectrum but are calibrated to give accurate readings in most outdoor light conditions. Silicon sensors are much cheaper than thermopile pyranometers and have found widespread use in field weather stations. Measurement of both incident (R_{si}) and reflected (R_{sr}) shortwave allows estimation of the albedo from

$$R_{si}(1 - \alpha) = R_{si} - R_{sr}. \tag{6.4}$$

This is done using upward and downward facing matched pyranometers (Table 6.1). Specially made albedometers are available for this purpose (e.g., Kipp & Zonen model CM-14; Figure 6.9).

In the absence of measurements, solar irradiance may be estimated from the sun's brightness above the atmosphere R_a (W m^{-2}) and reductions in surface irradiation mediated by latitude, time of day and year, and elevation. Such estimates also provide a useful check on measurements by providing an upper bound to R_{si}, the clear sky solar irradiance, R_{so}. The "solar constant" is the flux density of solar radiation on a plane surface perpendicular to the direction of radiation and outside the earth's atmosphere. It is 1366 W m^{-2} on average, with a within-year variation of about ±3.5%, being largest in January when the sun is closest to the earth and smallest in July (Jones, 1992). Several satellite observation platforms have recorded the value of solar irradiance, R_a, over a >20 year span (Figure 6.10) and clearly show the average solar cycle of 11 years (Fröhlich and Lean, 1998). The data range over about 0.36% of the mean value when corrected for earth–sun distance. Thus, considering the "solar constant" to be 1366 W m^{-2} will introduce a <1% error in calculations when corrected for earth–sun distance.

Irradiance at the earth's surface is somewhat less, due to absorption and scattering in the atmosphere and due to sun angle effects, not often exceeding 1000 W m^{-2}. The further the sun is from the zenith the longer the transmission path through the atmosphere, and the more absorption and scattering occurs. Also, as sun angle above the horizon, β (radians), decreases (it is
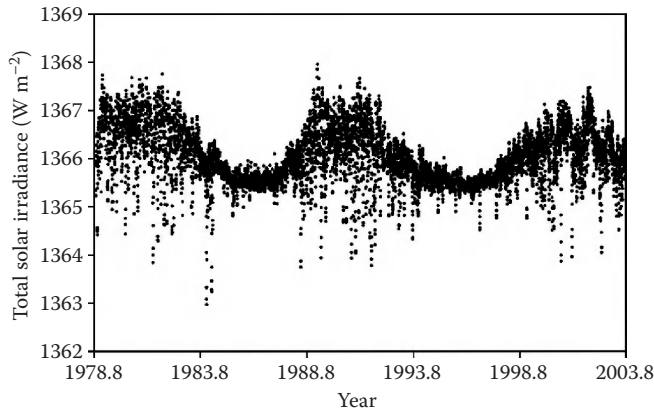
FIGURE 6.10 Composite of observations of total solar irradiance (R_a) outside the earth's atmosphere from six satellites: NIMBUS-7, SMM, ERBS, UARS, SOHO, and ACRIM-Sat (corrected for earth–sun distance). (ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR_IRRADIANCE/composite_d25_07_0310a.dat. Graph constructed using version d25_07_0310a of the dataset from PMOD/WRC, Davos, Switzerland, including unpublished data from the VIRGO experiment on the cooperative ESA/NASA Mission SoHO; Fröhlich, C., and J. Lean. 1998. The sun's total irradiance: Cycles, trends and related climate change uncertainties since 1976. Geophys. Res. Lett. 25:4377–4380.)

greatest at solar noon) the radiation density, I, on a horizontal surface decreases according to Lambert's law

$$I = I_o \sin(\beta), \tag{6.5}$$

where I_o is the flux density on a surface normal to the beam.

Sun angle (β) changes with time of day and year, and can be calculated from

$$\beta = \sin^{-1}[\sin(\delta)\sin(\phi) + \cos(\delta)\cos(\phi)\cos(t_{sa})], \tag{6.6}$$

where

ϕ is latitude
δ is solar declination
t_{sa} is solar time angle (all radians)

Solar time angle is defined as

$$t_{sa} = \frac{(t - t_{SN})2\pi}{24}, \tag{6.7}$$

where

t is time (h)
t_{SN} is the time (h) of solar noon

The time of solar noon varies with time of year and longitude according to (recall that 1° longitude = 4 min):

$$t_{SN} = 12 + \frac{4(\text{longitude} - \text{local meridian})}{60} - t_{EQ}, \tag{6.8}$$

where

t_{EQ} is the "equation of time" value (h)
longitude is in degrees
the local meridian is the longitude (°) for which standard time is calculated for the time zone in question

In the United States, the meridians for Eastern Standard Time (EST), Central Standard Time (CST), Mountain Standard Time (MST), and Pacific Standard Time (PST) are 75°, 90°, 105°, and 120°, respectively. Local or true solar time (t_{LS}, h) for any local standard time (t_{ST}) may be calculated with

$$t_{LS} = t_{ST} - 4\frac{(\text{longitude} - \text{local meridian})}{60} + t_{EQ}. \tag{6.9}$$

The declination (radians) may be calculated from (Rosenberg et al., 1983)

$$\delta = 0.4101\cos\left[\frac{2\pi(J - 172)}{365}\right], \tag{6.10}$$

where J is the day of the year.

List (1971) gave "equation of time" values to the nearest second for the first of each month and every 4 days after that for each month (95 values for the year). The following equation reproduces those values with a maximum error of 6 s and can be used to estimate t_{EQ} (h) for any day of the year.

$$t_{EQ} = b_0 + b_1\sin\left(\frac{J}{P_1}\right) + b_2\cos\left(\frac{J}{P_1}\right) + c_1\sin\left(\frac{J}{P_4}\right) + c_2\cos\left(\frac{J}{P_4}\right)$$
$$+ c_3\sin\left(\frac{2J}{P_4}\right) c_4\cos\left(\frac{2J}{P_4}\right) + c_5\sin\left(\frac{3J}{P_4}\right) + c_6\cos\left(\frac{3J}{P_4}\right)$$
$$+ c_7\sin\left(\frac{4J}{P_4}\right) + c_8\cos\left(\frac{4J}{P_4}\right), \tag{6.11}$$

where the coefficients b_1 and c_1 are given in Table 6.2, and $P_1 = 182.5/(2\pi)$ and $P_4 = 365/(2\pi)$.

Jensen et al. (1990) gave a simpler method for t_{EQ}:

$$t_{EQ} = 0.1645\sin(2b) - 0.1255\cos(b) - 0.025\sin(b), \tag{6.12}$$

where $b = 2\pi(J - 81)/364$. The maximum error compared against List's t_{EQ} values is 88 s.

Disregarding air quality, solar irradiance is affected by latitude, time of year and day, and elevation. Latitude and time affect the sun angle, β, and so affect both the path length of radiation through the atmosphere (and thus absorption and scattering losses) and the flux density at the surface through Equation 6.5. Elevation affects the path length. Methods for calculating extraterrestrial, R_{sa}, and clear sky solar irradiance at the surface, R_{so}, are given by ASCE (2005), Campbell (1977, Chapter 5),

TABLE 6.2 Coefficients for Calculating the "Equation of Time" Value from Equation 6.11

b_0	4.744×10^{-5}	c_2	9.19×10^{-3}	c_6	-1.29×10^{-3}
b_1	-0.157	c_3	-5.78×10^{-4}	c_7	-3.23×10^{-3}
b_2	-0.0508	c_4	3.61×10^{-4}	c_8	-2.1×10^{-3}
c_1	-0.122	c_5	-5.48×10^{-3}		

Jensen et al. (1990, Appendix B), Jones (1992, Appendix 7), and McCullough and Porter (1971). Calculation of R_{sa} depends on latitude and time of day. Once R_{sa} is calculated, R_{so} may be estimated from considerations of absorption and scattering in the atmosphere, which depend mainly on the path length through the atmosphere and its density. Thus latitude, time of day, and elevation are factors in estimating R_{so} from R_{sa}. The value of R_{so} is an important quantity against which to check measured R_{si}; and it can be used in estimates of R_n, either to replace R_{si} in Equation 6.2 or to use location-specific regression relationships of $R_n = f(R_{so})$ (see Jensen et al., 1990, Appendix B). Duffie and Beckman (1991) presented the following method of calculating R_{sa} (MJ m^{-2} h^{-1}) for any period, τ (h),

$$R_{sa} = \left[\frac{24(60)}{2\pi}\right] G_{SC} d_r \{\cos(\phi)\cos(\delta)[\sin(\omega_2) - \sin(\omega_1)]$$
$$+ (\omega_1 - \omega_2)\sin(\phi)\sin(\delta)\}, \qquad (6.13)$$

where

    G_{SC} is the solar constant (0.08202 MJ m^{-2} min^{-1})
    d_r is the relative earth–sun distance
    ω_1 and ω_2 are the solar time angles at the beginning and end of the period, respectively (all angles in radians)

The term $24(60)/(2\pi)$ is the inverse angle of rotation per minute. The relative earth–sun distance, d_r, is given by

$$d_r = 1 + 0.033\cos\left(\frac{2\pi J}{365}\right), \qquad (6.14)$$

where J is the day of year. The solar time angles at the beginning and end of the period in question are given by

$$\omega_1 = \omega - \frac{\pi}{(24/\tau)}, \qquad (6.15)$$

$$\omega_2 = \omega + \frac{\pi}{(24/\tau)}, \qquad (6.16)$$

where

    ω is the solar time angle at the center of the period (radians)
    τ is the length of the period in hours

Calculation of R_{sa} for an entire day requires knowledge of the sunrise and sunset times. The sunset time angle (angle from solar noon to sunset, radians) is given by

$$\omega_s = \cos^{-1}[-\tan(\phi)\tan(\delta)], \qquad (6.17)$$

from which it is clear that day length, t_D (h), is

$$t_D = \frac{24\omega_s}{\pi}. \qquad (6.18)$$

Equation 6.13 can be rewritten for total daily R_{sa} as

$$R_{sa} = \left[\frac{24(60)}{\pi}\right] G_{SC} d_r [\cos(\phi)\cos(\delta)\sin(\omega_s) + \omega_s \sin(\phi)\sin(\delta)]. \qquad (6.19)$$

For example, on day 119 at latitude 35°11′N, longitude 102°6′W, R_{sa} calculated using Equation 6.13 on a half-hourly basis totaled 38.097 MJ m^{-2} compared with 38.100 MJ m^{-2} calculated with Equation 6.19.

Jensen et al. (1990) recommended estimating daily total clear sky solar irradiance as

$$R_{so} = 0.75 R_{sa} \qquad (6.20)$$

Somewhat in agreement with this, Monteith and Unsworth (1990) stated that direct beam radiation rarely exceeded 1030 W m^{-2}, about 75% of the solar constant.

Jones (1992) and Monteith and Unsworth (1990) suggested

$$R_{si} = R_{si\_max} \sin\left(\frac{\pi t}{N}\right), \qquad (6.21)$$

for instantaneous values of R_{si} on clear days, where R_{si_max} is the maximum instantaneous irradiance occurring at solar noon, t is time after sunrise (h), and N is day length (h).

Rather than R_{si_max}, it is more common to know daily total R_{si}. Collares-Pereira and Rabl (1979) gave the ratio of hourly irradiance, $R_{si,h}$ to daily irradiance, $R_{si,d}$ as

$$\frac{R_{si,h}}{R_{si,d}} = \left(\frac{\pi}{24}\right)(a + b\cos\omega)\frac{(\cos\omega - \cos\omega_S)}{(\sin\omega_S - \omega_S \cos\omega_S)}, \qquad (6.22a)$$

where

$$a = 0.409 + 0.5016\sin\left(\omega_S - \frac{\pi}{3}\right) \qquad (6.22b)$$

and

$$b = 0.6609 + 0.4767\sin\left(\omega_S - \frac{\pi}{3}\right) \qquad (6.22c)$$

Equation 6.22 performed well when applied to data from Bushland, Texas (Figure 6.11).

More complex methods of estimating R_{so} account for attenuation of direct beam radiation using Beer's law, coupled with Lambert's law to calculate irradiance on a horizontal surface, plus an accounting of diffuse irradiance (see, e.g., List, 1971; Rosenberg et al., 1983; Jones, 1992). Beer's law describes the intensity I of radiation after passing a distance x through
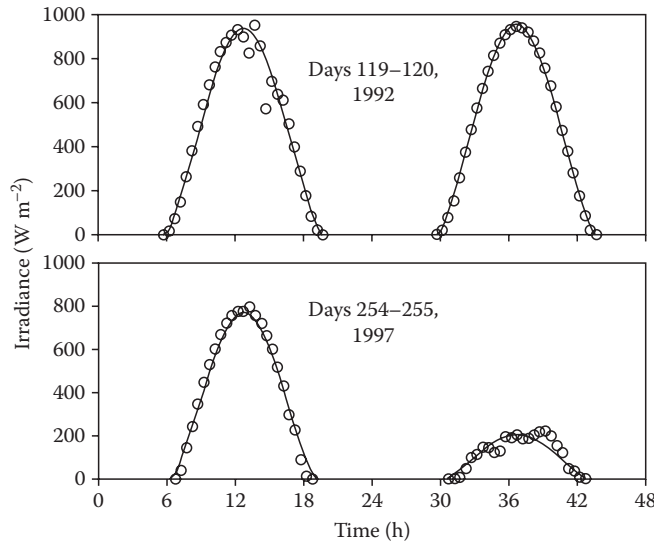
FIGURE 6.11 Solar irradiance measured at Bushland, Texas, in 1992 and 1997 on clear and cloudy days (open circles) and Equation 6.22a half-hourly predictions (lines).

a medium in terms of an extinction coefficient, k, and the initial intensity, I_a, as

$$I = I_a e^{kx}. \tag{6.23}$$

For solar radiation, the distance is expressed in terms of air mass number, m, as (List, 1971)

$$m = \sec\left(\frac{\pi}{2} - \beta\right). \tag{6.24}$$

The air mass is referenced to the length of the path when the sun is directly overhead. For β less than 0.175 radians (10°), the

measured air mass number is less than that given by Equation 6.24 due to refraction and reflection at these low angles. List (1971) gives corrections and notes also that for pressures, p, less than standard sea-level pressure, p_0, that m should be corrected by $m = m(p/p_0)$. Rewriting Equation 6.23, we have

$$I_o = I_a e^{k \sec(\pi/2 - \beta)}, \tag{6.25}$$

where I_o is direct beam radiation at the earth's surface. Monteith and Unsworth (1990) give a range of values of k for England as 0.07 for very clean air and 0.6 for very polluted air.

Assuming that both direct, I_o, and diffuse, I_d, radiation are known, the total irradiance at the surface is

$$R_{si} = I_o(\sin\beta) + I_d. \tag{6.26}$$

Diffuse radiation is difficult to estimate because it is so dependent on cloud cover and aerosol concentration in the air. Yet, summarizing several data sets, Spitters et al. (1986) found that the proportion of R_d to R_{si} is a function of the ratio of R_{si} to R_{sa} (Figure 6.12) described for daily total R_{si} by

$$R_{d,d} = 1, \quad \frac{R_{si,d}}{R_{sa,d}} < 0.07 \tag{6.27a}$$

$$\frac{R_{d,d}}{R_{si,d}} = 1 - 2.3\left(\frac{R_{si,d}}{R_{sa,d}} - 0.07\right)^2, \quad 0.07 \le \frac{R_{si,d}}{R_{sa,d}} < 0.35 \tag{6.27b}$$

$$\frac{R_{d,d}}{R_{si,d}} = 1.33 - 1.46\left(\frac{R_{si,d}}{R_{sa,d}}\right), \quad 0.35 \le \frac{R_{si,d}}{R_{sa,d}} < 0.75 \tag{6.27c}$$

$$\frac{R_{d,d}}{R_{si,d}} = 0.23\left(\frac{R_{si,d}}{R_{sa,d}}\right), \quad 0.75 \le \frac{R_{si,d}}{R_{sa,d}}, \tag{6.27d}$$
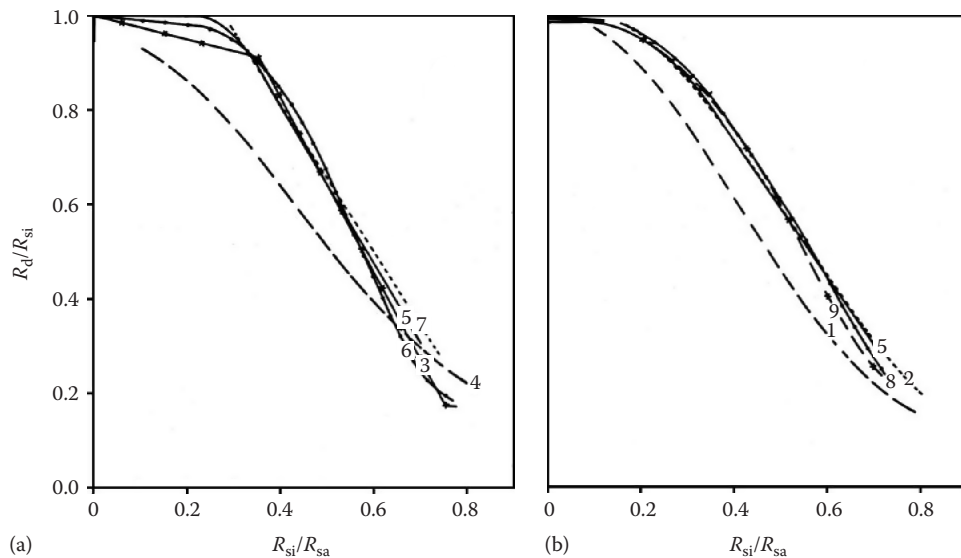
FIGURE 6.12 Daily (a) and hourly (b) relationships between R_d/R_{si} and R_{si}/R_{sa}. (Reprinted from Spitters, C.J.T., H.A.J.M. Toussaint, and J. Goudriaan. 1986. Separating the diffuse and direct component of global radiation and its implications for modeling canopy photosynthesis. Part I. Components of incoming radiation. Agr. Forest Meteorol. 38:217–229.)

and for hourly values by

$$R_{d,h} = 1, \quad \frac{R_{si,h}}{R_{sa,h}} \le 0.22 \qquad (6.28a)$$

$$\frac{R_{d,h}}{R_{si,h}} = 1 - 6.4\left(\frac{R_{si,h}}{R_{sa,h}} - 0.22\right)^2, \quad 0.22 < \frac{R_{si,h}}{R_{sa,h}} \le 0.35 \qquad (6.28b)$$

$$\frac{R_{d,h}}{R_{si,h}} = 1.47 - 1.66\left(\frac{R_{si,h}}{R_{sa,h}}\right), \quad 0.35 < \frac{R_{si,h}}{R_{sa,h}} \le K \qquad (6.28c)$$

$$\frac{R_{d,h}}{R_{si,h}} = R, \quad K < \frac{R_{si,h}}{R_{sa,h}}, \qquad (6.28d)$$

where

$$R = 0.847 - 1.61\sin\beta + 1.04\sin^2\beta \qquad (6.29)$$

and

$$K = \frac{(1.47 - R)}{1.66} \qquad (6.30)$$

6.2.1.3 Surface Albedo and Emissivity

Because R_{si} provides most of the energy that is partitioned at the earth's surface, albedo plays a major role in the energy balance. The mean albedo of the earth is 0.36 ± 0.06 (Weast, 1982). But albedo varies diurnally (Figure 6.13) with larger albedo corresponding to lesser sun angle (see also bare soil data of Monteith and Szeicz, 1961; Idso et al., 1974; Aase and Idso, 1975). Soil and plant

surfaces are often considered optically rough, but in some cases, specular (mirror-like) rather than diffuse reflection may occur. Some plant leaves are shiny and reflect specularly when the angle of incident radiation is small. Wet soil surfaces may also reflect specularly. These mechanisms lead to larger albedo when sun angle is smaller. The albedo of plant stands is also smaller in midday because more sunlight penetrates deeply within the canopy and is trapped by multiple reflections. Wilting and other physiological changes during the day may also contribute to changes in albedo.

Soil albedo decreases as water content increases. Bowers and Hanks (1965) found the relationship to be curvilinear, as did Skidmore et al. (1975). Working with thinner soil layers, Idso and Reginato (1974) found that bare soil albedo changed linearly with water content of the surface 2 mm of soil (smooth clay loam; Figure 6.14). For thicker layers, the relationship was curvilinear. The maximum albedo, 0.3, occurred for air dry soil, but the minimum albedo, 0.14, occurred at about $0.23\,\text{m}^3\,\text{m}^{-3}$ water content, much less than saturation. This represents field capacity (soil water tension of 30 kPa) for this soil; and Idso and Reginato (1974) postulated that the minimum albedo would occur at field capacity for all soils. Kondo et al. (1992) found a similar relationship for a bare loam with a maximum albedo of 0.24 and minimum of 0.13, and with the minimum attained when soil water content reached about $0.22\,\text{m}^3\,\text{m}^{-3}$. Data of Idso et al. (1974, 1975) show that the difference between wet and dry soil albedos was constant despite the diurnal variation of albedo or the day of year. Similarly, Monteith (1961) measured albedo of clay loam to be 0.18 when dry, decreasing to 0.11 when at field capacity water content of $0.35\,\text{m}^3\,\text{m}^{-3}$.

The interaction of sun angle and soil drying causes complex patterns of soil albedo change over time. Figure 6.13 illustrates low daytime wet soil albedos of 0.11 after irrigation and rain on days 191 and 192, 1992, at Bushland, Texas. Rapid soil surface drying on day 193 caused albedo to rise sharply during the day. Additional drying on day 194 completed the change, and
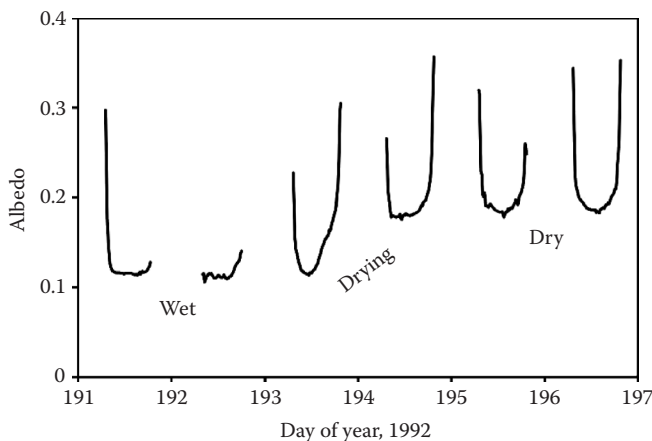
FIGURE 6.13 Albedo plotted as a function of time for smooth, bare Pullman clay loam at Bushland, Texas, in the daytime, when wet, during drying, and dry.
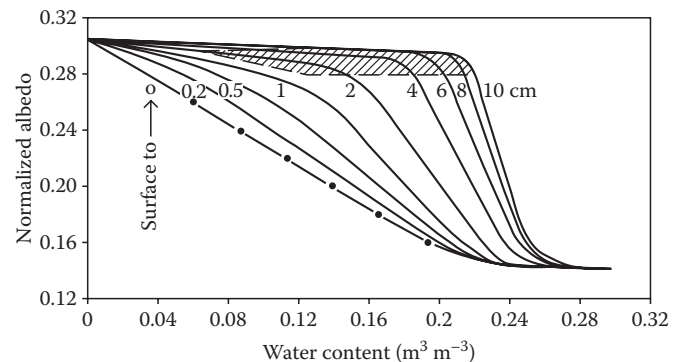
FIGURE 6.14 Albedo, normalized according to sun angle, versus soil water content for different surface layer thicknesses of Avondale clay loam at Phoenix, AZ. Solid circles describe albedo versus water content at the surface. Data for shaded areas are uncertain. (From Idso, S.B., and R.J. Reginato. 1974. Assessing soil-water status via albedo measurement. Hydrol. Water Resour. Arizona SW 4:41–54.)

diurnal albedo changes on days 195 and 196 reflected only sun angle effects, with a minimum albedo of 0.2 for this smooth soil surface. The same surface in a roughened condition earlier in the year never reached midday albedo values larger than 0.13.

Other than water content, major determinants of soil albedo are color, texture, organic matter content, and surface roughness. Dvoracek and Hannabas (1990) presented a model of albedo dependence on sun angle, surface roughness, and color

$$\alpha = p^{(c \sin \beta + 1)}, \tag{6.31}$$

where
   p is a color coefficient
   c is a roughness coefficient
   β is solar angle

They demonstrated good fits with measured data (Table 6.3). Albedo values modeled using p and c values from Table 6.3 for wheat and cotton (day of year 192, latitude 41°N) appear realistic (Figure 6.15). However, the physical meaning of the p and c coefficients is not well understood.

Daily mean albedos may be calculated as the ratio of daily total reflected shortwave energy to daily total R_{si}. Using data from Figures 6.5 and 6.6, daily mean albedos for fallow (soybean residue) and alfalfa differ by about 0.10 when the soil is very dry (Figure 6.16). The gradual decline in fallow albedo in early 1996 may be due to decomposition of the soybean residue.

TABLE 6.3 Color (p) and Roughness (c) Coefficients for Equation 6.31

Surface and Condition	Color Coefficient p	Roughness Coefficient c	Mean r^2
Lakes and ponds, clear water			
Waves, none	0.13	0.29	0.82
Waves, ripples up to 2.5 cm	0.16	0.70	0.74
Waves, larger than 2.5 cm with occasional whitecaps	0.23	1.25	0.83
Waves, frequent whitecaps	0.30	2.00	0.85
Lakes and ponds			
Green water, ripples up to 2.5 cm	0.22	0.70	0.90
Muddy water, no waves	0.19	0.29	0.76
Cotton			
Winds, calm to 4.5 m s⁻¹	0.27	0.27	0.80
Winds, over 4.5 m s⁻¹	0.27	0.43	0.88
Wheat			
Winds, calm to 4.5 m s⁻¹	0.31	0.92	0.85
Winds, over 4.5 m s⁻¹	0.37	1.30	0.85

Source: Dvoracek, M.J., and B. Hannabas. 1990. Prediction of albedo for use in evapotranspiration and irrigation scheduling. *In* Visions of the future, proceedings of the third national irrigation symposium held in conjunction with the 11th annual international irrigation exposition, October 28–November 1, Phoenix, AZ. ASAE, St. Joseph, MI.
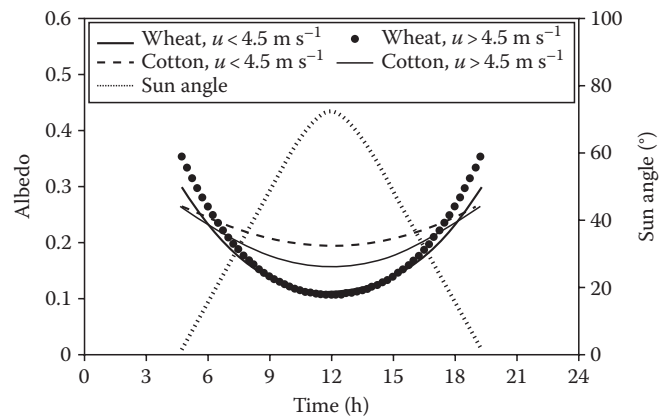
FIGURE 6.15 Albedo for wheat and cotton versus sun angle and wind speed (u) according to Equation 6.31.
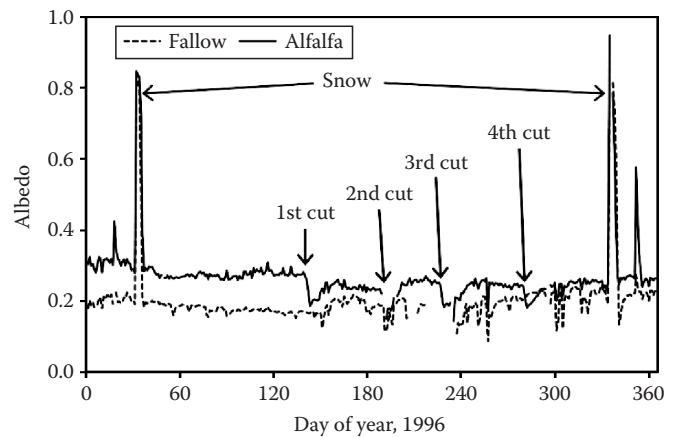
FIGURE 6.16 Daily mean albedos for irrigated alfalfa and fallow after soybean on Pullman clay loam at Bushland, Texas.

Albedo for the alfalfa field declined at each cutting to nearly that of the fallow field, which was initially rougher than the soil under the alfalfa. But, during heavy rains in the latter part of the year, the fallow soil surface was slaked and smoothed and its albedo increased to near that of the alfalfa. Thus, after the fourth cut, the alfalfa field albedo was lower than that of the fallow field for a brief time, probably because the alfalfa was irrigated and the fallow field had dried out again. Peaks of albedo exceeding 0.8 were due to snow early and late in the year. In contrast to soil, albedo of closed canopies (well watered) is relatively constant (Table 6.4).

Albedo values for many plant covers may be found in Gates (1980). For surfaces with plants, the amount of radiation reaching the soil surface, R_t (Figure 6.1), depends on the LAI and the canopy structure. Numerical models have been developed that take into account leaf orientation and distribution in the canopy to calculate absorption of radiation at different levels in the canopy (Goudriaan, 1977; Chen, 1984). Lascano et al. (1987) used Chen's model to calculate polynomials representing the dependence of albedo on LAI, as well as the dependence of the

TABLE 6.4 Some Albedo and Emissivity Values for Various Soil and Plant Surfaces
(Most are Daily Averages)

Surface	Albedo	Emissivity	Source
Soils, dark, wet to light, dry	0.05–0.50	0.90–0.98	Oke (1978)
Dry sandy soil	0.25–0.45		Rosenberg et al. (1983)
Bare dark soil	0.16–0.17		Rosenberg et al. (1983)
Dry clay soil	0.20–0.35		Rosenberg et al. (1983)
Quartz sand	0.35		Van Wijk and Scholte Ubing (1963)
Sand, wet	0.09	0.98	Van Wijk and Scholte Ubing (1963)
Sand, dry	0.18	0.95	Van Wijk and Scholte Ubing (1963)
Dark clay, wet	0.02–0.08	0.97	Van Wijk and Scholte Ubing (1963)
Dark clay, dry	0.16	0.95	Van Wijk and Scholte Ubing (1963)
Fields, bare	0.12–0.25		Van Wijk and Scholte Ubing (1963)
Fields, wet, plowed	0.05–0.14		Van Wijk and Scholte Ubing (1963)
Dry salt cover	0.50		Van Wijk and Scholte Ubing (1963)
Snow, fresh	0.80–0.95		Rosenberg et al. (1983)
Snow, old	0.42–0.70		Rosenberg et al. (1983)
Snow, fresh	0.95	0.99	Oke (1978)
Snow, old	0.40	0.82	Oke (1978)
Snow, fresh	0.80–0.85		Van Wijk and Scholte Ubing (1963)
Snow, compressed	0.70		Van Wijk and Scholte Ubing (1963)
Snow, melting	0.30–0.65		Van Wijk and Scholte Ubing (1963)
Grass, long (1 m)	0.16	0.90	Oke (1978)
Short (0.02 m)	0.26	0.95	Oke (1978)
Grass, green	0.16–0.27	0.96–0.98	Van Wijk and Scholte Ubing (1963)
Grass, dried	0.16–0.19		Van Wijk and Scholte Ubing (1963)
Prairie, wet	0.22		Van Wijk and Scholte Ubing (1963)
Prairie, dry	0.32		Van Wijk and Scholte Ubing (1963)
Stubble fields	0.15–0.17		Van Wijk and Scholte Ubing (1963)
Grain crops	0.10–0.25		Van Wijk and Scholte Ubing (1963)
Green field crops full cover, LAI > 3	0.20–0.25		Jensen et al. (1990)
Leaves of common farm crops		0.94–0.98	Jensen et al. (1990)
Most field crops	0.18–0.30		Rosenberg et al. (1983)
Field crops, latitude 22°–52°	0.22–0.26	0.94–0.99	Monteith and Unsworth (1990)
Field crops, latitude 7°–22°	0.15–0.21	0.94–0.99	Monteith and Unsworth (1990)
Deciduous forest	0.15–0.20	0.96[a]	Rosenberg et al. (1983)
Deciduous forest, bare	0.15	0.97	Oke (1978)
Leaved	0.20	0.98	Oke (1978)
Coniferous forest	0.10–0.15	0.97	Rosenberg et al. (1983)
Coniferous forest	0.05–0.15	0.98–0.99	Oke (1978)
Vineyard	0.18–0.19		Rosenberg et al. (1983)
Mangrove swamp	0.15		Rosenberg et al. (1983)
Grass	0.24		Jones (1992)
Crops	0.15–0.26		Jones (1992)
Forest	0.12–0.18		Jones (1992)
Water, high sun	0.03–0.10	0.92–0.97	Oke (1978)
Water, low sun	0.10–1.00	0.92–0.97	Oke (1978)
Sea, calm	0.07–0.08		Rosenberg et al. (1983)
Sea, windy	0.12–0.14		Rosenberg et al. (1983)
Ice, sea	0.30–0.45	0.92–0.97	Oke (1978)
Ice, glacier	0.20–0.40		Oke (1978)
Ice, lake, clear	0.10		Rosenberg et al. (1983)
Ice, lake, w/snow	0.46		Rosenberg et al. (1983)

[a] Van Wijk and Scholte Ubing (1963).

view factor (proportion of sky visible from the soil) on LAI, and incorporated these into their ENergy and WATer BALance model (ENWATBAL). Monteith and Unsworth (1990) presented equations describing the albedo of a deep canopy with a spherical distribution of leaves for sun angles greater than 25°. More discussion of these concepts can be found in Russell et al. (1989). For field studies we can either measure albedo, directly measure the components of net radiation, or use a net radiometer (Table 6.1).

Because of the largely unpredictable and time-varying position of sun flecks below the canopy, transmitted radiation can be measured below the canopy most effectively with one of the linear radiation sensors rather than with point sensors. Linear sensors are typically 1 m long and integrate radiation over their entire length. Some are equipped with diffusers. The tube solarimeter is enclosed in a clear envelope and accepts radiation from all directions. Some other linear sensors are configured to accept only down welling radiation; for example, transmitted photosynthetically active radiation (400–700 nm) can be measured below the canopy with a "line quantum sensor" (LI-Cor, Lincoln, Nebraska, USA) or with a "crop canopy absorption meter" (ICT International, Armidale, New South Wales, Australia). For either, the results may be usefully related to total shortwave irradiance through a linear regression, or the latter sensor may be ordered with sensitivity across the shortwave spectrum from 400 to 950 nm.

6.2.1.4 Spectral Reflection and Radiation

The discussion of emissivities and albedos of surfaces given here is based on a broadband view of irradiance, reflection, and emission that recognizes only shortwave and longwave radiation as presented in Figure 6.7. Although these are arguably the most important features from an energy balance perspective, there is much recent work on multispectral sensing of radiation reflected and emitted from vegetation and soil surfaces (Van Toai et al., 2003). This spectral sensing may be done for only a few relatively narrow bands of radiation in the visible and infrared (IR), or may involve hyperspectral scanning that provides sensing of radiation for every nanometer of the spectrum across a wide range. The advent of fiber optics capable of transmitting both visible and IR light and the development of miniaturized spectrometers (e.g., Ocean Optics, Inc., model S2000) has revolutionized the way researchers view plant and soil surfaces.

An example of multispectral sensing is the use of red and near infrared (NIR) reflectance from a cotton canopy (Figure 6.17). The ratio of NIR/red reflectances is clearly related to LAI. However, the relationship is not stable from year to year and research continues. Other uses include sensing of the onset and progression of plant disease and insect infestation, and sensing of plant water status. Much work remains to be done to make these techniques useful.

6.2.1.5 Incoming Longwave Radiation

Methods of estimating longwave irradiance from the sky, $L\downarrow$, usually take the form

$$L\downarrow = \varepsilon_a \sigma (T_a + 273.16)^4, \tag{6.32}$$

where T_a (°C) is air temperature at the reference measurement level (often 2 m), and the emissivity (ε_a) may be estimated from the vapor pressure of water in air at reference level (e_a) (kPa) or from both e_a and T_a.

The vapor pressure is

$$e_a = \frac{RH}{100} e_s, \tag{6.33}$$
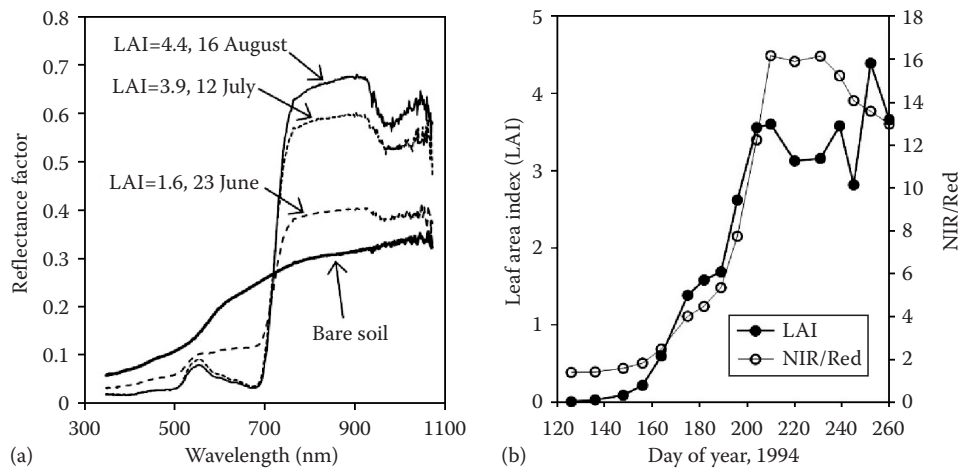
FIGURE 6.17 Reflectance factor versus wavelength for bare soil and cotton at three levels of LAI (a), and the relationship between the ratio of NIR to red reflectance and LAI (b) for cotton in Arizona. (Personal communication from P.J. Pinter Jr. and E.M. Barnes, U.S. Water Conservation Laboratory, USDA-ARS, Phoenix, AZ, May 2000.)

where

RH is the relative humidity of the air

e_s is the saturation vapor pressure (kPa) at T_a (°C) given by (Murray, 1967)

$$e_s = 0.61078 \exp\left(\frac{17.269 T_a}{237.3 + T_a} \right). \tag{6.34}$$

If the dew point temperature, rather than the RH, is known then

$$e_a = 0.61078 \exp\left(\frac{17.269 T_{dew}}{237.3 + T_{dew}} \right). \tag{6.35}$$

Hatfield et al. (1983) compared several methods for estimating ε and concluded that methods using only air temperature performed less well than those that used vapor pressure or both vapor pressure and air temperature. Among the best methods was Idso's (1981) equation

$$\varepsilon_a = 0.70 + 5.95 \times 10^{-4} e_a \exp\left[\frac{1500}{T_a + 273.1} \right], \tag{6.36}$$

where e_a is in kPa. Idso showed fairly conclusively that ε is a function of both e_a and T_a.

Howell et al. (1993) measured $L\downarrow$ (Table 6.1) and calculated ε_a by inverting Equation 6.32. Applying Equation 6.36 as well as Brunt's (1932) equation

$$\varepsilon_a = 0.52 + 0.206 e_a^{0.5} \tag{6.37}$$

and Brutsaert's (1982) equation

$$\varepsilon_a = 0.767 e_a^{1/7} \tag{6.38}$$

to their data shows that all three equations gave good predictions for clear sky conditions but probably underestimated ε_a for cloudy and nighttime conditions (Figure 6.18). For regressions of predicted versus measured ε, the Idso equation gave a slightly greater correlation coefficient and a slope closer to unity (Table 6.5). Under heavy clouds, sky emissivity approaches unity; and none of these models predicts this well.

Choi et al. (2008) found similar results in a humid climate (Florida), where the Brunt and Brutsaert models both performed well under clear sky conditions, but $L\downarrow$ was underestimated under cloudy conditions. They compared seven methods of estimating down welling longwave radiation under clouds, $L\downarrow_c$, and found that the Crawford and Duchon (1999) model $(L\downarrow_c = L\downarrow (1 - c_c) + c_c \sigma T_a^4)$ worked best where c_c was fractional cloud cover.
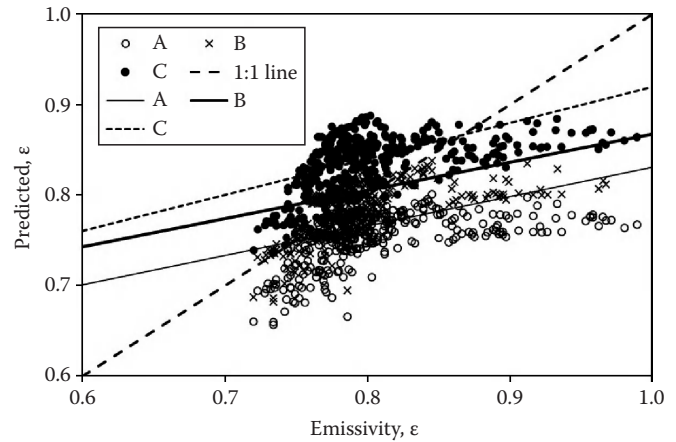
FIGURE 6.18 Comparison of predictions with measured emissivity for two periods in 1992 at Bushland, Texas. Points plotted at extreme right were associated with nighttime and overcast conditions. A, Equation 6.37; B, Equation 6.38; C, Equation 6.36. Lines are for regressions shown in Table 6.5.

TABLE 6.5 Regressions of Predicted Emissivity (ε_p) versus Measured Values (ε) for data from Day 133 through 140 and 192 through 197, 1992, at Bushland, Texas

Method	Regression Equation	r^2	SE
Brunt, Equation 6.37	$\varepsilon_p = 0.505 + 0.325\varepsilon$	0.33	0.024
Brutsaert, Equation 6.38	$\varepsilon_p = 0.556 + 0.311\varepsilon$	0.32	0.024
Idso, Equation 6.36	$\varepsilon_p = 0.522 + 0.398\varepsilon$	0.37	0.027

Despite the difficulty of estimating sky emissivity well, uncertainty in the value of $L\downarrow$ usually causes little difficulty in estimating net radiation for daytime; but $L\downarrow$ can be seriously underestimated at night with important consequences for models of chill stress or frost.

6.2.1.6 Comparison of Net Radiation Estimates with Measured Values

It has become commonplace to have data from field weather stations that include R_{si} (W m^{-2}), and air temperature, T_{az} (°C), wind speed, u_z (m s^{-1}), and relative humidity, RH$_z$ (%), measured at some reference height, z (often 2 m). Measurement of R_n is still not common, probably due to several factors including additional expense, fragility of the plastic domes used on some models of net radiometer, and problems with calibration, including wind speed effects. Net radiometer calibration changes with time; and experience shows that even new radiometers may not agree within 10%. If a net radiometer is used, it is prudent, as with all instruments, to check measured R_n values against estimated ones. Methods presented in previous sections can be used to estimate R_n, but simpler methods exist that are adequate for most cases. Jensen et al. (1990) compared four methods of estimating R_n, including those of Wright and Jensen (1972), Doorenbos and Pruitt (1977), a combination of Brutsaert (1975) and Weiss (1982), and Wright (1982),

against values measured at Copenhagen, Denmark, and Davis, California. The Wright (1982) method was overall best, but underestimated R_n in the peak month at Copenhagen by 9%. The Wright and Jensen (1972) method was almost as good. These methods all assume that surface temperature is not measured, so that only air temperature is used in the calculations.

Jensen et al. (1990) calculated net longwave radiation, R_{nl}, as

$$R_{nl} = -\left[a\left(\frac{R_{si}}{R_{so}}\right) + b\right](a_1 + b_1 e_d^{0.5})\sigma T_{az}^4, \quad (6.39)$$

where

e_d is the saturation vapor pressure of water in air at dew point temperature (kPa)

the term $(a_1 + b_1 e_d^{0.5})$ is a "net emittance," ε', of the surface

The "net emittance" attempts to compensate for the fact that surface temperature is not measured, the assumption being that T_{az} can substitute reasonably well for both sky and surface temperature. The coefficients a, b, a_1, and b_1 are climate specific, with a and b being cloudiness factors. Some values are presented by Jensen et al. (1990, Table 6.3.3).

Many weather stations report only daily totals of solar radiation; and maximum and minimum of air temperature, T_x and T_n, respectively (K). If this is the case, the term σT^4 can be estimated as

$$\sigma T^4 \cong \frac{\sigma(T_x^4 - T_n^4)}{2}. \quad (6.40)$$

If mean dew point temperature is not available, it may be estimated as equal to T_n in humid areas.

Allen et al. (1994a, 1994b) gave slightly modified versions of the methods presented by Jensen et al. (1990) in a proposed FAO standard for reference ET estimation. As an example, estimates of daily total net radiation were made for Bushland, Texas, using the following equations

$$R_n = (1-\alpha)R_s - \left[a_c\left(\frac{R_{si}}{R_{so}}\right) + b_c\right](a_1 + b_1 e_d^{0.5})\sigma\left(\frac{T_m^4 + T_n^4}{2}\right), \quad (6.41)$$

where the cloud factors were $a_c = 1.35$ and $b_c = -0.35$, the emissivity factors were $a_1 = 0.35$ and $b_1 = -0.14$, the albedo was $\alpha = 0.23$, R_{si} was measured, e_d was calculated from mean dew point temperature, and R_{so} was calculated from

$$R_{so} = (0.75 + 0.00002\,EL_{msl})R_{sa}, \quad (6.42)$$

where

R_{sa} is from Equation 6.19

EL_{msl} is elevation (m) above mean sea level

This is similar to Equation 6.20 but with a correction increasing R_{so} for higher elevation sites. The mean daily saturated vapor pressure at dew point temperature was estimated from mean daily dew point temperature, T_d (°C)

$$e_d = 0.611\exp\left[\frac{17.27T_d}{T_d + 237.3}\right]. \quad (6.43)$$

Additionally, half-hourly R_n estimates were calculated from half-hourly measured values of R_{si}, T_a, and T_d using equations given by Allen et al. (1994a, 1994b) equivalent to Equations 6.7, 6.8, 6.10, and 6.12 through 6.16 to estimate half-hourly R_{sa}, and Equation 6.42 to estimate half-hourly R_{so}. Equation 6.43 was applied to half-hourly dew point temperatures to estimate half-hourly e_d values. Equation 6.41 was written for half-hourly values of air temperature, T_a, as

$$R_n = (1-\alpha)R_s - \left[a_c\left(\frac{R_{si}}{R_{so}}\right) + b_c\right](a_1 + b_1 e_d^{0.5})\sigma(T_a^4), \quad (6.44)$$

where the ratio of R_{si} to R_{so} was set to 0.7 for nighttime estimates of R_n.

Comparison of daily R_n estimates, calculated using half-hourly data means, with measurements made with a Radiation and Energy Balance Systems (REBS) Q*5 (Seattle, Washington) net radiometer over irrigated grass show excellent agreement for alfalfa (Figure 6.19) and grass (Figure 6.20) at Bushland, Texas. But, there was a consistent bias for R_n estimated from daily means, with underestimation of R_n at large measured values,
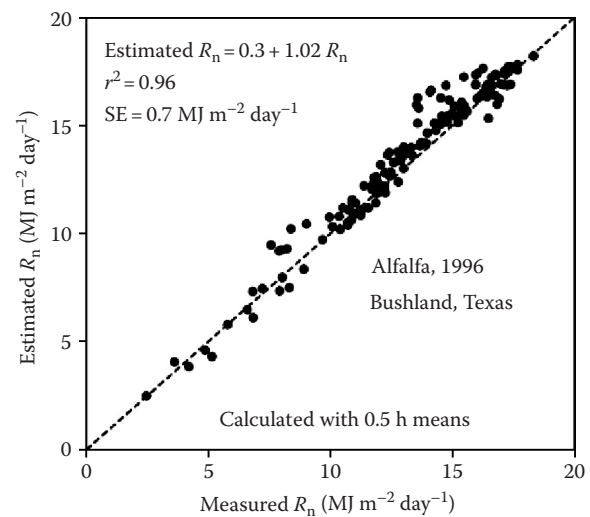
FIGURE 6.19 Daily net radiation, estimated with methods from Allen, R.G., M. Smith, A. Perrier, and L.S. Pereira. 1994a. An update for the definition of reference evapotranspiration. ICID Bull. 43:1–34; Allen, R.G., M. Smith, A. Perrier, and L.S. Pereira. 1994b. An update for the calculation of reference evapotranspiration. ICID Bull. 43:35–92. Using half-hourly data, compared with measurements with a REBS Q*5 net radiometer over sprinkler irrigated alfalfa in 1996 at Bushland, Texas.
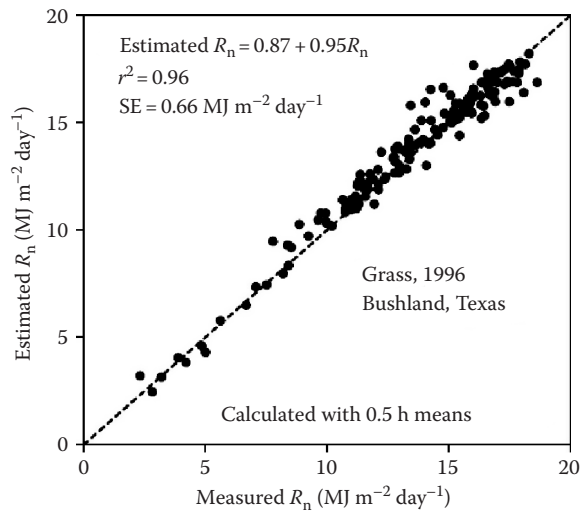
FIGURE 6.20 Net radiation estimated with methods from Allen et al. (1994a, 1994b) compared with measurements over drip irrigated grass in 1996 at Bushland, Texas.
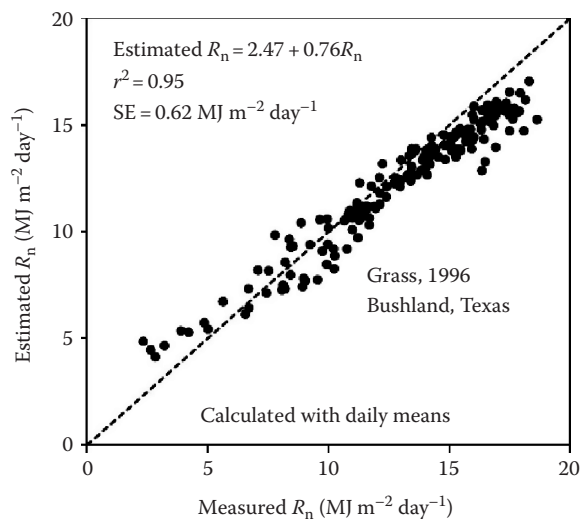
FIGURE 6.21 Daily net radiation, estimated with methods of Allen et al. (1994a, 1994b) using daily means and maxima and minima, compared with measurements with a REBS Q*5 net radiometer over drip irrigated grass in 1996 at Bushland, Texas.

and overestimation at small measured values (Figure 6.21). The bias evident when daily means and maximum/minimum temperatures were used is probably tied to both poor estimates of vapor pressure from the max/min temperature data, and the inadequacy of Equation 6.40.

Estimates of half-hourly net radiation for alfalfa at Bushland, Texas, made using half-hourly data and these methods also gave good results (Figure 6.22). Allen et al. (1994a, 1994b) give detailed methods for estimating R_n when measurements are missing for R_{si} and/or e_d. Irmak et al. (2003) give useful methods for estimating R_n when only measured T_x, T_n, and R_{si} are available, or when only measured T_x, T_n, and RH are available and R_{si} is estimated.
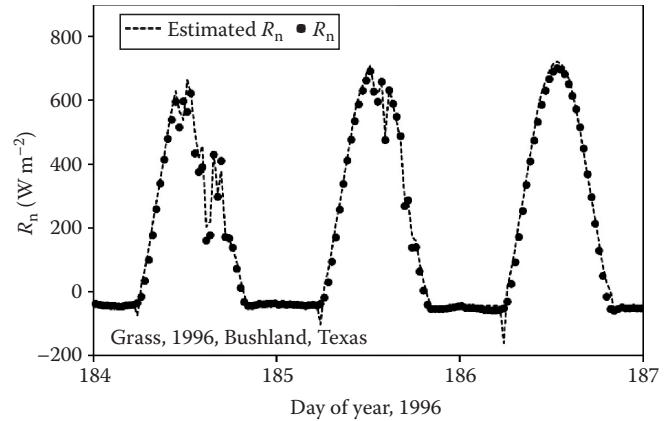
FIGURE 6.22 Half-hourly measured net radiation compared with values estimated with methods of Allen et al. (1994a, 1994b) using half-hourly data for drip irrigated grass at Bushland, Texas, 1996.

6.2.2 Latent Heat Flux

Latent heat flux, LE (W m^{-2}), is the product of the evaporative flux, E (kg s^{-1} m^{-2}), and the latent heat of evaporation, λ (2.44 × 10^6 J kg^{-1} at 25°C). The value of λ is temperature dependent, but is well described (in J kg^{-1} × 10^6) by

$$\lambda = 2.501 - 2.370 \times 10^{-3} T \ \ (r^2 = 0.99995), \tag{6.45}$$

where T is in °C. Methods of E measurement include weighing lysimeters (including microlysimeters), and other mass balance techniques that rely on measurements of change in soil water storage, ΔS. Evaporative flux may also be estimated from other measurement methods such as eddy covariance (EC) and Bowen ratio measurements or from models ranging from the relatively simple analytical form of the Penman–Monteith (P–M) equation to complex field surface energy balance (FSEB) models. The surface renewal method is also a possibility (Snyder et al., 1997), but may be biased in some conditions (Zapata and Martinez-Cob, 2002) and depends on measurements above the crop and knowledge of the measurement height and crop height. Castellví (2004) addressed problems with surface renewal analysis by combining it with similarity theory. The surface renewal approach has so far not become widely used in agricultural micrometeorology and will not be further discussed herein. Because ΔS is a component of the soil water balance, and weighing lysimetry is a key tool for investigations of soil water balance, discussion of lysimetric techniques will be deferred to Section 6.3.

6.2.2.1 Boundary Layers

Evaporative fluxes move between plant or soil surfaces and the air by both diffusion and convection. Diffusive processes prevail in the laminar sublayer close (millimeters) to these surfaces. In this layer, air movement is parallel to the surface and little mixing occurs. Vapor flux across the laminar sublayer is well described by a Fickian diffusion law relating flux rate to vapor pressure gradient factored by a conductance term. But in the

turbulent layer beyond the laminar layer, the flux is mostly convective in nature so that water vapor is moved in parcels of air that are emitted and mixed into the atmosphere in turbulent flow. These moving parcels of air are often referred to as eddies and are analogous to eddies seen in a stream. Usually the eddies are not visible; but in foggy, smoky, or dusty conditions, they may be apparent. Certainly, anyone who has felt the buffeting of the wind can attest to the force of eddies and the turbulence of the airstream in which they occur. As wind speeds increase, the depth of the laminar sublayer decreases. Surface roughness enhances this process, resulting in thinner laminar sublayers. Because the resistance to vapor transport across the laminar airstream is much larger than the resistance across a turbulent airstream of similar dimension, increasing roughness and wind speed both tend to enhance vapor transport. If the air is still, then eddies due to turbulent flow do not exist, but eddies due to free convection may well be present. Free convection occurs when an air parcel is warmer (or colder) than the surrounding air and thus moves upward (or downward) because it is lighter (or heavier). These buoyancy effects can predominate at very low wind speeds when the surface is considerably warmer than the air. As opposed to free convection, transport in eddies due to wind is called forced convection. A full discussion of the fluid mechanics of laminar and turbulent flow, Fickian diffusion, and forced and free convection is well beyond the scope of this chapter. Discussions relevant to soil and plant surfaces are presented in Chapters 7–9 of Monteith and Unsworth (1990), Chapters 3 and 4 of Rosenberg et al. (1983), and Chapter 3 of Jones (1992). Relevant micrometeorological measurement methods are explained in detail in Chapters 16–21 of Hatfield and Baker (2005). Here we will concentrate on some results and methods of measurement. These methods are valid within the constant flux layer (Figure 6.23), a layer of moving air that develops from the point at which the airstream first reaches a surface of given condition, for example, the wheat field shown in Figure 6.23. The air mixes as it moves over the field, equilibrating with the new surface condition, and forming a layer of gradually increasing thickness, δ_e, within which the vertical flux of heat and vapor is constant with height. This is the fully adjusted or equilibrium layer. Within this layer is a layer, extending from the roughness
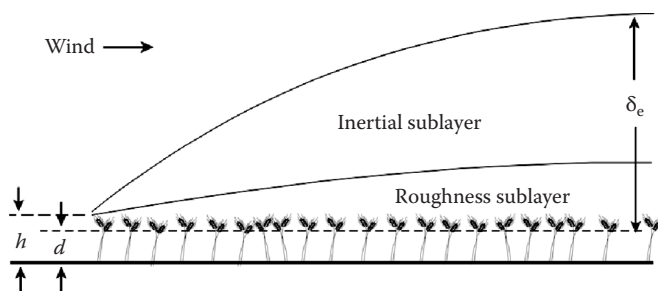
elements (wheat plants in this schema) upward, within which air flow is more turbulent due to the influence of the roughness elements. This is called the roughness sublayer (Monteith and Unsworth, 1990). For any measurement of air temperature, humidity, or wind speed, its fetch is the distance upwind from the point of observation to the edge of the new surface. The ratio of the fetch to the value of δ_e is dependent on the roughness of the surface, the stability of the air, and the wind speed. For many crop surfaces, it may be as small as 20:1 or as large as 200:1. For smooth surfaces such as bare soil, the ratio may well be larger than 200:1. See Section 6.2.2.4 for further discussion of fetch. Measurements should be made in the constant flux layer but above the roughness sublayer.

6.2.2.2 Eddy Covariance

The observation of turbulent flow and concept of eddies lead to the EC (also referred to as eddy correlation) method of turbulent flux measurement, including latent heat flux measurement. It has been described as the most physically correct method of providing a direct measurement of the vertical turbulent flux across the mean horizontal streamlines (Swinbank, 1951), provided that fast response sensors (10–20 Hz) are available for the velocity vector and scalar entity of interest. Conceptually, the vertical velocity component (w) of upward moving eddies covaries with a scalar or mass density (s) and the same is true for the vertical velocity component and scalar of downward moving eddies; so, if the instantaneous product ws is on average larger for the upward moving eddies than for the downward moving ones, then the net flux of the scalar or mass of interest is upward. In this method, fast response sensors are used to measure w and s simultaneously at a rate of, for example, 10–20 Hz, producing estimates of the instantaneous deviations w' and s'. This provides an estimate of the vertical flux at the measurement height (but note fetch requirements below). The generic expression for EC can be described as

$$F_s = \overline{w's'},\qquad(6.46a)$$

where

F_s is the flux of a scalar or mass of interest, normally water vapor, sensible heat, or carbon dioxide

w and s are vertical velocity and scalar measurements, respectively

Primes represent instantaneous deviations from a mean and the over bar indicates a time average operator, typically 30 min. With respect to the turbulent fluxes of the energy balance equation, Equation 6.46a can be recast to represent evaporation, LE, and sensible heat, H, fluxes as

$$\text{LE} = \lambda \overline{w'\rho_v'}\qquad(6.46b)$$

$$H = \rho_a c_p \overline{w'T_a'},\qquad(6.46c)$$

FIGURE 6.23 Schematic of sublayers of the surface boundary layer over a wheat crop. The height (h) of the crop and the depth (δ_e) of the constant flux layer are noted. The height (d) is the zero plane displacement height, which is the height to which a logarithmic wind profile, measured above the crop, would extrapolate to zero wind speed.

respectively, where

λ is the latent heat of vaporization

w' has been defined above

ρ'_v is the estimate of instantaneous water vapor density

ρ_a is the density of air

c_p is the specific heat of moist air

T'_a is the estimate of instantaneous air temperature

The rate of data acquisition must be faster for measurements nearer the surface. Monteith and Unsworth (1990) state that eddy sizes increase with surface roughness, wind speed, and height above the surface; and they suggest 1 kHz rates may be needed near a smooth surface, while 10 Hz or less may be adequate at several meters above a forest canopy. Because the measurements should take place within the fully adjusted boundary layer, simply increasing sensor height will not eliminate the need for fast sensor response. EC methods are challenging to conduct due to the data handling and sensor requirements. Data-processing requirements are large, but modern data logging and computing equipment are capable of handling these. Commercial systems including data-processing software are now available, although expensive (Campbell Scientific Inc., Logan, Utah; and The Institute of Ecology and Resource Management at the University of Edinburgh, Scotland). The sonic anemometer is the wind sensor of choice for EC work due its fast response and sensitivity. At this time, a single-axis unit costs about $2500 and a 3D sonic anemometer costs about $8000. Suitable vapor pressure sensors include the krypton hygrometer and infrared gas analyzer (IRGA) available at this writing in the $6,000–$15,000 range.

When LE and *H* are measured by EC, the performance of this approach may be evaluated (Houser et al., 1998) by rearranging Equation 6.1 to

$$LE + H = -R_n - G \tag{6.47}$$

and measuring R_n and *G* (Figure 6.24). This is referred to as evaluating the energy balance closure. If all measurements in Equation 6.47 are perfectly made, then the sum of the turbulent fluxes (LE + *H*) will equal the available energy ($R_n + G$). But in reality, turbulent flux estimates are usually less than the available energy. The degree of energy imbalance is an indication of how well the EC measurements were made. Energy balance closure can also be expressed as the ratio of the sum of the turbulent fluxes to the available energy, where the typical range considered as generally achievable for LE and *H* fluxes is 0.8–0.9.

Fast response thermocouples for measuring air temperature are used in EC systems for measuring *H*. Because these are very much less expensive than fast response vapor pressure sensors, it is sometimes sensible to measure R_n and *G*, and *H* by EC, and to compute LE as the residual

$$LE = -R_n - G - H. \tag{6.48}$$

While EC is in principle a physically correct approach to measuring turbulent flux exchange between a surface and the boundary
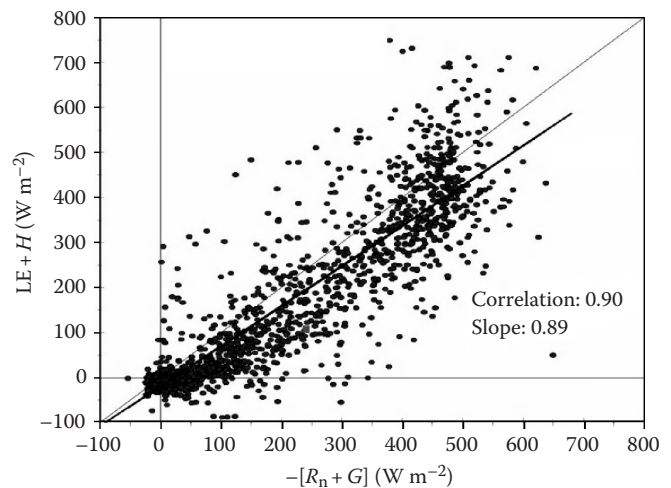
FIGURE 6.24 Check of eddy correlation system LE and *H* values against measured R_n and *G* values, all measured at the Walnut Gulch Experimental Watershed, Lucky Hills Site, on days 194–225, 1996. (Adapted from Houser, P.R., C. Harlow, W.J. Shuttleworth, T.O. Keefer, W.E. Emmerich, and D.C. Goodrich. 1998. Evaluation of multiple flux measurement techniques using water balance information at a semi-arid site, p. 84–87. *In* Proc. Spec. Symp. Hydrol. American Meteorological Society, Phoenix, AZ.)

layer of the atmosphere and in practice is relatively straightforward to make, there are a number of important assumptions that must be considered prior to and during data acquisition. These assumptions are well described by Foken and Wichura (1996) and should always be considered prior to any field campaign. Equally important are the corrections applied to the turbulent fluxes after the measurements are made. It is significantly beyond the scope of this chapter to describe in detail the corrections now considered standard for turbulent flux measurements using EC but we list the more important corrections with references. These include coordinate rotation of the velocity and scalar components (Tanner and Thurtell, 1969; Kaimal and Finnigan, 1994; Paw et al., 2000; Wilczak et al., 2001; Lee et al., 2004), the proportionality of a gas concentration to density effects due to changes in temperature and humidity commonly referred to as the WPL correction or adjustment (Webb et al., 1980; Webb, 1982; Leuning, 2004; Massman, 2004), and frequency response and sensor separation corrections (Moore, 1986; Leuning and Judd, 1996; Massman, 2000; Finnigan et al., 2003; Massman and Clement, 2004). All these need to be considered and implemented to ensure the best possible data quality. Note that Thurtell questioned the use of coordinate rotations at a 2002 symposium in his honor.

Despite 40 years of advances in EC equipment and theory since the work of Tanner and Thurtell (1969), including the corrections described above, closure errors of 10%–30% are still common and the error is a systematic underestimation of fluxes, particularly LE and CO_2 fluxes (Twine et al., 2000; Norman and Baker, 2002). This outcome leads one to question either (1) the description of this method by Swinbank (1951) as the most physically correct method of providing a direct measurement of the vertical turbulent flux across the mean horizontal streamlines or

(2) the relevance of such a measurement to determination of LE. Some of the mechanisms possibly causing lack of closure, such as flux divergence and nonstationarity of flow, are intractable, leading Twine et al. (2000) to recommend that closure be forced routinely in order to avoid cumulative errors in flux estimates over time. Closure can be forced either by calculating LE as the residual of Equation 6.1 or by using the measured Bowen ratio; Twine et al. (2000) preferred the latter, but see the next section for possible problems. Comparisons of EC and Bowen ratio systems are found in Dugas et al. (1991) and Houser et al. (1998). Some specifics of EC system design are given in Unland et al. (1996) and Moncrieff et al. (1997).

6.2.2.3 Bowen Ratio

The Bowen ratio is the ratio of sensible to latent heat flux, $\beta_r = H/\mathrm{LE}$ (Bowen, 1926). Introducing this into Equation 6.1 and rearranging gives the Bowen ratio method for estimating LE

$$\mathrm{LE} = \frac{-(R_n + G)}{\beta_r + 1}. \tag{6.49}$$

In the constant flux layer, it is possible to measure temperature and water vapor pressure differences at two heights, z_1 and z_2 and to evaluate β_r from a finite difference form

$$\begin{aligned} \beta_r &= \frac{K_H \rho_a c_p (T_{a\_z_2} - T_{a\_z_1})/(z_2 - z_1)}{K_V (M_w/(M_a P))\rho_a \lambda (e_{a\_z_2} - e_{a\_z_1})/(z_2 - z_1)} \\ &\approx \frac{c_p (T_{a\_z_2} - T_{a\_z_1})}{(M_w/(M_a P))\lambda (e_{a\_z_2} - e_{a\_z_1})} = \gamma \frac{(T_{a\_z_2} - T_{a\_z_1})}{(e_{a\_z_2} - e_{a\_z_1})}, \end{aligned} \tag{6.50}$$

where the second and third entities assume equivalency of the exchange coefficients for sensible heat flux, K_H, and latent heat flux, K_V; and $\gamma = c_p P/(\lambda M_w/M_a)$ is the psychrometric "constant," so-called because its value changes little with temperature, humidity, and normal variations in air pressure. Commonly, values of T_a and e_a are half-hourly or hourly means. Because the sensor response time does not have to be very short, Bowen ratio equipment is much less expensive than that for EC, with complete systems available for under $10,000. Bowen ratio systems are available from REBS, Seattle, Washington; Campbell Scientific, Inc., Logan, Utah, and others.

Because slight differences in instrument calibration may lead to large errors, it is advisable to switch instruments between the measurement heights. The moving arm system popularized by REBS is one way to do this. Bowen ratio measurements are usually valid only during daylight hours. Near sunset and at night, the sum of R_n and G in Equation 6.49 approaches zero even though LE may continue at considerable rates, particularly in advective environments (Tolk et al., 2006), in which most of the irrigated western United States is included. For periods just after sunrise and before sunset, the gradients of

T_a and e_a may become small at the same time that R_n becomes small, leading to instability in Equation 6.50 and irrational estimates of LE. Also near sunset, the value of β_r may approach −1, leading to a near-zero denominator in Equation 6.49 and poorly defined LE (Todd et al., 2000). Bowen ratio systems tend to underestimate LE when regional and local sensible heat advection occurs (Blad and Rosenberg, 1974; Todd et al., 2000), probably because $K_H/K_V > 1$ under the stable conditions that prevail then (Verma et al., 1978). Four Bowen ratio systems were compared by Dugas et al. (1991) who discuss the merits of different designs. Three EC systems agreed well with each other; but LE measurements from them were consistently lower than those from the four Bowen ratio systems. Note that R_n measurement errors can have a large impact on LE calculated using Equation 6.49.

6.2.2.4 Fetch Requirements

Both EC and Bowen ratio methods are sensitive to upwind surface conditions. The LE and H values from these methods represent an areal mean for a certain upwind surface area, often called the "footprint." Both methods require considerable representative upwind surface or fetch, often extending to hundreds of meters, of surface that is similar to that where the measurement is made, if the measurement is to be representative of that surface. Also, the longer the same-surface fetch is, the deeper is the fully adjusted layer, and the higher the instruments can be placed above the surface. Issues of instrument height and fetch are discussed by Savage et al. (1995, 1996) who recommended placing the sonic anemometer of an EC system no closer than 0.5 m above a short grass cover. Because eddies are smaller nearer the surface, placement of the sonic anemometer too near the surface may lead to eddies being smaller than the measurement window of the anemometer. Fetch requirements may be stated as a ratio of fetch distance to instrument height. Heilman et al. (1989) studied fetch requirements for Bowen ratio systems and concluded that a fetch to height ratio of 20:1 was adequate for many measurements, down from the 100:1 ratios reported earlier. Fetch requirements increase as measurement height, z_m, increases. This poses some additional problems for Bowen ratio systems because these incorporate two sensors and the sensors must be sufficiently separated vertically so that the vapor pressure and temperature gradients between them are large enough to be accurately determined. The rougher the surface is, the smaller the gradients are. For many surfaces, and common instrument resolution, this results in separation distances on the order of a meter. The lower measurement should be above the roughness sublayer, typically at least 0.5 m above a crop (more for a very rough surface such as a forest), so the upper measurement may well be nearly 2 m above the crop surface. This could easily lead to a fetch requirement of 100 m. Analysis of relative flux and cumulative relative flux for an alfalfa field under moderately stable conditions using the methods of Schuepp et al. (1990) leads to rather large fetch requirements (Todd et al., 2000; Figure 6.25). For unstable conditions, mixing is enhanced and the boundary
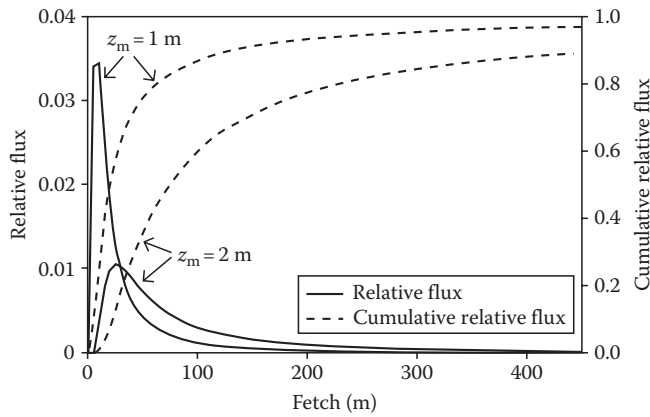
FIGURE 6.25 Relative and cumulative relative flux of an alfalfa field for measurement heights (z_m) of 1 and 2 m, moderately stable thermal conditions, and canopy height of 0.5 m. Cumulative relative flux reaches 0.8 at 65 m for $z_m = 1$ m, and at 225 m for $z_m = 2$ m. (Courtesy of Richard W. Todd, personal communication.)

layer becomes adjusted more quickly over a new surface so that fetch requirements are decreased. However, stable conditions often prevail for irrigated crops in semiarid and arid regions, suggesting that EC and Bowen ratio methods are ill-suited for estimating ET in these settings. Schmid (1997) concluded that fetch requirements are more severe for Bowen ratio than for EC measurements. However, Stannard (1997) argues that Schmid's analysis ignored the gradient measurement approach of Bowen ratio systems and that the fetch requirements are similar for EC and Bowen ratio.

EC methods are sometimes described as making direct measurements of fluxes, and thus are stated to be the only "direct" measure of latent (or sensible) heat flux, particularly of areal extent greater than a few square meters. However, the humidity sensors in EC and Bowen ratio systems do not measure e_a directly, but infer it from measurements of other physical properties, typically light absorption or capacitance changes. Moreover, consideration of fetch requirements leads to a conclusion that both EC and Bowen ratio measurements are representative only for a constantly changing footprint area upwind of the measurement location. The footprint area and the "true" flux are ill-defined because the location and size of the footprint change with wind direction and speed. There is strength in this kind of areal averaging because it reduces noise due to the spatial variability of evaporation. But the measurement cannot be said to be true for any specific location. Indeed, as wind direction changes, the area contributing to the measurement may change completely. By contrast, the soil water balance methods of estimating E, discussed in Section 6.2, provide measures for specific locations. In the case of weighing lysimeters, these are in fact direct measurements of E, specific to a well-defined location, for all times during which precipitation and runoff are not occurring (neglecting the negligible change in plant mass over short periods). Both EC and Bowen ratio measurements are based on the assumption that, within the fully adjusted layer, the vertical mass or energy fluxes are uniform with height, that is, there is no

vertical flux divergence. However, we see from footprint analysis that, for nonzero wind speeds, the upwind area contributes to the measured flux. Even in supposedly uniform fields, there is spatial variability in soil properties and plant responses, so that there is almost always some horizontal flux divergence, and therefore necessarily some vertical flux divergence. Although both methods have been described as "point" measurements, they are really averages over an area, with closer upwind subareas being weighted more heavily and with wind speed and atmospheric stability causing changes in the relative weighting of subareas and the total area involved.

Interest in the spatial variability of the energy balance, particularly the LE component, and in precision farming technologies aimed at addressing crop requirements for water and nutrients at scales well below the field size, has led to a need to measure the spatial variability of LE at scales smaller than can reasonably be addressed with Bowen ratio and EC systems. The radiometric surface temperature can be remotely sensed to provide the spatial variability of LE based on Equations 6.1 and 6.87 (in Section 6.2.4). However, this approach is challenged by difficulties in quantifying surface and aerodynamic resistances and discriminating between crop and soil contributions to the radiometric temperature. There remains a need for ground truth measurements giving LE for a well-defined area. The water balance methods discussed in Section 6.3.1 are capable in many cases, but costs of deployment are practically insurmountable. Thus, there remains a need for inexpensive, accurate, unattended soil profile water content measurement methods for implementation of the water balance method of LE measurement.

6.2.2.5 Field Surface Energy Balance and Remote Sensing

The FSEB has long been recognized for its promise in calculating the latent heat flux, LE, due to ET, from the other energy balance terms in Equation 6.1. When R_n, G, T_s, T_a, h_c, and u_z are measured in the field, Kimball et al. (1999) found that ET (m s^{-1}, positive toward the atmosphere) can be estimated with fair accuracy for full-cover crop surfaces as the residual in

$$\text{ET} = \frac{-\text{LE}}{(\lambda \rho_w)} = \frac{(R_n + G + H)}{(\lambda \rho_w)}, \qquad (6.51)$$

where

$H = \rho_a c_p (T_s - T_a)/r_a$

T_s is assumed equal to the aerodynamic surface temperature, T_o (°C)

r_a is aerodynamic resistance, which is calculated as a function of crop height, h_c, and wind speed, u_z (m s^{-1}), measured at an elevation z (m)

c_p is the heat capacity of air (~1003 kJ kg^{-1} K^{-1})

λ is the latent heat of vaporization (~2.45 × 10^6 J kg^{-1})

ρ_w is the density of water (~1000 kg m^{-3})

ρ_a is air density (kg m^{-3}, $\rho_a \approx 1.291 - 0.00418 T_a$)

In field research, many or all of these fluxes are measured; but from a remote sensing perspective, the surface brightness in the thermal IR (used to estimate T_s) and the brightness in the NIR and visible bands are used to estimate R_n and H. The value of G is taken as a fraction of R_n or some function of R_n and plant cover, which in turn is estimated from a vegetative index based on reflectance in discrete NIR and visible bands. The ET is evaluated as the residual.

Remote sensing–based approaches to FSEB evaluation for ET include both single-surface approaches (e.g., Bastiaanssen et al., 2005; Allen et al., 2007a, 2007b) and two-surface approaches that evaluate the energy and water balances for both canopy and soil surfaces and thus must estimate the crop cover fraction (e.g., Norman et al., 1995; Kustas and Norman, 1999). While providing useful knowledge of regional ET and its spatial and temporal distribution, FSEB predictions that rely on satellite images are not used for infield management due to lack of daily data with sufficient resolution to match appropriate scales for management (Gowda et al., 2008). Attempts to resolve this problem with existing satellite data involve using infrequent, greater resolution data such as that from Landsat (90 m resolution in the thermal IR, 16 d repeat time) in combination with lower resolution more frequent data such as that from MODIS (1000 m resolution in the thermal IR, daily). Such a combination was demonstrated with the DisALEXI (disaggregated atmosphere land exchange inverse) model (Norman et al., 2003; Anderson et al., 2007; Kustas et al., 2007), but no testing of this approach for management has ensued. While aircraft platforms could resolve some of these problems, cost has prevented widespread use of aircraft platforms to provide imaging for management.

The FSEB using remotely sensed data typically provides an instantaneous value of ET, which must be scaled to daily ET using various methods such as the evaporative fraction (which is assumed constant during daylight hours) or reference ET (which is also assumed constant relative to latent heat flux during daylight hours; Colaizzi et al., 2006). Models of the FSEB that use remote sensing include the two-source model (TSM, Norman et al., 1995; Kustas and Norman, 1999), the surface energy balance algorithm for land (SEBAL, Bastiaanssen et al., 1998), and the mapping ET with internalized calibration model (METRIC, Allen et al., 2007a, 2007b). Gowda et al. (2008) reviewed these and several other approaches and reported that daily ET estimation errors ranged from 3% to 35% but that almost all studies compared FSEB ET with ET sensed by Bowen ratio or EC methods, which themselves may have large errors.

Estimates of ET from regional FSEB models using remotely sensed data are often inaccurate for particular field locations due to problems with correctly estimating the surface radiation balance components (Berbery et al., 1999). However, studies using either ground-based or airborne sensor platforms to achieve suitable spatial scales (i.e., a few meters or less), have shown some success with energy balance models for ET (Colaizzi et al., 2003; French et al., 2007).

6.2.2.6 Penman–Monteith Estimates of Latent Heat Flux

Since Penman (1948) published his famous equation describing evaporation from wet surfaces based on the surface energy balance, there have been developments, additions, and refinements of the theory too numerous to mention. Notable examples are the van Bavel (1966) formulation, which includes a surface roughness length term, z_o, and the P–M formula (Monteith, 1965), which includes aerodynamic and surface resistances. The van Bavel equation tends to overestimate in windy conditions and is very sensitive to the value of z_o (Rosenberg, 1969). Howell et al. (1994) compared several ET equations for well-watered, full-cover winter wheat and sorghum and found that the P–M formula performed best. Because it is widely used in agricultural and environmental research, and because it has been presented by ASCE (Jensen et al., 1990; ASCE, 2005) and FAO (Allen et al., 1994a, 1994b, 1998) as a method of computing estimates of reference crop water use, we will discuss the P–M equation, which is expressed as

$$LE = -\frac{\Delta(R_n + G) + \rho_a c_p (e_s - e_a)/r_a}{\Delta + \gamma((1 + r_s)/r_a)}, \qquad (6.52)$$

where

  LE is latent heat flux

  R_n is net radiation

  G is soil heat flux (all in MJ m^{-2} s^{-1})

  Δ is the slope of the saturation vapor pressure–temperature curve (kPa °C^{-1}) commonly evaluated at air temperature

  ρ_a is air density (kg m^{-3})

  c_p is the specific heat of air (kJ kg^{-1} °C^{-1})

  e_a is vapor pressure of the air at reference measurement height z_m

  e_s is the saturated vapor pressure at a dew point temperature equal to the air temperature at z_m (kPa)

  $(e_s - e_a)$ is the vapor pressure deficit

  r_a is the aerodynamic resistance (s m^{-1})

  r_s is the surface (canopy) resistance (s m^{-1})

  γ is the psychrometric constant (kPa °C^{-1})

Penman's equation and those derived from it eliminated canopy temperature from energy balance considerations. Besides measurements of R_n and G, the user must know the vapor pressure of the air, e_a, and air temperature (from which e_s may be calculated) at reference measurement height, z_m (often 2 m). The use of e_s as a surrogate for the (unknown) substomatal vapor pressure introduces the assumptions that osmotic potential of the leaf water has little affect on the substomatal vapor pressure and that the difference between air and canopy temperature does not introduce much error in the estimation of vapor pressure. To the extent that these assumptions are not true, the errors are merged into the resistance terms in Equation 6.52. The values of r_a and r_s may be difficult to obtain. The surface or canopy resistance is known for only a few crops

and is dependent on plant height, leaf area, irradiance, water status of the plants, species, and probably variety.

Jensen et al. (1990) and Allen et al. (1994a, 1994b) presented methods of calculating LE for well-watered, full-cover grass and alfalfa. The following example, drawn from recent studies at Bushland, Texas (Evett et al., 1998, 2000; Todd et al., 2000), employs those methods. Aerodynamic resistance was estimated for neutral atmospheric conditions from

$$r_a = \frac{\ln((z_m - d)/z_{om})\ln((z_H - d)/z_{oH})}{k^2 u_z},$$ (6.53)

where

z_m (m) is the measurement height for wind speed, u_z (m s^{-1})

z_H (m) is measurement height for air temperature and relative humidity

k is 0.41

z_{om} and z_{oH} are the roughness length parameters for momentum (wind) and sensible heat transport

d is the zero plane displacement height

The value of r_a calculated from Equation 6.53 will be too large for highly unstable conditions and too small for very stable conditions. Stability corrections should be made to Equation 6.53 for those conditions (see Monteith and Unsworth, 1990, p. 234 for some examples) but were not made for this example.

Surface resistance was calculated from

$$r_s = \frac{r_l}{(0.5\text{LAI})},$$ (6.54)

where r_l is the stomatal resistance taken as 100 s m^{-1}, and the LAI was taken as

$$\text{LAI} = 5.5 + 1.5\ln(h_c),$$ (6.55)

where the crop height, h_c, was taken as 0.12 cm for grass and 0.5 m for alfalfa.

The zero plane displacement height, d, was calculated as

$$d = \frac{2}{3}h_c.$$ (6.56)

The roughness length for momentum, z_{om}, was calculated as

$$z_{om} = 0.123h_c$$ (6.57)

and the roughness length for sensible heat transport was

$$z_{oH} = 0.1z_{om}.$$ (6.58)
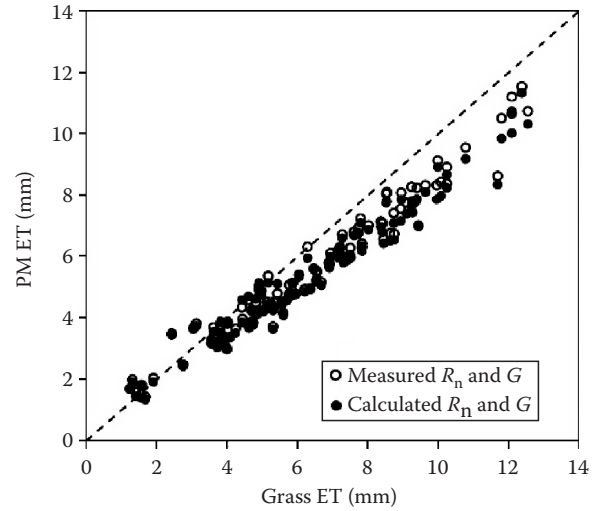
FIGURE 6.26 Daily P–M estimates of ET using both measured and estimated values of R_n and G were not significantly different from each other for well-watered, full-cover mixed fescue grown at Bushland, Texas, in 1996. Both P–M and ET values were less than values measured by a weighing lysimeter for values above 4 mm day^{-1}.

Net radiation was calculated as shown in Section 6.2.1.6. All calculations were on a half-hourly basis. For well-watered mixed fescue grass in 1996, the P–M equation underestimated ET, as measured by a weighing lysimeter, at ET rates exceeding 4 mm day^{-1} (Figure 6.26), even though R_n and G were well estimated. The underestimation of ET was due to systematic error in the surface and/or aerodynamic resistances. For well-watered, full-cover alfalfa in 1996, the P–M estimates of ET were close to values measured with a weighing lysimeter (Figure 6.27). Because R_n and G were well estimated, it is presumed that r_a and r_s were predicted well also. Examination of diurnal dynamics showed that the P–M method was capable of closely reproducing those dynamics.
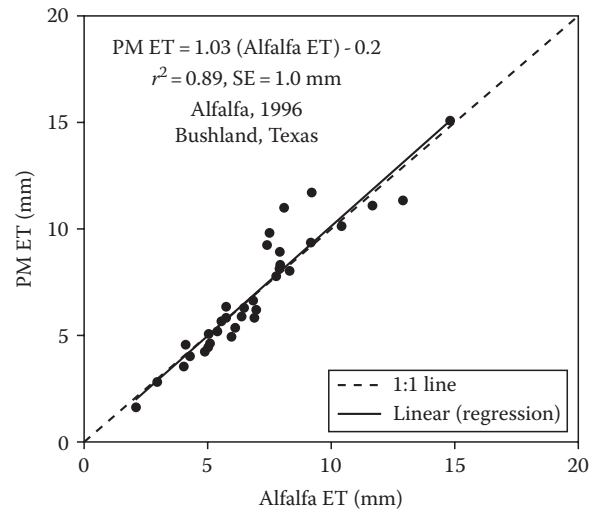
FIGURE 6.27 Daily P–M estimates of ET using estimated R_n and G for well-watered, full-cover (LAI > 3) alfalfa at Bushland, Texas, in 1996.

Although important as a research model, the P–M method is not much used for direct prediction of LE due to the difficulty of knowing r_a and r_s. However, it is commonly used to predict a theoretical reference ET, ET_o for grass and ET_r for alfalfa, for use in irrigation scheduling (Allen et al., 1994a, 1994b, 1998). In this application, crop water use or ET is estimated from daily values of ET_r and a dimensionless crop coefficient (K_c), which is dependent on the crop variety, crop growth, and the reference ET (ET_o or ET_r) used, and which is often cast as a function of time since planting or growing degree days.

$$ET = K_c ET_r. \qquad (6.59)$$

The crop coefficients are determined from experiments that measure daily crop water use, ET, and that measure or, more commonly, estimate ET_r and then compute

$$K_c = \frac{ET}{ET_r}. \qquad (6.60)$$

Many details on this methodology are found in Jensen et al. (1990) and ASCE (2005).

6.2.2.7 Limitations of the Penman Approximation

At the time that Penman (1948), Monteith (1965), and van Bavel (1966) developed their equations for LE, it was very difficult to measure surface temperature of water or plant canopies. All of these equations are called combination equations because they derive from the combination of the energy balance terms given in Equation 6.1 with heat and mass transfer mechanisms. The transfer mechanisms are usually stated as flux equations in terms of resistance(s) or conductance(s) and a gradient of temperature or vapor pressure from the surface to the atmosphere. Penman (1948) stated Equation 6.1 for a wet surface and used transport mechanisms for LE and H to derive a combination equation

$$LE = f(u)(e_z - e_o^*) = -\left[R_n + G + \gamma f(u)(T_z - T_o) \right], \quad (6.61)$$

where

$f(u)$ is a wind speed dependent conductance or transport coefficient
γ is the psychrometric constant ($c_p P/(0.622\lambda)$)
e_o^* is the saturation vapor pressure at the surface temperature
e_z is the air vapor pressure at measurement height z_m
T_z is the air temperature at measurement height
T_o is the surface temperature
the transport mechanism for H is analogous to Equation 6.89

The equation can use e_o^* because it is assumed that the surface is "wet," that is, a free water surface. With an appropriate choice of $f(u)$, the surface can be the canopy of a crop that is full cover and well supplied with water and thus freely transpiring. Because surface temperature was difficult to measure, Penman (1948) introduced an approximation for $T_o - T_z$ that was derived from
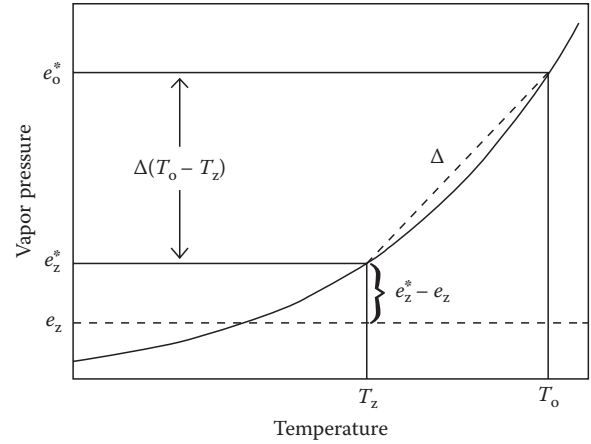
FIGURE 6.28 Quantities used in the derivation of Penman's (1948) combination equation.

the slope of the saturation vapor pressure versus temperature curve (Figure 6.28), which is

$$\Delta = \frac{(e_o^* - e_z^*)}{(T_o - T_z)}. \qquad (6.62)$$

From Equation 6.62 and Figure 6.28, we see that

$$e_o^* - e_z = e_z^* - e_z + \Delta(T_o - T_z). \qquad (6.63)$$

Rearranging to find $(T_z - T_o)$ and substituting into Equation 6.61 gives

$$f(u)(e_z - e_o^*) = -\left[R_n + G + \left(\frac{\gamma}{\Delta}\right) f(u)(e_z - e_o^* + e_z^* - e_z) \right] \qquad (6.64)$$

or

$$f(u)(e_z - e_o^*) = -\left[R_n + G + \left(\frac{\gamma}{\Delta}\right) f(u)(e_z - e_o^*) + \left(\frac{\gamma}{\Delta}\right) f(u)(e_z^* - e_z) \right], \qquad (6.65)$$

both sides of which hold the identity for $LE = f(u)(e_z - e_o^*)$. Rearranging gives

$$LE = -\frac{[\Delta(R_n + G) + \gamma f(u)(e_z^* - e_z)]}{(\Delta + \gamma)}, \qquad (6.66)$$

which is the Penman equation and is analogous to Equation 6.52. Rather than using wind functions of the transport coefficient, Monteith recognized the importance of surface resistance

and derived the equation in terms of the aerodynamic and surface resistances shown in Equation 6.52.

Although there is much evidence that the Penman equation and Equation 6.52 are useful for estimating LE from free water and from well-watered, full-cover crops, the results have been disappointing for transfer of wind functions and r_s and r_a formulations between different regions and crops. Some facts about the underlying assumptions should lend insight. First, the value of Δ is usually evaluated at the air temperature, T_z. In humid climes, there may be little difference between canopy and air temperatures, particularly if skies are often overcast. But in more arid regions, the canopy temperature of a freely transpiring crop may be several degrees cooler than air temperature, causing Δ to be overestimated—more so at the hottest part of the day, when transpiration rates are greatest, than near sunrise and sunset. Second, the net radiation and soil heat fluxes are modified by the quotient $(1 + \gamma/\Delta)$. This is fundamentally incorrect. The value of γ/Δ ranges from 0.80 at 10°C to 0.17 at 40°C; so the modification of R_n and G is not small and will change with air temperature over the course of a day. Third, the surrogate in the Penman equation for sensible heat flux, $\gamma f(u)(e_z^\star - e_z)/(\Delta + \gamma)$, is parameterized by vapor pressure terms defined at the same height. Thus, unlike real sensible heat flux, the surrogate quantity $\gamma f(u)(e_z^\star - e_z)/(\Delta + \gamma)$ can never reverse sign (necessarily, $e_z^\star \geq e_z$). Also, over the range from 10°C to 40°C, the value of γ varies from 65.5 to 67.5 Pa K^{-1}, while the value of Δ varies from 85 to 402 Pa K^{-1}. To the extent that γ is constant, the division of the vapor pressure deficit by $(\Delta + \gamma)$ causes the surrogate term to vary as H would vary. To the extent that the sum $(\Delta + \gamma)$ varies from Δ, and to the extent that $e_z^\star$ differs from $e_o^\star$, the surrogate differs in value from H.

In Equation 6.52, the sum of R_n and G is further modified by the surface and aerodynamic resistances to mass and heat transfer, even though those resistances have negligible effect on the fundamental mechanisms affecting either quantity. (The effect of r_s and r_a on canopy surface temperature has a negligible effect on R_n.) These facts have much to do with the difficulties encountered in determining appropriate values of r_a and r_s and in transferring these values from one region to another. Also, the fact that wind functions for the Penman equation have been determined to be different for different climates is certainly related to the approximations used in the derivation of the Penman and related equations.

At the time that the combination equations were being developed, the instrumentation for measuring net radiation was crude but, compared with that for measuring surface temperature, effective. In the 1960s, the development of infrared radiometers allowed the first radiometric measurements of surface temperature on a large scale, leading to much research on the use of surface temperatures to solve H using forms of Equation 6.89 (see Section 6.2.4), and thus to estimate LE from Equation 6.1. McNaughton (1988) pointed out problems with this method that persist to this day. They include the fact that radiometric surface temperature often differs from aerodynamic surface temperature (the surface temperature that works in Equation 6.89), the difficulty

of evaluating r_a in Equation 6.89 for many surfaces (e.g., partial or mixed canopies), and the spatial heterogeneity of surfaces that leads to spatial heterogeneity of air temperature. However, the continued development of IRTs has led to easy and reliable surface temperature measurement with solid state devices such as the thermocouple IRT (e.g., the model IRt/c from Exergen, Inc.). Meanwhile, the technology for net radiation measurements has improved, but calibration standards do not exist and it is still common to find differences of 10% or more between competing instruments. Just as the lack of adequate instrumentation for measurement of surface temperature affected the development of theory and practice in the mid-twentieth century, the shift in instrumentation capabilities is now affecting much of the experimental physics and development of theory and practice for calculation of surface energy and water balances as discussed in Section 6.2.2.5.

6.2.2.8 Recursive Estimates of Latent Heat Flux

In order to avoid the limitations of the Penman (1948) approximation for $(T_o - T_z)$, several efforts have focused on iterative or recursive solution of the surface energy balance equations, which are implicit in T_o, without resorting to any assumptions. It has long been recognized that only by iterative solution of the implicit energy balance equations can these be solved with complete accuracy (Budyko, 1958; Tracy et al., 1984; McArthur, 1990, 1992; Milly, 1991). Iterative solutions have been used in computer models of the general surface energy balance (Bristow, 1987), of evaporation from bare soil (Lascano and van Bavel, 1983, 1986), and of ET from plant and soil surfaces (Lascano et al., 1987; Evett and Lascano, 1993).

Even though iterative solutions have long been available on personal computers and even possible on hand-held calculators, they have not yet supplanted the P–M approach for calculating reference ET, ET_o, although that is changing. As an alternative to the P–M equation, Lascano and van Bavel (2007) applied a recursive method, attributed to Budyko (1958), in which ET and T_o were found by iteration, satisfying the surface energy balance. Particularly, when $T_a \gg T_o$ and evaporative demand was large, the P–M method underestimated reference ET by as much as 25%. They concluded that the P–M method will underestimate ET in most cases, with the error increasing as evaporative demand increases (larger values of $T_a - T_o$ and smaller values of RH). Widmoser (2009) compared an iterative solution with the P–M method and found the P–M solutions for ET to deviate by as much as −40% to +9% and deviation was greater for smaller time steps (e.g., hourly versus daily). Negative errors were larger when T_a was larger, RH was smaller and the available energy $(R_n + G)$ was smaller. Positive errors increased when RH and T_a were both large while $(R_n + G)$ was small, or when $(R_n + G)$ and T_a were both large and the ratio r_s/r_a was large (large surface resistance and small aerodynamic resistance, for example, tall, stressed plants, and windy conditions).

These analyses give further insight into the problems encountered when transferring crop coefficients between regions with different climates when those K_c values were determined using

P–M based reference ET values. Allen et al. (1994a) provided evidence for this lack of transferability by comparing the estimated ratio of alfalfa to grass reference ET across six arid and five humid locations. The ratio varied considerably across locations, most dramatically between arid and humid locations. For most locations, there was also a difference between the ratio for the peak month and the mean ratio for that location. Note that this variance of ratios applies equally as well to the ratio of a particular crop ET to reference ET (i.e., the crop coefficient, ET/ET_r) thus calling into question the transferability of crop coefficients. Evett et al. (2000) compared alfalfa and grass P–M reference ET formulas to measured ET for alfalfa and grass grown under reference ET conditions and found that the ratio of alfalfa to grass reference ET was not well predicted by the P–M formulations for their windy, semiarid advective environment, thus supporting the findings of Allen et al. (1994a). Shuttleworth (2006) presented an analysis of K_c as used with the P–M method in which he showed that K_c values depend explicitly on the climatic conditions at the time of K_c determination. For a variety of different plant heights and canopy resistances, Annandale and Stockle (1994) used an energy balance model to study variability of full-canopy-cover K_c, as influenced by changes in solar radiation, air temperature, the vapor pressure deficit ($e_z − e_o^*$), and wind speed. Variability in K_c increased as crop height increased and as r_s decreased. Variability in K_c decreased if an alfalfa reference ET was used rather than a grass reference ET; and they recommended (1) using alfalfa reference ET and (2) development of methods for directly estimating crop ET. In light of the discussion in Sections 6.2.2.7 and 6.2.2.8, we add to this list a recommendation to pursue a reference ET based on iterative solution of the energy balance equations.

6.2.2.9 Bare Soil Evaporation Estimates

Evaporation from bare soil (E_S) may be estimated using the EC and Bowen ratio methods previously described with consideration given to the reduced surface roughness of many bare soil surfaces when compared with vegetated surfaces. Also, Section 6.3 describes measurements of E_S using weighable lysimeters of various sizes. Here we describe some alternatives based on energy balance solutions.

Fox (1968) and later Ben-Asher et al. (1983) and Evett et al. (1994) described a bare soil evaporation (E_S) prediction method based on subtracting the energy balance equations (Equation 6.1) written for a dry and a drying soil. Because E_S is zero for a dry soil, this gives an expression for E_S from the drying soil in terms of the other energy balance terms. The method requires a column of dry soil embedded in the field of drying soil and measurements of the surface temperatures of the dry soil and of the drying field soil. The surface temperature difference between the dry and drying soils explains most of E_S, but prediction accuracy is only moderately good ($r^2 = 0.82$ for daily predictions, Evett et al., 1994). Evett et al. showed that the aerodynamic resistance over the dry soil surface was reduced and that the resistance was relatively independent of wind speed, probably due to buoyancy of air heated over the relatively hot, small surface. They also

showed that consideration of the soil albedo change with drying could improve the E_S estimates. Qiu et al. (1999) replaced the aerodynamic resistance with a soil evaporation transfer coefficient: $h_a = (T_s − T_a)/(T_d − T_a)$, where T_s is the surface temperature of a drying soil, T_d is the surface temperature of a dry soil, and T_a is air temperature. They found good agreement dynamically, although their model consistently overestimated soil E_S by 5%. Although the method shows promise, it does not provide an estimate of surface soil water content that would be needed to calculate the albedo change.

When soil is wet, the evaporative flux can be estimated using the Penman or P–M equations with surface resistance set to an appropriate low value (Howell et al., 1993). This wet period is the energy-limited stage of evaporation. As the soil dries, E_S becomes limited by soil properties. Van Bavel and Hillel (1976) addressed this using a finite difference model of soil water and heat flux that later was developed into the CONSERVB model of evaporation from bare soil (Lascano and van Bavel, 1986). This model described 1D soil water movement with Darcy's law, including the dependence of hydraulic conductivity, K (m s^{-1}), on soil water potential, h (m), and the soil water retention function, $θ_v(h)$. The surface energy balance was solved implicitly for surface temperature (T_s), resulting in calculated values of E_S, H, R_n, and G at each time step. The value of E_S was used as the upper boundary condition for soil water flux at the next time step. The elements of CONSERVB were included in the ENWATBAL model by Lascano et al. (1987) who included a vapor transport algorithm as a function of soil drying. The latter model was upgraded to model albedo changes dependent on surface soil water content by Evett and Lascano (1993). The 1993 version of ENWATBAL was shown to more accurately predict E_S than either the Penman or P–M equations (Howell et al., 1993). These detailed models of the soil–plant–atmosphere continuum include numerical simulation of soil heat and water fluxes that depend on detailed information on the soil hydraulic properties (characteristic curves) and property changes with depth.

Kustas (2002) summarizes the bulk transport approaches to E_s estimation that depend on the vapor pressure gradient between the soil surface and the air and may include a drying front algorithm; and he contrasts these more mechanistic approaches with several analytical methods that involve a bulk soil resistance and are more amenable to inclusion in models using remotely sensed data for large area estimation of evaporation.

6.2.2.10 Transpiration

Transpiration, the water evaporated from the plants primarily through stomata, is a significant contributor to total latent heat flux and is directly related to the total dry matter production of plant communities (de Wit, 1958). Methods of determining transpiration (E_T) vary widely. The simplest approach has been to weigh single plants in sealed containers (Briggs and Shantz, 1913, 1914). Canopy level measurement of the total evaporative flux (ET) using such methods as lysimetry, Bowen ratio, and EC requires that soil water evaporation (E_S) be determined, so that E_T can be estimated as the residual, or $E_T = ET − E_S$.

The E_S component may be measured using soil microlysimeters (see Section 6.3.1). This is the most direct method of determining E_T without sealing the soil surface.

Another method has been to estimate E_T using equations based on the surface energy balance and gradients in water vapor concentrations and resistances to vapor and heat exchange between the plant canopy and the air. A representative equation is the widely used P–M formula (1965). The difficulty in using such equations is derivation of bulk canopy resistances to vapor and heat transport that also exclude contributions from the soil. One approach to measure canopy resistance has been to first make stomatal resistance measurements on a portion of a leaf using a porometer (Turner, 1991) or gas analyzer system (Rochette et al., 1991). The measurement strategy must somehow capture the totality of the water loss from the plant canopy as affected by differences in leaf illumination, leaf angles, microclimates, and aerodynamic resistances throughout the canopy. The stratified measurements must then be integrated to a whole canopy resistance, which is a process that has met with limited success (Baldocchi et al., 1991; Rochette et al., 1991). The bulk aerodynamic resistance must also be correctly estimated for the P–M equation, which is also complicated by its possible interactive relationship with bulk canopy resistance (Paw and Meyers, 1989). Furon et al. (2007) used numerical simulations to assess the leaf-to-canopy scale translation of surface resistances and their impact on the performance of the P–M equation. Raupach and Finnigan (1988) investigated both single- and multilayer models of evaporation.

Because of these difficulties, techniques to measure plant sap flow as an approximation of whole plant transpiration have advanced. Sap flow measurement systems use the application of heat to the plant stem to serve as a tracer of the sap movement through the xylem tissue. Currently, there are three groups of heat systems to determine sap flow: stem heat balance, heat dissipation, and heat pulse (Gonzalez-Altozano et al., 2008). The stem heat balance method is the most suitable method for herbaceous annual plants with smaller stem diameters, while the heat pulse and heat dissipation methods are primarily used for woody plants, especially those with large trunk diameters. In the stem heat balance method, a heater strip constantly heats a small section of the entire plant stem, and temperature sensors below and above the strip measure how much heat is moved away from the region by the sap. The specific heat capacity and density of the sap are then used to convert the measurements into volumetric mass sap flow. Over- or underestimating sap flow can occur during periods of small or large flow rates. The heat pulse and heat dissipation methods use probes inserted into the stem, with the heat applied either as a pulse or at a constant rate. Temperature sensors in the probes determine sap velocity. Probe location and cross-sectional area and geometry of the stem must be accounted for to extrapolate probe measurements into whole plant sap flow for both methods, and calibration may be required. In all methods, a strategy for transforming single plant measurements to field scale E_T must be developed. Commonly used strategies include weighting by leaf area or stem or trunk diameter (Cohen and Li, 1996).

6.2.3 Soil Heat Flux

Soil heat flux is discussed in detail in Chapter 9. Additional information related to the surface energy balance is given here. Briefly, heat conduction in one dimension is described by a diffusion equation:

$$C\frac{\partial T}{\partial t} = k_T \frac{\partial}{\partial z}\left[\frac{\partial T}{\partial z}\right], \qquad (6.67)$$

where the volumetric heat capacity, C (J m^{-3} K^{-1}), and the thermal conductivity, k_T (J s^{-1} m^{-1} K^{-1}), are assumed constant in space; and vertical distance is denoted by z, time by t, and temperature by T.

The 1D soil heat flux, G, for a homogeneous medium is described by

$$G = -k_T \frac{\partial T}{\partial z}. \qquad (6.68)$$

The thermal conductivity is a single-valued function of water content and is related to the thermal diffusivity, D_T (m^2 s^{-1}), by

$$k_T = D_T C, \qquad (6.69)$$

where the volumetric heat capacity, C (J m^{-3} K^{-1}), can be calculated with reasonable accuracy from the volumetric water content, θ_v (m^3 m^{-3}), and the soil bulk density, ρ_b (Mg m^{-3}), by

$$C = \frac{2.0 \times 10^6 \rho_b}{2.65} + 4.2 \times 10^6 \theta_v + 2.5 \times 10^6 f_o \qquad (6.70)$$

for a soil with a volume fraction, f_o, of organic matter (Hillel, 1980).

Table 6.6 lists thermal conductivities at "wet" and "dry" points for several soils. For coarse soils, the thermal conductivity versus water content curve is S-shaped (see e.g., Campbell et al., 1994), with a rapid rise at water contents corresponding to about 33 kPa soil water tension (about "field capacity"). For fine soils, the relationship is more linear; and the thermal conductivity between dry and wet conditions in Table 6.6 can be linearly interpolated from the values given, with reasonably small errors. But for water contents below the "dry" value, the thermal conductivity should be taken as the value corresponding to the "dry" state.

De Vries (1963) developed a method of estimating soil thermal conductivity from soil texture, bulk density, and water content. The method, while including most important soil properties affecting conductivity, is limited in that it requires knowledge of parameters called shape factors that describe how the soil particles are packed together. The shape factors are specific to a given soil and perhaps pedon and must be measured. They are, in effect, fitting parameters (e.g., Kimball et al., 1976). De Vries' method tends to overestimate thermal conductivity at

TABLE 6.6 Thermal Conductivity, k_T, of Some Soil Materials

Soil	Dry θ_v	k_T W m^{-1} K^{-1}	Wet θ_v	k_T W m^{-1} K^{-1}	ρ_b Mg m^{-3}	Source
Fairbanks sand	0.003	0.33	0.18	2.08	1.71	de Vries (1963)
Quartz sand	0.00	0.25	0.40	2.51	1.51	de Vries (1963)
Sand	0.02	0.9	0.38	2.25	1.60	Riha et al. (1980)
Sand	0.00	0.27	0.38	1.77	1.64	Watts et al. (1990)
Sand	0.003	0.32	0.38	2.84	1.66	Howell and Tolk (1990)
Gravelly coarse sand (pumice)	0.02	0.13	0.40	0.52	0.76	Cochran et al. (1967)
Medium and coarse gravel (pumice)	0.01	0.09	0.43	0.39	0.44	Cochran et al. (1967)
Loamy sand	0.01	0.25	0.40	1.59	1.69	Sepaskhah and Boersma (1979)
Loam	0.01	0.20	0.60	1.05	1.18	Sepaskhah and Boersma (1979)
Avondale loam	0.08	0.46	0.23	0.88	1.35–1.45	Kimball et al. (1976)
Avondale loam	0.03	0.31	0.30	1.20	1.40	Asrar and Kanemasu (1983)
Silt loam	0.09	0.40	0.50	1.0	1.25	Riha et al. (1980)
Yolo silt loam	0.14	0.49	0.34	1.13	1.25	Wierenga et al. (1969)
Muir silty clay loam	0.03	0.30	0.30	0.90	1.25	Asrar and Kanemasu (1983)
Silty clay loam	0.01	0.20	0.59	1.09	1.16	Sepaskhah and Boersma (1979)
Pullman silty clay loam	0.07	0.16	0.29	0.89	1.3	Evett (1994)
Healy clay	0.04	0.30	0.30	0.91	1.34	de Vries (1963)
Fairbanks peat	0.03	0.06	0.61	0.37	0.34	de Vries (1963)
Forest litter	0.02	0.10	0.55	0.40	0.21	Riha et al. (1980)

water contents above about 0.15 (Asrar and Kanemasu, 1983; Evett, 1994). Campbell et al. (1994) developed modifications of de Vries' theory that allowed them to match measured values well. They showed that, as temperature increased, the thermal conductivity versus water content curve assumed a pronounced S-shape for the eight soils in their study, with the curve deviating from monotonicity at temperatures above 50°C.

Horton et al. (1983) developed a measurement method for D_T based on harmonic analysis. The method entailed fitting a Fourier series to the diurnal soil temperature measured at 1 h intervals at 0.01 m depth followed by the prediction of temperatures at a depth, z (0.1 m), based on the Fourier series solution to the 1D heat flux problem using an assumed value of D_T. The value of D_T was changed in an iterative fashion until the best fit between predicted and measured temperatures at z was obtained. The best fit was considered to occur when a minimum in the sum of squared differences between predicted and measured temperatures was found (i.e., minimum sum of squared error, [SSE]). Poor fits with this and earlier methods are often due to the fact that field soils usually exhibit increasing water content with depth and changing water content with time, while the method assumes a homogeneous soil. Costello and Braud (1989) used the same Fourier series solution and a nonlinear regression method, with diffusivity as a parameter to be fitted, for fitting the solution to temperatures measured at depths of 0.025, 0.15, and 0.3 m.

Neither Horton et al. (1983) nor Costello and Braud (1989) addressed the dependency of diffusivity on water content or differences in water content between the different depths. Other papers have dealt with thermal diffusivity in nonuniform soils but did not result in functional relationships

between thermal properties and water content, probably due to a paucity of depth-dependent soil water content data (Nassar and Horton, 1989, 1990). Soil water content often changes quickly with depth, time, and horizontal distance. Moreover, diffusivity is not a single-valued function of soil water content and so is difficult to directly use in modeling. The ability of time domain reflectometry (TDR) to measure water contents in layers as thin as 0.02 m (Baker and Lascano, 1989; Alsanabani, 1991) provided the basis for design of a system that simultaneously measures water contents and temperatures at several depths.

Evett (1994) used measurements of soil temperature at several depths (e.g., 2, 4, 6, 8 … cm), coupled with TDR measurements of soil water content at the same depths, to find a relationship between thermal conductivity and water content in a field soil. He used the minimum SSE method of Horton to find the thermal diffusivity for each soil layer between vertically adjacent measurements of water content and temperature. The water content for this layer was used to calculate C and thus λ corresponding to that water content. A function of k_T versus θ_v was developed by regression analysis on the k_T and θ_v data (Figure 6.29). Because both C and k_T were known for each layer, this method also gave the soil heat flux.

The complicated methods of measuring thermal diffusivity and conductivity mentioned here use computer programs and nonlinear regression fitting of multiterm sine series in order to handle diurnal temperature waves that differ from simple sinusoidal waves, as well as to incorporate measurements of soil water content that vary with depth. However, rough estimates of the thermal parameters may be made from phase differences and amplitude differences observed for temperatures measured
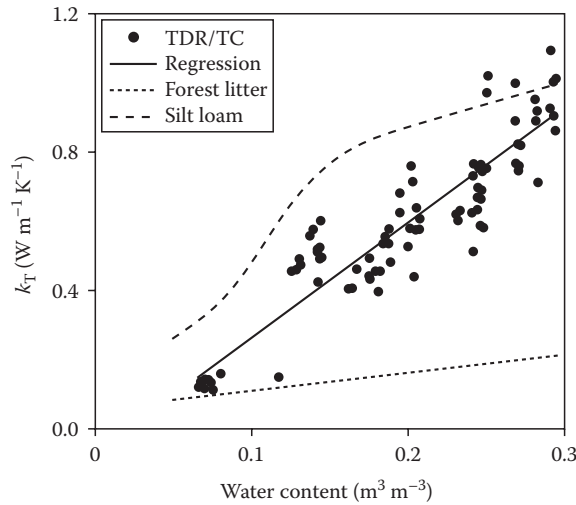
FIGURE 6.29 Thermal conductivity (k_T) of Pullman silty clay loam, determined from TDR probe and thermocouple arrays, compared with functions from Campbell (1985) for forest litter and silt loam.

FIGURE 6.30 Temperatures at six depths in a Pullman clay loam for 2 days in 1993.

at only two depths, with the application of simplifying assumptions of homogeneous water content and soil properties, and a simple sinusoidal diurnal temperature wave describing the temperature T at depth z and time t

$$T(z,t) = \overline{T} + A_z \sin[\omega t + \phi(z)], \qquad (6.71)$$

where
$\overline{T}$ is the mean temperature (i.e., the mean of maximum and minimum for a sine wave)
A_z is the amplitude of the wave (i.e., the difference between maximum and minimum temperatures)
ω is $2\pi/\tau$, where τ is the period (e.g., 24 h)
$\phi(z)$ is the phase angle at depth z (difference in time between the occurrence of the maxima or minima at depth 0 and depth z, units of radiance)

Equation 6.67 can be solved using the above equation for $T(z, t)$ yielding

$$T(z,t) = \overline{T} + \frac{A_0 \sin[\omega t - (z/z_d)]}{e^{z/z_d}}, \qquad (6.72)$$

where
A_0 is the amplitude at depth zero
z_d is called the damping depth and is a function of the thermal conductivity and volumetric heat capacity

$$z_d = \left(\frac{2k_T}{C\omega}\right)^{1/2}. \qquad (6.73)$$

When z equals z_d, the amplitude is $1/e = 0.37$ of the amplitude at depth zero. Writing Equation 6.72 for two depths, z_1 and z_2, and solving for k_T/C gives

$$\frac{k_T}{C} = \frac{\pi(z_2 - z_1)^2}{\tau[2.3\log(A_1/A_2)]^2}. \qquad (6.74)$$

Figure 6.30 shows measured diurnal temperature waves for several depths in a Pullman clay loam on two different days. Table 6.7 gives the mean water content as measured by TDR for each depth. Also shown are the thermal conductivities obtained by fitting a sine series to each depth and evaluating the thermal conductivity of the layer between depths by using the coefficients of the sine series in a sine series solution to $T(t)$ at the next depth and then performing a nonlinear fit of that solution to obtain the thermal conductivity.

Single-probe heat pulse methods have been developed to measure thermal diffusivity; and a dual-probe heat pulse method (Campbell et al., 1991) can measure the thermal diffusivity, D_T, as well as k_T and C (Kluitenberg et al., 1995). Noborio et al. (1996) demonstrated a modified trifilar (three-rod) TDR probe that measured θ_v by TDR, and k_T by the dual probe heat pulse method. Their measured k_T compared well with values calculated from de Vries (1963) theory.

Soil heat flux is more commonly determined using heat flux plates (Table 6.1). These are thermopiles that measure the temperature gradient across the plate, and, knowing the conductivity of the plate, allow calculation of the heat flux from Equation 6.74. Heat flux plates are impermeable and block water movement. Because of this, the plates should be installed a minimum of 5 cm below the soil surface, so that the soil above the plate does not dry out or wet up appreciably more than the surrounding soil. Typical installation depths are 5 or 10 cm. Even at these shallow depths, the heat flux is greatly reduced from its value at

TABLE 6.7 Mean Soil Water Contents, θ (m³ m⁻³), and Thermal Conductivities, k_T (J m⁻¹ s⁻¹ K⁻¹), Calculated from a Nonlinear Fit of a Sine Series Solution Compared with k_T Estimated using Equation 6.74 for a Pullman Clay Loam for 2 Days in 1993

Depth (cm)	θ (m³ m⁻³)	Amplitude Based, Equation 6.74		Nonlinear Fit—Sine Series Soln.
		$k_{T \text{ at } \theta=\text{interlayer mean}}$	$k_{T \text{ at } \theta=0.23}$	$k_T\theta$
Day 308				
2	0.09			
		0.176	0.232	0.138
4	0.14			
		0.518	0.577	0.485
6	0.22			
		0.74	0.717	0.709
10	0.27			
		1.128	1.003	1.221
15	0.31			
		1.295	1.151	1.271
20	0.27			
Day 309				
2	0.08			
		0.082	0.111	0.117
4	0.13			
		2.273	2.566	0.363
6	0.22			
		3.85	3.772	0.596
10	0.26			
		3.008	2.7	0.916
15	0.31			
		2.006	1.784	1.098
20	0.27			

FIGURE 6.31 Four methods of correcting heat flux, measured with plates at 5 cm depth, to surface heat flux.

the soil surface; and corrections must be applied to compute surface heat flux. The most common correction involves measuring the temperature and water content of the soil at midlayer depths, z_j, in N layers (j to N) between the plate and surface and applying the combination equation over some time period, τ_t, defined by beginning and ending times t_i and t_{i+1}

$$G = G_z + \frac{\sum_{j=1}^{N}(T_{zj\_i+1} - T_{zj\_i})\Delta z_j C_{zj}}{(t_{i+1} - t_i)}, \quad (6.75)$$

where

    G is the surface heat flux during τ_t
    G_z is the flux at depth z
    T_{zj} are temperatures at the N depths, z_j, at times t_i and t_{i+1}
    Δz_j is the depth of the layer with midpoint z_j

where the volumetric heat capacities, C_{zj}, at depths z_j are calculated from Equation 6.70, rewritten as

$$C_{zj} = \frac{2.0 \times 10^6 \rho_{bzj}}{2.65} + 4.2 \times 10^6 \theta_{vzj} + 2.5 \times 10^6 f_{ozj}, \quad (6.76)$$

where θ_{vzj}, ρ_{bzj}, and f_{ozj} are the water contents, soil bulk densities, and volume fractions of organic matter, respectively, at depths z_j. The estimate of G is not much changed by the exact form of the combination equation as shown by data from Bushland, Texas, for four forms of Equation 6.75 (Figure 6.31). For situations where water content and temperature change rapidly with depth, or bulk density or f_o changes rapidly with depth, the multiple layer approach will work better.

As an example, we compare four methods of combining temperature and water content data to correct heat flux for bare soil at Bushland, Texas, in 1992 (Figure 6.31). Temperatures were measured at 2 and 4 cm depths (two replicates) with thermocouples, T_2 and T_4; at the surface with a single IRT, T_0; and as a mean temperature of the surface to 5 cm depth soil layer using thermocouples wired in series and buried (four replicates) at 1 and 4 cm depths, T_{1_4}. Water contents were measured by TDR probes (two replicates) inserted horizontally at 2 and 4 cm depths, θ_2 and θ_4. Soil heat flux at 5 cm depth, G_5, was measured with heat flux plates (four replicates). For all methods, the product of soil bulk density and heat capacity of soil solids were set to 1.125 MJ m⁻³. For the first method, the surface heat flux, G_0, was described as

$$G_0 = G_5 + \frac{(1.125 + 4.2\theta_w) \times 10^6 (0.05)(T_{w+1} - T_w)}{1800}, \quad (6.77)$$

where 1800 was the period in seconds; the weighted water content for the surface to 5 cm depth layer, θ_w, was

$$\theta_w = \frac{3\theta_2}{5} + \frac{2\theta_4}{5} \quad (6.78)$$

the weighted temperature for the surface to 5 cm depth layer, T_w, was

$$T_w = \frac{3T_2}{5} + \frac{2T_4}{5} \quad (6.79)$$

and T_{w+1} was calculated the same way, but for the previous measurement.

For the second method, θ_w and the series-wired thermocouple temperature were used:

$$G_0 = G_5 + \frac{(1.125 + 4.2\theta_w) \times 10^6 (0.05)(T_{1\_4+1} - T_{1\_4})}{1800}. \quad (6.80)$$

For the third method, θ_w was used:

$$G_0 = G_5 + \frac{(1.125 + 4.2\theta_w) \times 10^6 (0.05)(T_{024+1} - T_{024})}{1800}, \quad (6.81)$$

but the depth-weighted mean, T_{024}, of IRT temperature and the temperatures measured at 2 and 4 cm was used.

$$T_{024} = \frac{\left[\dfrac{(T_0 + T_2)/2 + T_0}{2}\right]}{5} + \frac{2T_2}{5} + \frac{2T_4}{5}. \quad (6.82)$$

For the fourth method, a modified layer approach was used:

$$G_0 = G_5 + \frac{(1.125 + 4.2\theta_2) \times 10^6 (0.05)(T_{02+1} - T_{02})}{1800}$$
$$+ \frac{(1.125 + 4.2\theta_4) \times 10^6 (0.05)(T_{4+1} - T_4)}{1800}. \quad (6.83)$$

where the depth-weighted mean temperature in the surface to 3 cm deep layer, T_{02}, was

$$T_{02} = \frac{\left[\dfrac{(T_2 + T_0)/2 + T_0}{2}\right]}{3} + \frac{2T_2}{3}. \quad (6.84)$$

All of these methods produced similar values of G_0, but those using a depth-weighted water content tended to overestimate extreme values, probably because the 2 cm water content was less than that at 4 cm (Figure 6.31). The weighted mean approach for both water content and temperature, with surface temperature included (Equation 6.81), produced generally the largest diurnal swing in G_0. Methods that did not include the surface temperature, but used the weighted mean approach for both water content and temperature (Equations 6.77 and 6.80), produced intermediate results. The layer approach (Equation 6.83), produced the smallest diurnal swing in G_0, despite using the surface temperature, and is probably the most accurate approach. All methods corrected both the amplitude and the phase of the diurnal cycle of G_0 appropriately.

Spatial variability of G can be considerable due to shading of soil by vegetation and due to micro-relief that both lead to variable soil wetness, thermal conductivity, and E_s rates.

For row crop investigations, at least four soil heat flux plates should be used to obtain a representative mean. Spatial and temporal variation of G will depend on row orientation, crop height and cover factor, and the time since wetting and the pattern of soil water uptake by roots. For sparse canopies, G measurement may require many more heat flux plates (Kustas et al., 2000).

6.2.3.1 Bernoulli, Soil Air, and Convective Heat Flux

In addition to diffusive heat flux, convective heat flux plays an important and often overlooked role in soil heating or cooling. This is the heat transported by moving air or water, the latter denoted in Figure 6.1 by G_{Jw} for heat transported by infiltrating water. Convective heat flux due to infiltration of water can be much larger than that due to diffusion on a diurnal basis. For example, irrigation with 5 cm of water at 15°C on a soil at 25°C with an initial water content of 0.1 m³ m⁻³ and a bulk density of 1.48 would immediately lower the temperature of the 11.6 cm deep, wetted layer to 20°C (assuming negligible heat of wetting, the soil brought to saturation, and a heat capacity of 1.54 MJ m⁻³ K⁻¹). Although soil temperature may be changed strongly by infiltration of water, there is another mechanism for soil temperature changes, the heat of wetting. The heat of wetting is heat energy released when a soil is wetted and is usually not large enough to be important in heat balance calculations. It can be large for clays with large surface area if they are extremely dry, ranging from 40 J g⁻¹ for kaolinitic clays to 125 J g⁻¹ for allophanic clays (Iwata et al., 1988). But, it decreases quickly as the initial water content of the soil increases and is not likely to be important for the normal range of field water contents.

Convective heat flux due to air movement into and out of soil surfaces is commonly ignored in energy balance considerations. Most people do not see any reason for air to move into or out of soil other than changing atmospheric air pressure. However, there are other forces at play that may significantly increase air flow into or out of soil surfaces. Consider Bernoulli's theorem,

$$\frac{v_1^2}{2g} + \frac{p_1}{\rho_a g} + z_1 = \frac{v_2^2}{2g} + \frac{p_2}{\rho_a g} + z_2, \quad (6.85)$$

which is an equation of conservation of energy where v_1 and v_2 are fluid (e.g., air) velocities at two points, p_1 and p_2 are the respective fluid pressures normal to the fluid flow direction, ρ_a is the density of air, g is the acceleration due to gravity, and z_1 and z_2 are the elevations of the two points. Placing point 1 in the soil and point 2 in the atmosphere directly above the soil, we see that $v_1 \approx 0$, and $v_2 \geq 0$, so that the equation may be rewritten as

$$p_1 - p_2 = \frac{\rho_a v_2^2}{2} + (z_2 - z_1)\rho_a g. \quad (6.86)$$

That is, the pressure differential from soil to air is equal only to the elevation difference multiplied by $\rho_a g$ when wind speed is zero, and it increases above that value as the square of wind speed. During sustained winds across flat surfaces, there is a

sustained pressure gradient driving air movement from soil to air. For air movement over nonplanar surfaces, the situation is complicated by other aerodynamic effects such as drag, which may increase or decrease the pressure gradient at various places over the surface. This has implications for soil gas efflux rates and measurements (Rudolf et al., 1996). Turbulent air flow across soil was investigated by Farrell et al. (1966), who showed with field measurements and theoretical analysis that the effective gaseous diffusion coefficient in soil could be increased by 100 times for wind speed of $7\,m\,s^{-1}$. Bulk air movement into and out of soils can substantially alter the energy balance as evidenced by such phenomena as ice caves in permeable lava rocks overlain by soils.

6.2.4 Sensible Heat Flux

Sensible heat flux is the transfer of heat energy away from or to the surface by conduction or convection. Because air is a poor conductor of heat, most sensible heat flux is by convection (movement) of air. This occurs in eddies of different scales depending on the turbulence of the atmosphere near the surface. Turbulence is influenced by the aerodynamic roughness of the surface, the wind speed, and the temperature differential between the surface and the air. Perhaps the most common method of evaluating sensible heat flux is to measure the other terms in Equation 6.1 as accurately as possible and then set H equal to the residual

$$H = -R_n - G - LE. \qquad (6.87)$$

Of course, this approach lumps all the errors in the other terms into H. More importantly, it does not allow for a check on the accuracy of the energy balance. By definition, if H is defined by Equation 6.87 then Equation 6.1 will sum to zero. Only an independent measure of H can provide a check sum for Equation 6.1. As noted in Section 6.2.2.2, EC is a direct method of measuring H.

The Bowen ratio method can be applied to sensible heat flux as well as to latent heat flux as outlined in Section 6.2.2.3. Sensible heat flux is calculated from the Bowen ratio as (following Rosenberg et al., 1983, p. 256)

$$H = \frac{-(R_n + G)}{(1 + (1/\beta_r))}. \qquad (6.88)$$

The considerations of fetch, measurement height, equipment, etc., mentioned in Section 6.2.2.4 for Bowen ratio and EC measurements apply as well to sensible heat flux measurements made with these methods.

Though obviously a dynamic and complex process, sensible heat flux, H ($W\,m^{-2}$), is sometimes estimated using a straightforward resistance equation

$$H = \frac{\rho_a c_p (T_z - T_0)}{r_{aH}}, \qquad (6.89)$$

where

ρ_a is the density of air ($\rho_a = 1.291 - 0.00418 T_a$, with less than $0.005\,kg\,m^{-3}$ error in the $-5°C$ to $40°C$ range, T_a in °C)

c_p is the heat capacity of air ($1.013 \times 10^3\,J\,kg^{-1}\,K^{-1}$)

T_z is the air temperature at measurement height

r_{aH} is the aerodynamic resistance to sensible heat flux ($s\,m^{-1}$)

T_0 is the temperature of the surface

For vegetation, the "surface" for aerodynamic resistance is the height at which the logarithmic wind speed profile, established by measurements of wind speed above the surface, extrapolates to zero. This height is $d + z_{om}$ and is often well below the top of the canopy, typically at 2/3 to $3/4 h_c$. Measurements of surface temperature (with, for instance, an IRT) may not be the mean temperature at the same height as the aerodynamic "surface," thus introducing bias in the r_{aH} estimation. Also, the roughness length for momentum, z_{om}, may be different from that for sensible heat, z_{oH}.

A general form for aerodynamic resistance to sensible, latent energy, and momentum fluxes is

$$r_a = \frac{1}{k^2 u_z} \left[\ln\left(\frac{z_m - d}{z_o}\right) \right]^2, \qquad (6.90)$$

where

k is the von Kármán constant = 0.41

z_o is the roughness length (m)

z_m is the reference measurement height (m)

u_z is the wind speed ($m\,s^{-1}$) at that height

d is the zero plane displacement height (m)

Equation 6.90 only holds for neutral stability conditions. Unstable conditions occur when the temperature (and thus air density) gradient from the surface upward is such that there is warm air rising through the atmosphere (forced convection). Stable conditions prevail when the air is much cooler and denser near the surface, thus inhibiting turbulent mixing due to forced convection. Neutral conditions obtain when neither stable nor unstable conditions do.

For bare soil, Kreith and Sellers (1975) simplified Equation 6.90 as follows:

$$r_a = \frac{1}{k^2 u_z} \left[\ln\left(\frac{z_m}{z_o}\right) \right]^2. \qquad (6.91)$$

They found a value of $z_o = 0.003\,m$ worked well for smooth bare soil.

For nonneutral conditions, a variety of stability corrections have been proposed. See Rosenberg et al. (1983, pp. 140–144) and Monteith and Unsworth (1990, pp. 234–238). Because many models of the soil–plant–atmosphere continuum use Equation 6.89 to model H, it is important to note that, while stability corrections can improve model predictions of H and surface

temperature, the stability corrections are implicit in terms of H. This leads to a requirement for iterative solution of sensible heat flux at each time step in these models.

Knowledge of appropriate values for d and z_o in the above equations can be hard to come by since their determination involves measurements of wind speed and temperature at multiple heights within and above the canopy (Monteith and Unsworth, 1990). Campbell (1977) suggests estimating these from plant height, h_c, as

$$d = 0.64h_c \qquad (6.92)$$

for densely planted agricultural crops and

$$z_{om} = 0.13h_c \qquad (6.93)$$

for the roughness length for momentum for the same condition. Campbell (1977) gives the roughness length parameters for sensible heat, z_{oH}, and vapor transport, z_{ov}, as

$$z_{oH} = z_{ov} = 0.2z_{om}. \qquad (6.94)$$

Note that Equation 6.94 differs from Equation 6.58, where Jensen et al. (1990) used $z_{oH} = 0.1\,z_{om}$. For coniferous forest, Jones (1992) gives

$$d = 0.78h_c \qquad (6.95)$$

and

$$z_{om} = 0.075h_c \qquad (6.96)$$

for these parameters. As wind speeds increase, many plants change form and height, with resulting decrease in h, d, and z_{om}. It is unlikely that the relationships given in Equations 6.92 through 6.96 hold true for large wind speeds.

6.3 Water Balance Equation

The water balance equation is written for a control volume of unit surface area and with a vertical dimension that extends from the soil surface to a lower boundary that is commonly assigned a depth below the bottom of the root zone (Figure 6.1) but may be much deeper:

$$0 = \Delta S - P + R - F - \frac{LE}{(\lambda \rho_w)}, \qquad (6.97)$$

where
   ΔS is the change in soil water storage in the profile (taken as positive when water storage increases over time)
   P is precipitation and irrigation
   R is the sum of runoff and runon
   F is flux across the lower boundary of the profile
   $LE/(\lambda \rho_w)$ is water lost to the atmosphere through evaporation from the soil or plant or gained by dew formation

Units are typically millimeter depth of water per unit time, equivalent to volume of water per unit area per unit time. The value of P is always positive or zero, but values of ΔS, R, F, and $LE/(\lambda \rho_w)$ may have either sign. By convention, R is taken to be positive when there is more runoff than runon. As in Equation 6.1, $LE/(\lambda \rho_w)$ is positive when flux is toward the surface of the soil. The equation is often rearranged to provide values of $LE/(\lambda \rho_w)$ when suitable measurements or estimates of the other terms are available; but it can and has been used to estimate runoff, soil water available for plants, and deep percolation losses (flux downward out of the control volume). Here, we take F as positive when flux is upward across the lower boundary into the control volume. The term F is used rather than P for deep percolation, both to avoid confusion with precipitation, and to avoid the common misconception that deep flux is only downward since P is often used to indicate deep percolation.

In the case of $LE/(\lambda \rho_w)$, the units of LE (W m^{-2}) are conveniently converted to depth of water in mm s^{-1} by dividing by the latent heat of vaporization (λ, ~2.45 × 10^6 J kg^{-1}) and by the density of water (ρ_w, ~1000 kg m^{-3}), with little loss of accuracy because the changes in density of water and the latent heat of vaporization with temperature are small. For precise work, both parameters can be temperature corrected. The change in storage (ΔS) is often determined by measuring soil water content changes by methods that give volumetric water content, θ_v (m^3 m^{-3}). Multiplying the water content by the depth of the layer gives the depth of water stored, and the first derivative of storage over time is ΔS. In the United States, the term ET is used to represent the sum of evaporative fluxes from the soil and plant. By convention, ET is taken as positive for fluxes from plant or soil surface to the atmosphere. Thus, ET = $-LE/(\lambda \rho_w)$ and the water balance may be rearranged as

$$\Delta S = P - R + F - \text{ET}. \qquad (6.98)$$

This provides a use for the ET term for those who prefer to say evaporation rather than ET. Examination of Equation 6.98 will satisfy the reader that soil water storage increases with precipitation, decreases if runoff from precipitation occurs, decreases with increasing ET, and increases with flux upward into the control volume. Although Equations 6.97 and 6.98 are instantaneous rate equations, they are often written for an assumed period of time such that the variables are in units of depth of water for that time period.

6.3.1 Measuring ΔS and ET

Probably the most accurate method of measuring ΔS is the weighing lysimeter (Wright, 1991). Although large weighing lysimeters involve considerable expense, they can give very precise measurements (0.05 mm = 0.05 kg m^{-2}) (Howell et al., 1995). An excellent review of the use of weighing lysimeters is given by Howell et al. (1991). Careful design, installation, and operation will overcome any of the serious problems reported with some lysimeters including disturbance of the soil profile (less with monolithic lysimeters),
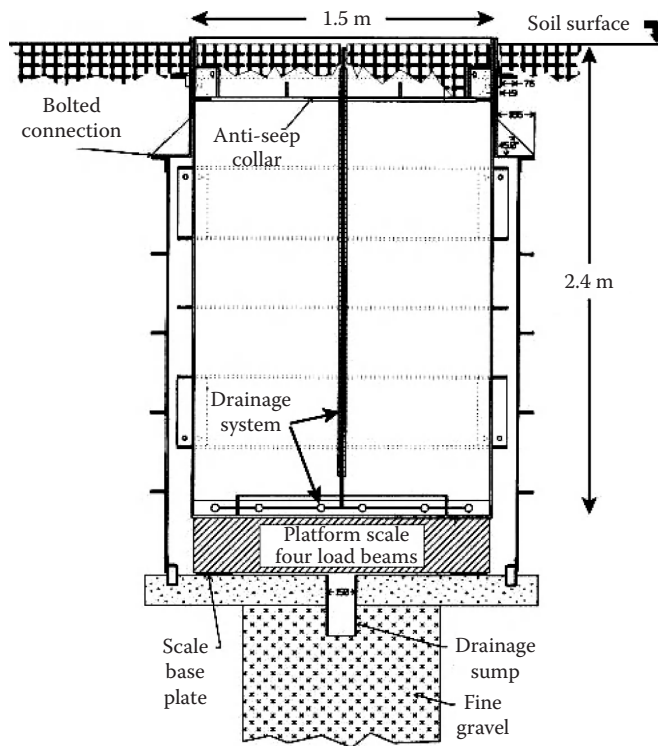
FIGURE 6.32 Cross-sectional view of simplified weighing lysimeter installed for grass reference ET measurements at Bushland, Texas. (From Schneider, A.D., T.A. Howell, A.T.A. Moustafa, S.R. Evett, and W.S. Abou-Zeid. 1998. A simplified weighing lysimeter for monolithic or reconstructed soils. Appl. Eng. Agr. 14:267–273.)

interruption of deep percolation and horizontal flow components, uneven management of lysimeter compared with field soil (Grebet and Cuenca, 1991), and other sources of bias (Ritchie et al., 1996). Other drawbacks include heat flux distortions caused by greatly conductive steel walls (Black et al., 1968; Dugas and Bland, 1991), but minimal for large lysimeters, and high cost, for example, $65,000 (Lourence and Moore, 1991), $80,000 (Marek et al., 1988), and $65,000 (Evett et al., 2009a). All lysimeters have walls, one that retains the soil and is part of the soil container, and another that separates the soil container from the surrounding field soil. A gap between the walls is necessary to allow weighing. The wall-gap surface area is the sum of inside and outside wall horizontal surface areas plus the gap horizontal surface area. Well-designed lysimeters may have a wall-gap surface area as little as 1.5% of the lysimeter soil surface area (Howell and Tolk, 1990).

Schneider et al. (1996) described simplified monolithic weighing lysimeters (Figure 6.32) that were considerably less expensive than, and nearly as accurate as, the monolithic weighing lysimeter described by Evett et al. (2009a) (Figure 6.33). These two designs represent contrasts in mode of operation and each presents some advantages and disadvantages. The design in Figure 6.33 allows access to all sides and the bottom of the lysimeter for installation or repair of sensors and weighing or drainage systems. The data logger that handles all measurements is installed in the underground chamber and typically is subject to only a small diurnal

temperature swing of 1°C, reducing temperature-induced errors in low-level measurements such as load cell transducer bridges and thermocouples. Other equipment that may be installed in the chamber includes a system for TDR measurements of soil water content and bulk electrical conductivity and concurrent measurements of soil temperature, and an automatic vacuum drainage system that continuously monitors drainage rate. The drainage tanks were suspended by load cells from the bottom of the lysimeter tank, allowing measurement of tank mass change without changing the mass of the lysimeter until the tanks are drained (manual but infrequent). The design illustrated in Figure 6.33 overcomes the main disadvantages of the design of Marek et al. (1988), which were the shallow soil depth over the ceiling around the periphery of the chamber and the nearness of the entrance hatch to the lysimeter surface. The shallow soil depth can cause uneven plant growth next to the lysimeter, but this problem can be eliminated with the installation of a drip irrigation system to apply additional water to this area. Access to the lysimeter must be carefully managed to avoid damage to the crop.

Figure 6.32 shows a weighing lysimeter design that minimizes disturbance to the field during both installation and operation. The monolith was collected a short distance away, and the outer box was installed in a square hole that disturbed only a 15 cm perimeter of soil outside the lysimeter. Because there is no access to the sides or bottom of the lysimeter, there is no reason for personnel to visit the lysimeter area except for crop management and the occasional manual drainage accomplished with a vacuum pump and collection bottle. Disadvantages of this design are considerable, and include the lack of continuous drainage data, the fact that manual drainage may not be timely or complete, and the fact that the weighing system is impossible to repair without disassembling the entire lysimeter, which can be costly and time consuming (sometimes months; e.g., Schneider et al., 1998; Marek et al., 2006). This lower cost, direct weighing lysimeter built with no access below the surface at Bushland, Texas (Howell et al., 2000), experienced a weighing system failure due to lightning that required removal of the soil container and scale, which took many weeks, was costly, and disrupted an entire measurement season. A scale failure in a similar lysimeter at Uvalde, Texas, caused Marek et al. (2006) to point out that this "closed" type of design has the potential to result in costly repairs and lengthy loss of data when scales fail. In addition, the remote location of the data logger for the "closed" design can cause problems of electrical interference or signal degradation. The long cable lengths to the external data logger used with the lysimeter in Figure 6.32, and the large diurnal temperature swing to which cables and data logger were exposed, both caused a six-wire bridge to be needed to reduce errors due to temperature-induced resistance changes when reading the platform scale load cells. By contrast, the location of the data logger inside the lysimeter chamber in Figure 6.33 results in diurnal temperature variations being <1°C, allowing a four-wire bridge to be used for reading the weighing system load cell. Measurement precision with the lever beam scale in Figure 6.33 is 0.06 mm, while that with the platform scale in Figure 6.32 is 0.1 mm.
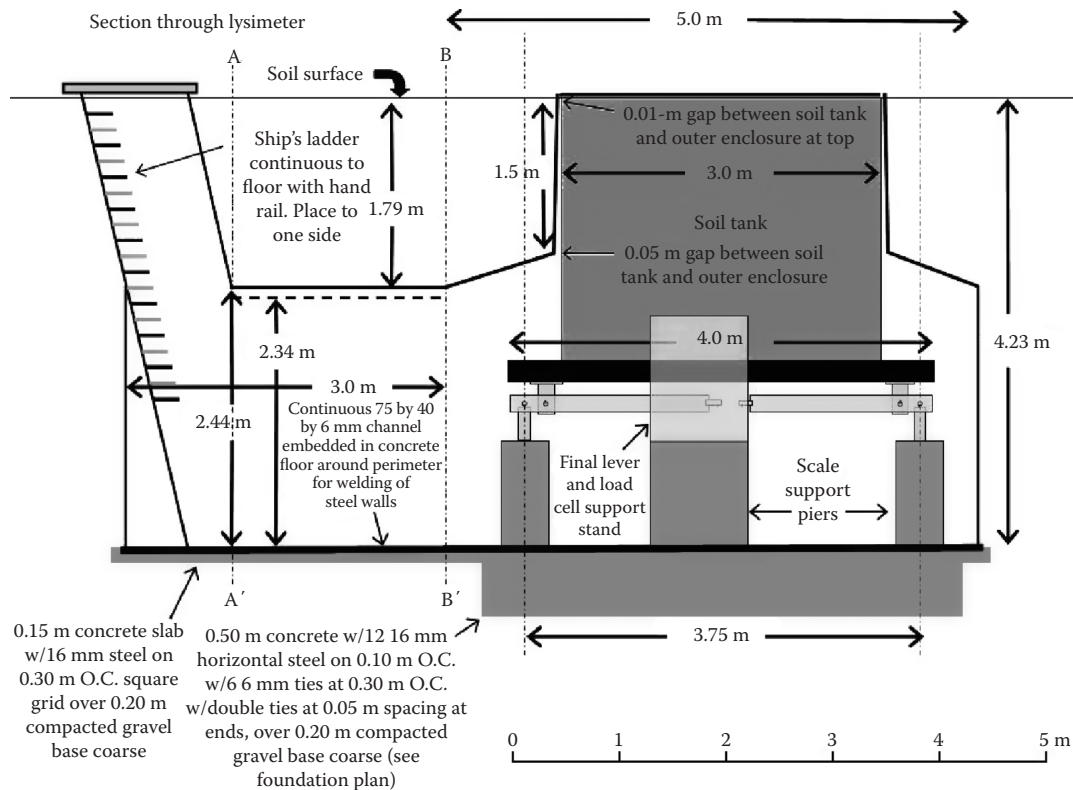
FIGURE 6.33 Cross-sectional view of the large weighing lysimeter at Dayr Alla in the Jordan Valley, Jordan. (From Evett, S.R., N.T. Mazahrih, M.A. Jitan, M.H. Sawalha, P.D. Colaizzi, and J.E. Ayars. 2009a. A weighing lysimeter for crop water use determination in the Jordan Valley, Jordan. Trans. ASABE 52:155–169.)

Weighing lysimeters measure mass change over a given time. If mass is measured in kilograms, then dividing the mass change by the surface area in meter squares of the lysimeter will give the change in water storage, ΔS, of the lysimeter as an equivalent depth of water in millimeter, with only slight inaccuracy due to the density of water not being quite equal to $1\,Mg\,m^{-3}$. If evaporation from bare soil is being measured, then the surface area of the soil within the lysimeter should be used in the calculation. However, if a closed crop canopy is present, then the surface area to be used should be representative of the canopy surface that is supported by the lysimeter. This may be equal to the sum of the lysimeter soil surface area plus one half of the wall-gap surface area, depending on the lysimeter design and crop planting pattern.

If only daily ET values are needed, then ΔS is computed from the 24 h change in lysimeter mass, usually midnight to midnight. Some averaging of readings around midnight may be needed to smooth out noise. The daily ET is computed by summing ΔS with any precipitation or drainage. Data from a continuously weighing lysimeter may be presented as a time sequence of mass (or depth of water storage) referenced to an arbitrary zero (Figure 6.34). Irrigation or precipitation events often will appear as obvious increases in storage (Figure 6.34), and drainage events will show up as decreases in storage. Adjusting the sequential record of storage amount by subtracting the rainfall or irrigation depth, and adding the depth of drainage water, at the time that these occurred, will remove these changes in storage, and is equivalent

to the operations defined by the $+P$ and $+F$ in Equation 6.98; resulting in the monotonically decreasing storage shown in Figure 6.35. Taking the first derivative of the adjusted storage with respect to time gives the adjusted ΔS rate, and thus ET rate (Figure 6.36) (R assumed to be zero). In order to compute ET rates on the same time interval as lysimeter mass measurements are made, we must have concurrent measurements of irrigation, precipitation, and drainage on the same or a finer recording interval.

Weighing lysimeters are subject to wind loading, more so when the soil surface is bare, as evidenced by Figure 6.37. In windy
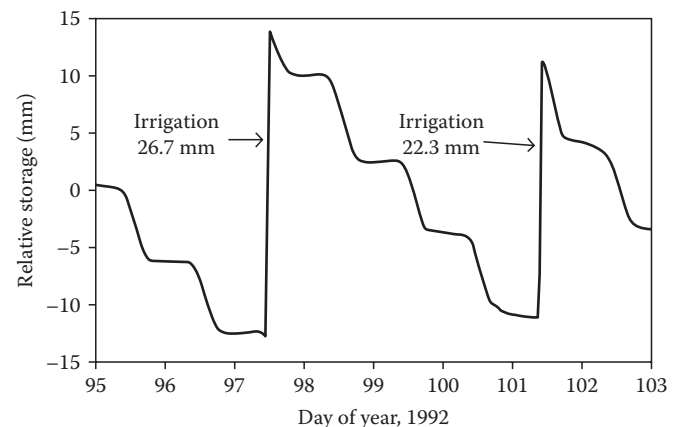
FIGURE 6.34 Unadjusted weighing lysimeter storage for winter wheat at Bushland, Texas, showing storage increases due to two irrigations.
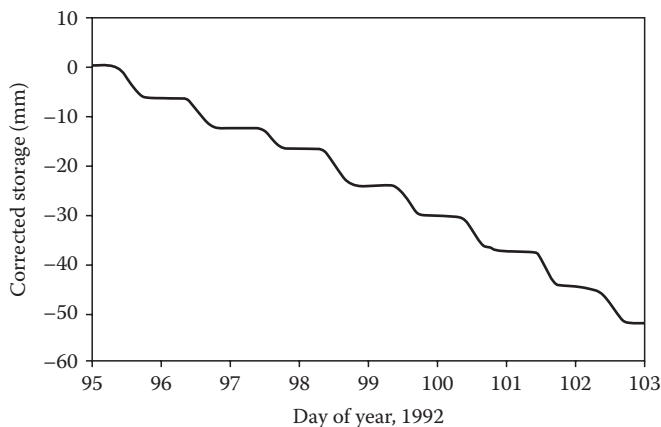
FIGURE 6.35 Lysimeter storage values from Figure 6.34 adjusted by subtracting irrigation amounts.
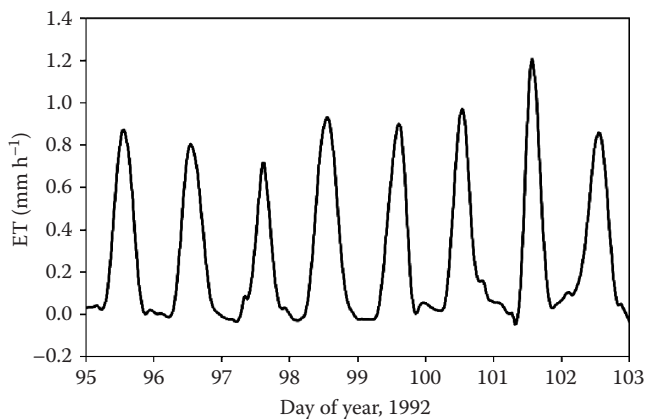
FIGURE 6.36 ET rate calculated by taking the negative of the first derivative of adjusted storage from Figure 6.35. The negative ET rates shown for some nights are caused by dew formation.
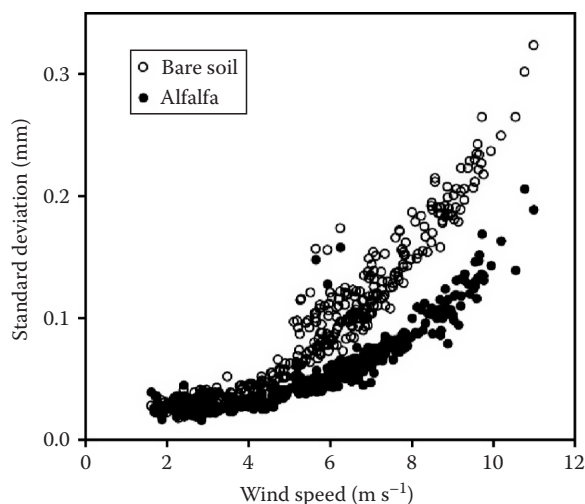
FIGURE 6.37 Half-hourly standard deviations of lysimeter storage (mm) as affected by wind speed over contrasting surfaces for days 97–105 at Bushland, Texas, in 1994.

regions, it may be necessary to smooth the data to remove noise when calculating the ET rate. Gorry (1990), following Savitsky and Golay (1964), described a method for general least squares smoothing that allowed application of different levels of smoothing to both the raw data and their first derivative. Application of this method to post-processing of data is preferable to real-time smoothing that may eliminate detail in the data. With post-processing, we can apply only the amount of smoothing needed to reduce noise to acceptable levels. A computer program to apply Savitsky–Golay smoothing is available (http://www.cprl.ars.usda.gov/programs/).

Microlysimeters are lysimeters small enough to be installed and removed by hand for weighing daily or more often. They have been used to measure the water use of short grass, but lack the soil depth needed for soil water redistribution and drainage that would reproduce the soil profile water content in the adjacent field over the long term. They can give good precision but are sensitive to spatial variability of soil water characteristics, soil profile water content distribution, and surface condition. Lascano and Hatfield (1992) showed that 182 microlysimeters were required to measure field average E with precision of 0.1 mm day^{-1} at a 90% confidence level when their soil was wet but only 39 when dry. This was due to the greater variability of E for wet soil. For a precision of 0.5 mm day^{-1}, only seven microlysimeters were required for any soil wetness. To avoid heat conduction to and from the surface, microlysimeter walls should be made of low thermal conductivity materials such as plastic; and, to avoid trapping heat at the bottom of the microlysimeter, the bottom end cap should be made of a thermally conductive material such as metal (Evett et al., 1995). Plastic pipe makes good microlysimeters. Typical dimensions are 7.6 or 10 cm in diameter and from 10 to 40 cm in height. Beveling the bottom end eases insertion into the soil. Typical practice is to insert the microlysimeter vertically until its top is level with the soil surface; then dig it out, or rotate it to shear the soil at the bottom, and pull it out. After capping the bottom with a water-tight seal, it is weighed before reinsertion into the original or a new hole; sometimes lined with a material (e.g., plastic sheet or bag) to prevent sticking of the soil to the microlysimeter outside surface. After a period of time, the microlysimeter is reweighed and the difference in initial and final weights is the evaporative loss. Short microlysimeters should be replaced daily, as the water supply is soon used up to the point that the soil inside the lysimeter is no longer at the same water content as the soil outside, biasing the measured E. In a study of spatial variability of evaporation from bare soil, Evett et al. (1995) used 30 cm high microlysimeters to avoid daily replacement so that the spatial relationship would not be changed. They showed that for their clay loam soil the 30 cm height was adequate for 9 days. If plant roots are present, it is recommended to replace microlysimeters daily to lessen errors associated with root water uptake that occurs elsewhere in the field but not in the microlysimeters.

Alternatives to weighing lysimetry include soil water measurement methods for assessing ΔS for a soil profile of given depth over a given time. In this case, the soil volume of interest is

unbounded below the surface and F is, strictly speaking, uncontrolled. Measurements of soil water content can give the change in soil water stored in a profile of given depth with good accuracy; and can give good values for E if water flux across the bottom of the profile is known or can be closely estimated. Baker and Spaans (1994) described a microlysimeter with TDR probe installed vertically to measure the water content. Comparison of E calculated from the change in storage determined with TDR closely matched the E measured by weighing the microlysimeter. Young et al. (1997) showed that a single 800 mm long probe installed vertically from the surface could account for 96% of ET from weighing lysimeters irrigated on a 6 day interval; but standard error for the probe was about four times larger than that for the lysimeter (0.46 and 0.07 mm, respectively). In a container study with a sorghum plant, Wraith and Baker (1991) showed that a TDR system could measure ET with good resolution and provide measurements of change in storage on a 15 min interval that compared very well with those measured by an electronic scale. Despite the good laboratory performance of TDR for soil E and ET determinations, outdoor installations are limited by the depth to which TDR probes can be installed without disturbing the soil profile unduly, and the diurnal and seasonal temperature effects on TDR derived water contents, which can bias E and ET determinations.

Evett et al. (1993) showed that change in storage in the upper 35 cm of the profile under winter wheat could be accurately tracked with horizontally placed TDR probes, with an average of 88% of daily ΔS occurring in the upper 30 cm. But, E estimates were incorrect (compared with a weighing lysimeter) when flux across the 35 cm lower boundary occurred. However, combination of the TDR system with neutron moisture meter (NMM) measurements of deeper soil moisture allowed measurement of E to within 0.7 mm of lysimeter measured ET over a 16 day period; five times better than the accuracy achieved using only NMM measurements.

Figure 6.38 shows the soil water storage (referenced to arbitrary zero) as measured for winter wheat by weighing lysimeter and two TDR arrays. Each TDR array consisted of seven probes inserted horizontally into the side of a pit and the pit backfilled after wheat planting. Probe depths were 2, 4, 6, 10, 15, 20, and 30 cm, and the probes were read every half hour. Rains on days 101, 104, and 106 can be seen as increases in the storage amount. Changes in storage as measured by the two systems were nearly identical in the 7 day period shown (Figure 6.39); and ET amounts were closely similar (Figure 6.40).

Water balance measurement intervals commonly range between hours and weeks and are usually no smaller than the required period of ET measurement. Measurement of each variable in Equation 6.97 presents its own unique problems. These include measurement errors in determination of lysimeter mass or ΔS, and errors in P and R measurements. Problems of P and R measurement are essentially identical for either weighing lysimetry or soil profile water content methods because the surface area of the control volume can be defined for both methods with a water-tight border, often consisting of a sheet metal square
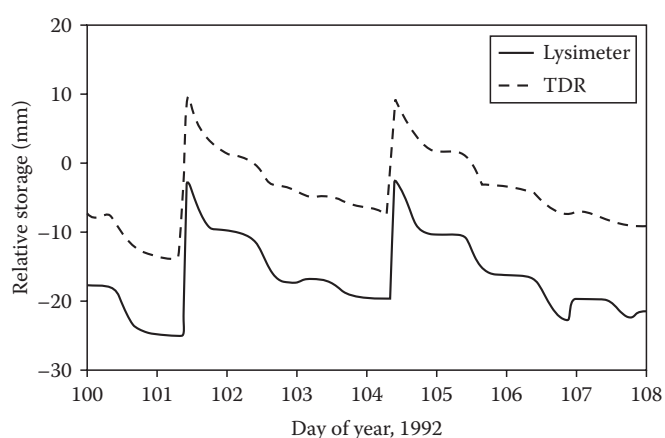
FIGURE 6.38 Soil water storage in the upper 35 cm of the soil profile, as measured by two TDR probe arrays, compared with storage in a 2.4 m deep weighing lysimeter. Zero reference is arbitrary. Winter wheat, Bushland, Texas, 1992.
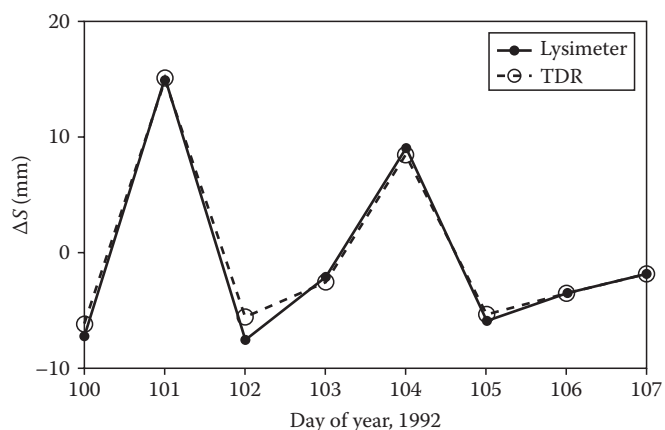
FIGURE 6.39 Daily change in storage for the 35 cm and 2.4 m profiles from Figure 6.38.
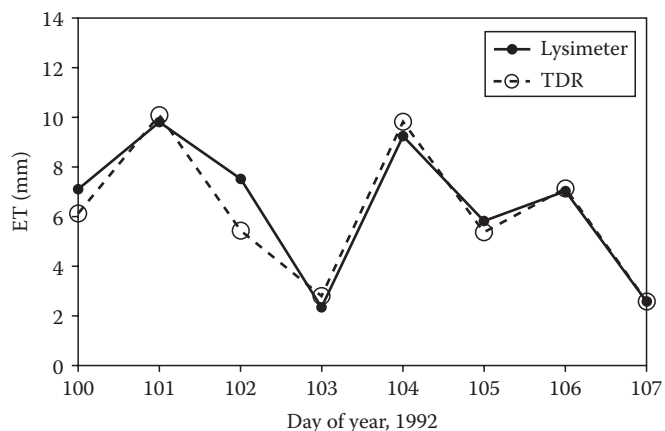
FIGURE 6.40 Daily ET as calculated from the TDR and weighing lysimeter data shown in Figure 6.38 and assuming $F = 0$.

or rectangle pressed into or partially buried in the soil surface. When the soil volume is unbounded below the surface, as in the soil profile water content method, there are additional errors due to uncontrolled horizontal flow components and deep percolation that are difficult to measure or estimate. Nevertheless, the profile water balance technique is applicable in many situations for which lysimetry is inappropriate or impossible and is, in addition, much less expensive. In many cases, the horizontal flow components may be assumed to sum to zero; and deep percolation may be nil if the soil profile water content measurements are made to sufficient depth (Wright, 1990).

When the NMM alone is used to measure soil water content, the soil water balance method is suitable for periods of several days or more if closure ($F \approx 0$) at the bottom of the measured soil profile can be obtained (Wright, 1990). The NMM has been commonly used (Cuenca, 1988; Wright, 1990) but due to the small changes in water content associated with daily ET, this water balance method has usually been restricted to determination of ET over several-day periods (Carrijo and Cuenca, 1992). Evett et al. (1993) showed that TDR measurements of soil water content near the surface could be coupled with deeper water content measurements by NMM to close the water balance considerably more precisely than with the NMM alone, opening up the prospect for daily ET measurements by this method.

Although several alternatives to the NMM have been introduced, all depending on sensing a change in the soil's electrical properties due to water content changes, none of these has proven useful for ET determinations due to interference from soil temperature, bulk electrical conductivity, and bound water effects on the soil dielectric permittivity, which are particularly important in superactive clay soils but present in nearly all soils (Evett et al., 2008), the limited measurement depth of some of the available sensors (Mazahrih et al., 2008), and the problems with small-scale spatial variability and nonrepresentativeness of readings uncovered by Evett et al. (2009b). The latter authors found that, in the field, the dielectric sensors used in access tubes were less precise than either the NMM or gravimetric sampling. Mazahrih et al. (2008) pointed out several studies in which soil water uptake was substantial to depths >3 m for crops such as corn, cotton, sugar beet, and winter wheat. A soil water sensing system for ET determination must be capable of determining the water content of the entire soil profile from the surface to well below the zone of root water uptake if the control volume is to be closed and ET estimates assured without errors due to deep F.

Surface storage has been ignored in the discussion of soil water balance here, but it not only has a large influence on the amount of water infiltrated from a precipitation or irrigation event, but it is also one of two factors that are the most amenable to human control both on a small and large scale. The other is infiltration rate, a soil property that may be increased by soil tillage over the short term, and decreased by tillage in the long term, and also may be influenced by no-tillage or minimum-tillage practices, by tree plantings and ground covers designed to protect the soil surface from slaking and crusting, and many other practices. Many farming efforts are aimed at increasing surface storage through practices like plowing to roughen the soil surface or furrow diking to create infiltration basins and decrease runoff, to name two. Soil conservation measures such as terraces, contour bunds, etc., are aimed at decreasing runoff or runoff velocity, and thus increasing the opportunity time for infiltration. On the other hand, water harvesting is a practice of increasing runoff by reducing both surface storage and infiltration rate, with the aim of using the runoff water elsewhere. Among the oldest artifacts of human cultivation are water harvesting systems that included the removal of gravel from desert surfaces to improve runoff, terrace systems for guiding and reducing runoff, and in some cases, the combined use of both technologies to concentrate water on an area chosen for cultivation (Shanan and Tadmor, 1979). To the extent that such practices influence infiltration, they will affect the timing, magnitude, and spatial patterns of such energy balance terms as sensible and latent heat fluxes, and the convective heat flux accompanying infiltrating water.

6.3.2 Estimating Flux across the Lower Boundary

One of the great advantages of lysimeters is that they control the soil water flux, F, into and out of the control volume. To date, a reliable soil water flux meter has not been developed, so F must be estimated if it is not controlled. If water flux across the lower boundary of the control volume is vertical, it may sometimes be estimated by measurements (preferably multiple) of soil water potential, h, at different depths separated by distance, Δz, and knowledge of the dependence of hydraulic conductivity, K (m s^{-1}), on soil water potential, the $K(h)$ curve. The potential difference, Δh, coupled with the unit hydraulic gradient for vertical flux, gives the hydraulic head difference, ΔH, driving soil water flux. Averaging the measurements allows estimation of the mean hydraulic conductivity for the soil layer between the measurement depths from the $K(h)$ curve, and thus estimation of the soil water flux, J_w (m s^{-1}), from a finite difference form of Darcy's law

$$J_w = -K\left(\frac{\Delta H}{\Delta z}\right). \tag{6.99}$$

Soil water potential may be measured by tensiometer or other means described in Chapter 3 (or see van Genuchten et al., 1991). Methods of measuring or estimating the $K(h)$ curve may be found in Chapter 4. For fluxes across boundaries too deep for the installation of tensiometers, the soil water content may be measured at two or more depths by the NMM (see Chapter 3) and the soil water potential inferred by inverting the $\theta_v(h)$ relationship, which may be estimated or measured (see Chapter 3 or van Genuchten et al., 1991). Due to the hysteresis of the $\theta_v(h)$ relationship, there is more room for error when basing J_w estimates on θ_v measurements. But, for many cases, the soil water potential

will be in the range where hysteresis is not a large source of error (drier soils), and hydraulic conductivity is not large either. Thus, both the value of J_w and the error in J_w may be small enough for practical use in determinations of ET.

6.3.3 Precipitation and Runoff

An in-depth discussion of precipitation and runoff measurement and modeling is beyond the scope of this chapter. A classic and still valuable reference on field hydrologic measurements is the *Field Manual for Research in Agricultural Hydrology* (Brakensiek et al., 1979). A more up-to-date and extensive reference is the *ASCE Hydrology Handbook*, 2nd edition (ASCE, 1996). Flow measurement in channels is detailed in *Flow Measuring and Regulating Flumes* (Bos et al., 1983). The monograph *Hydrologic Modeling of Small Watersheds* (Hann et al., 1982) included useful chapters on stochastic modeling, precipitation and snowmelt modeling, runoff modeling, etc., and listed some 75 hydrologic models available at that time. For soil water balance measurements, runoff is often controlled with plot borders or edging driven into the ground or included as the above-ground extension of a lysimeter. Steel borders driven into the soil to a depth of 20 cm will suffice in many situations. Sixteen gauge-galvanized steel in rolls 30 cm wide is useful for this and can be reinforced by rolling over one edge. If runoff must be measured, this can be done with flumes such as the H-flume and recording station shown in Figure 6.41.

Precipitation varies so much from location to location that it is rarely useful to attempt estimating it. Measurement methods include standard U.S. Weather Bureau rain gauges read manually, various tipping bucket rain gauges, heated gauges to capture snow fall (e.g., Qualimetrics model 6021, Sacramento, CA), snow depth stations, etc. A rain gauge should be located as close to the surface as possible and be surrounded by a wind shield to avoid catch loss associated with wind flow over the gauge (Figure 6.42). A standard for the capture area or throat of a rain gauge is that it should be 20 cm in diameter because smaller throats lead to more variability in amount captured. Various designs of tipping bucket rain gauge have become

FIGURE 6.41 H-flume and recorder (in white box) for measuring runoff rate from a graded bench terrace at Bushland, Texas. Note dike diverting flow from uphill flume.

FIGURE 6.42 Wind shield installed around a heated, tipping bucket rain gauge.

standard equipment on field weather stations. These are capable of providing precipitation data needed to solve the soil water balance for short intervals. Two problems are sometimes associated with tipping bucket type gauges. First, most of these devices count the tips using a Hall-effect sensor for detecting the magnetic field of a magnet attached to the tipping bucket; and the sensing system is sometimes susceptible to interference from sources of electromagnetic noise such as vehicle ignition systems, generators, lightning strikes, etc. Second, tipping bucket gauges do not keep up with very large rainfall rates. In regions with predominantly small rainfall events, a third error due to the water left in the tipping bucket at the end of precipitation may result in significant accumulative errors in seasonal precipitation. At Bushland, Texas, we have observed tipping bucket errors of 10%–15% for totals of rainfall from high intensity convective thunderstorms compared with amounts collected in standard rain gauges and sensed by weighing lysimeters. If accuracy is very important, then a tipping bucket gauge should be supplemented with a standard gauge that captures and stores all the rainfall. For solving the soil water balance, experience shows that the rain gauge(s) should be placed directly adjacent to the location of ΔS measurement. Separation of even 100 m can lead to large errors due to the spatial variability of precipitation.

For studies and operations at scales larger than small field size, there are now precipitation estimates from Doppler radar-based systems that offer calibrated rainfall data on a 24 h basis (Figure 6.43; see also Legates et al., 1996; Vieux and Farajalla, 1996). Although Figure 6.43 shows large grid sizes and only 16 levels of precipitation, grid sizes of 4 km on a side, with 256 levels of rainfall, are available. Data for these maps are generated by the WSR-88D radar system, usually known as the NEXRAD weather radar system, in widespread use in the United States. The Center for Computational Geosciences at the University of Oklahoma has developed a radar-based precipitation interface (RPI) for the radar data to generate the maps. Radar data are used from two or more stations and calibrated against rain gauge measurements available from, for example, the Oklahoma MESONET system of weather stations.
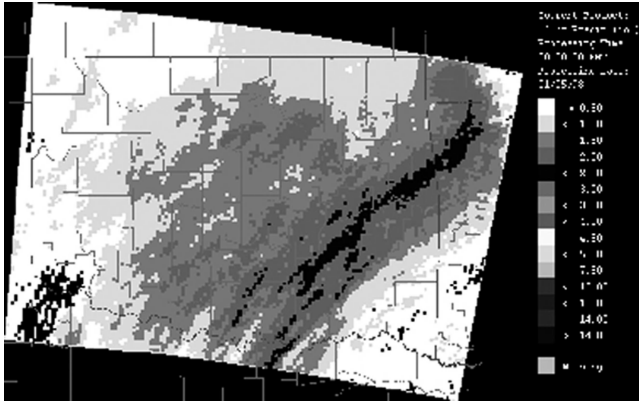
FIGURE 6.43 Map of 24h rainfall accumulation over Oklahoma. From NEXRAD radar data processed by the RPI. (Image downloaded from http://ccgwww.ou.edu/rip_images and converted to gray scale. With permission.)

Acknowledgments

References

Aase, J.K., and S.B. Idso. 1975. Solar radiation interactions with mixed prairie rangeland in natural and denuded conditions. Arch. Meteorol. Geophys. Bioklimatol. Ser. B. 23:255–264.

Allen, R.G., L.S. Pereira, D. Raes, and M. Smith. 1998. Crop evapotranspiration: Guidelines for computing crop water requirements. FAO irrigation and drainage paper 56. FAO, Rome, Italy.

Allen, R.G., M. Smith, A. Perrier, and L.S. Pereira. 1994a. An update for the definition of reference evapotranspiration. ICID Bull. 43:1–34.

Allen, R.G., M. Smith, A. Perrier, and L.S. Pereira. 1994b. An update for the calculation of reference evapotranspiration. ICID Bull. 43:35–92.

Allen, R.G., M. Tasumi, A. Morse, R. Trezza, J.L. Wright, W. Bastiaanssen, W. Krambler, I. Lorite, and C.W. Robinson. 2007b. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—Applications. J. Irrigat. Drain. Eng. 133:395–406.

Allen, R.G., M. Tasumi, and R. Trezza. 2007a. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)-model. J. Irrigat. Drain. Eng. 133:380–394.

Alsanabani, M.M. 1991. Soil water determination by time domain reflectometry: Sampling domain and geometry. Ph.D. Thesis. Department of Soil and Water Sciences, University of Arizona, Tucson, AZ.

Anderson, C.M., W.P. Kustas, and J.M. Norman. 2007. Upscaling flux observations from local to continental scales using thermal remote sensing. Agron. J. 99:240–254.

Anlauf, R., K.Ch. Kersebaum, L.Y. Ping, A. Nuske-Schüler, J. Richter, G. Springob, K.M. Syring, and J. Utermann. 1990. Models for processes in the soil—Programs and exercises. Catena Verlag, Cremlingen, West Germany.

Annandale, J.G., and C.O. Stockle. 1994. Fluctuation of crop evapotranspiration coefficients with weather: A sensitivity analysis. Irrigat. Sci. 15:1–7.

ASAE. 1988. Modeling agricultural, forest, and rangeland hydrology. Proceedings of the 1988 International Symposium, December 12–13, 1988, Chicago, IL. American Society of Agricultural Engineering, St. Joseph, MI.

ASCE. 1996. Hydrology handbook. 2nd edn. American Society of Civil Engineers, New York.

ASCE. 2005. The ASCE standardized reference evapotranspiration equation. American Society of Civil Engineers, Reston, VA.

Asrar, G., and E.T. Kanemasu. 1983. Estimating thermal diffusivity near the soil surface using Laplace transform: Uniform initial conditions. Soil Sci. Soc. Am. J. 47:397–401.

Baker, J.M., and R.J. Lascano. 1989. The spatial sensitivity of time-domain reflectometry. Soil Sci. 147:378–384.

Baker, J.M., and E.J.A. Spaans. 1994. Measuring water exchange between soil and atmosphere with TDR-microlysimetry. Soil Sci. 158:22–30.

Baldocchi, D.D., R.J. Luxmoore, and J.L. Hatfield. 1991. Discerning the forest from the trees: An essay on scaling canopy stomatal conductance. Agr. Forest Meteorol. 54:197–226.

Bastiaanssen, W.G.M., M. Menenti, R.A. Feddes, and A.A. Holtslang. 1998. A remote sensing surface energy balance algorithm for land (SEBAL): 1. Formulation. J. Hydrol. 212–213:198–212.

Bastiaanssen, W.G.M., E.J.M. Noordman, H. Pelgrum, G. Davids, B.P. Thoreson, and R.G. Allen. 2005. SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. J. Irrigat. Drain. Eng. 131:85–93.

Ben-Asher, J., A.D. Matthias, and A.W. Warrick. 1983. Assessment of evaporation from bare soil by infrared thermometry. Soil Sci. Soc. Am. J. 47:185–191.

Berbery, E.H., K.E. Mitchell, S. Benjamin, T. Smirnova, H. Ritchie, R. Hogue, and E. Radeva. 1999. Assessment of land-surface energy budgets from regional and global models. J. Geophys. Res. 104:19329–19348.

Black, T.A., G.W. Thurtell, and C.B. Tanner. 1968. Hydraulic load cell lysimeter, construction, calibration, and tests. Soil Sci. Soc. Am. Proc. 32:623–629.

Blad, B.L., and N.J. Rosenberg. 1974. Lysimetric calibration of the Bowen ratio energy balance method for evapotranspiration estimation in the central Great Plains. J. Appl. Meteorol. 13:227–236.

Bos, M.G., J.A. Replogle, and A.J. Clemmens. 1983. Flow measuring and regulating flumes. I.L.R.I., Wageningen, the Netherlands and U.S. Water Conservation Laboratory, Phoenix, AZ.

Bowen, I.S. 1926. The ratio of heat losses by conduction and by evaporation from any water surface. Phys. Rev. 27:779–787.

Bowers, S.A., and R.J. Hanks. 1965. Reflection of radiant energy from soils. Soil Sci. 100:130–138.

Brakensiek, D.L., H.B. Osborn, and W.J. Rawls. 1979. Field manual for research in agricultural hydrology. USDA agricultural handbook 224. U.S. Government Printing Office, Washington, DC.

Briggs, L.J., and H.L. Shantz. 1913. The water requirement of plants. I. Investigations in the Great Plains in 1910 and 1911. Bull. USDA Bur. Plant Ind. no. 284. USDA, Washington, DC.

Briggs, L.J., and H.L. Shantz. 1914. Relative water requirement of plants. J. Agr. Res. 3:1–64.

Bristow, K.L. 1987. On solving the surface energy balance equation for surface temperature. Agr. Forest Meteorol. 39:49–54.

Brunt, D. 1932. Notes on radiation in the atmosphere. Quart. J. Roy. Meteorol. Soc. 58:389–418.

Brutsaert, W. 1975. The roughness length for water vapor, sensible heat and other scalars. J. Atmos. Sci. 32:2028–2031.

Brutsaert, W. 1982. Evaporation into the atmosphere. D. Reidel Publishing Company, Boston, MA.

Budyko, M.I. 1958. The heat balance of the earth's surface (English translation by N.A. Stepanova). U.S. Department of Commerce, Washington, DC.

Campbell, G.S. 1977. An introduction to environmental biophysics. Springer-Verlag, New York.

Campbell, G.S. 1985. Soil physics with BASIC—Transport models for soil–plant systems. Elsevier, New York.

Campbell, G.S., C. Calissendorff, and J.H. Williams. 1991. Probe for measuring soil specific heat using a heat-pulse method. Soil Sci. Soc. Am. J. 55:291–293.

Campbell, G.S., J.D. Jungbauer, Jr., W.R. Bidlake, and R.D. Hungerford. 1994. Predicting the effect of temperature on soil thermal conductivity. Soil Sci. 158:307–313.

Carrijo, O.A., and R.H. Cuenca. 1992. Precision of evapotranspiration measurements using neutron probe. J. Irrigat. Drain. Eng. ASCE 118:943–953.

Castellví, F. 2004. Combining surface renewal analysis and similarity theory: A new approach for estimating sensible heat flux. Water Resour. Res. 40:1–20.

Chen, J. 1984. Mathematical analysis and simulation of crop micrometeorology. Ph.D. Thesis. Agricultural University, Wageningen, the Netherlands.

Choi, M., J.M. Jacobs, and W.P. Kustas. 2008. Assessment of clear and cloudy sky parameterizations for daily downwelling longwave radiation over different land surfaces in Florida, USA. Geophys. Res. Lett. 35:L20402, doi:10.1029/2008GL035731.

Cochran, P.H., L. Boersma, and C.T. Youngberg. 1967. Thermal conductivity of a pumice soil. Soil Sci. Soc. Am. Proc. 31:454–459.

Cohen, Y., and Y. Li. 1996. Validating sap flow measurement in field-grown sunflower and corn. J. Exp. Bot. 47:1699–1707.

Colaizzi, P.D., E.M. Barnes, T.R. Clarke, C.Y. Choi, P.M. Waller, J. Haberland, and M. Kostrzewski. 2003. Water stress detection under high frequency sprinkler irrigation with water deficit index. J. Irrigat. Drain. Eng. 129:36–43.

Colaizzi, P.D., S.R. Evett, T.A. Howell, and J.A. Tolk. 2006. Comparison of five models to scale daily evapotranspiration from one time of day measurements. Trans. ASABE 49:1409–1417.

Collares-Pereira, M., and A. Rabl. 1979. The average distribution of solar radiation–correlations between diffuse and hemispherical and between daily and hourly insolation values. Solar Energy 22:155–164.

Costello, T.A., and H.J. Braud, Jr. 1989. Thermal diffusivity of soil by nonlinear regression analysis of soil temperature data. Trans. ASAE 32:1281–1286.

Crawford, T.M., and C.E. Duchon. 1999. An improved parameterization for estimating effective atmospheric emissivity for use in calculating daytime downwelling longwave radiation. J. Appl. Meteorol. 38:474–480.

Cuenca, R.H. 1988. Model for evapotranspiration using neutron probe data. J. Irrigat. Drain. Eng. ASCE 114:644–663.

de Vries, D.A. 1963. Thermal properties of soils, p. 210–235. *In* W.R. Van Wijk (ed.) Physics of plant environment. North-Holland Publishing Company, Amsterdam, the Netherlands.

de Wit, C.T. 1958. Transpiration and crop yields. Versl. Landbouwk. Onderz. no. 64.6. Institute of Biological and Chemical Research on Field Crops and Herbage, Wageningen, the Netherlands.

Doorenbos, J., and W.O. Pruitt. 1977. Guidelines for predicting crop water requirements. FAO irrigation and drainage paper 24. FAO, Rome, Italy.

Duffie, J.A., and W.A. Beckman. 1991. Solar engineering of thermal processes. 2nd edn. John Wiley & Sons, New York.

Dugas, W.A., and W.L. Bland. 1991. Springtime soil temperatures in lysimeters in Central Texas. Soil Sci. 152:87–91.

Dugas, W.A., L.J. Fritschen, L.W. Gay, A.A. Held, A.D. Matthias, D.C. Reicosky, P. Steduto, and J.L. Steiner. 1991. Bowen ratio, eddy covariance, and portable chamber measurements of sensible and latent heat flux over irrigated spring wheat. Agr. Forest Meteorol. 56:1–20.

Dvoracek, M.J., and B. Hannabas. 1990. Prediction of albedo for use in evapotranspiration and irrigation scheduling. *In* Visions of the future, proceedings of the third national irrigation symposium held in conjunction with the 11th

annual international irrigation exposition, October 28–November 1, Phoenix, AZ. ASAE, St. Joseph, MI.

Evett, S.R. 1994. TDR–temperature arrays for analysis of field soil thermal properties, p. 320–327. *In* Proc. Symp. Time Domain Reflectometry Environ. Infrastruct. Min. Appl. September 7–9th, 1994. Northwestern University, Evanston, IL.

Evett, S.R., L.K. Heng, P. Moutonnet, and M.L. Nguyen (eds.). 2008. Field estimation of soil water content: A practical guide to methods, instrumentation, and sensor technology. IAEA-TCS-30. Int. At. Energy Agency, Vienna, Austria. ISSN 1018-5518. Available online at: http://www-pub.iaea.org/mtcd/publications/PubDetails.asp?pubId=7801

Evett, S.R., T.A. Howell, J.L. Steiner, and J.L. Cresap. 1993. Evapotranspiration by soil water balance using TDR and neutron scattering, p. 914–921. *In* R.G. Allen and C.M.U. Neale (eds.) Management of irrigation and drainage systems, integrated perspectives. Nat. Conf. Irrigat. Drain. Eng. Proc. ASCE, St. Joseph, MI.

Evett, S.R., T.A. Howell, R.W. Todd, A.D. Schneider, and J.A. Tolk. 1998. Evapotranspiration of irrigated alfalfa in a semi-arid environment. ASAE paper no. 98-2123.

Evett, S.R., T.A. Howell, R.W. Todd, A.D. Schneider, and J.A. Tolk. 2000. Alfalfa reference ET measurement and prediction, p. 266–272. *In* R.G. Evans, B.L. Benham, and T.P. Trooien (eds.) Proc. 4th Decennial Nat. Irrigat. Symp. November 14–16, Phoenix, AZ. ASAE, St. Joseph, MI.

Evett, S.R., and R.J. Lascano. 1993. ENWATBAL: A mechanistic evapotranspiration model written in compiled BASIC. Agron. J. 85:763–772.

Evett, S.R., A.D. Matthias, and A.W. Warrick. 1994. Energy balance model of spatially variable evaporation from bare soil. Soil Sci. Soc. Am. J. 58:1604–1611.

Evett, S.R., N.T. Mazahrih, M.A. Jitan, M.H. Sawalha, P.D. Colaizzi, and J.E. Ayars. 2009a. A weighing lysimeter for crop water use determination in the Jordan Valley, Jordan. Trans. ASABE 52:155–169.

Evett, S.R., R.C. Schwartz, J.A. Tolk, and T.A. Howell. 2009b. Soil profile water content determination: Spatio-temporal variability of electromagnetic and neutron probe sensors in access tubes. Vadose Zone J. 8:1–16.

Evett, S.R., A.W. Warrick, and A.D. Matthias. 1995. Wall material and capping effects on microlysimeter performance. Soil Sci. Soc. Am. J. 59:329–336.

Farrell, D.A., E.L. Greacen, and G.C. Gurr. 1966. Vapor transfer in soil due to air turbulences. Soil Sci. 102:305–313.

Finnigan, J.J., R. Clement, Y. Malhi, R. Leuning, and H.A. Cleugh. 2003. A re-evaluation of long-term flux measurement techniques, part I: Averaging and coordinate rotation. Bound. Layer Meteorol. 107:1–48.

Foken, T., and B. Wichura. 1996. Tools for quality assessment of surface-based flux measurements. Agr. Forest Meteorol. 78:83–105.

Fox, M.J. 1968. A technique to determine evaporation from dry stream beds. J. Appl. Meteorol. 7:697–701.

French, A.N., D.J. Hunsaker, T.R. Clarke, G.J. Fitzgerald, W.E. Luckett, and P.J. Pinter, Jr. 2007. Energy balance estimation of evapotranspiration for wheat grown under variable management practices in Central Arizona. Trans. ASABE 50:2059–2071.

Fröhlich, C., and J. Lean. 1998. The sun's total irradiance: Cycles, trends and related climate change uncertainties since 1976. Geophys. Res. Lett. 25:4377–4380.

Furon, A.C., J.S. Warland, and C. Wagner-Riddle. 2007. Analysis of scaling-up resistances from leaf to canopy using numerical solutions. Agron. J. 99:1483–1491.

Gates, D.M. 1980. Biophysical ecology. Springer-Verlag, New York.

Gonzalez-Altozano, P., E.W. Pavel, J.A. Oncins, J. Doltra, M. Cohen, T. Paco, R. Massai, and J.R. Castel. 2008. Comparative assessment of five methods of determining sap flow in peach trees. Agr. Water Manage. 95:503–515.

Gorry, P.A. 1990. General least-squares smoothing and differentiation by the convolution (Savitsky–Golay) method. Anal. Chem. 62:570–573.

Goudriaan, J. 1977. Crop micrometeorology: A simulation study. PUDOC, Wageningen, the Netherlands.

Gowda, P.H., J.L. Chavez, P.D. Colaizzi, S.R. Evett, T.A. Howell, and J.A. Tolk. 2008. ET mapping for agricultural water management: Present status and challenges. Irrigat. Sci. 26:223–237.

Grebet, R., and R.H. Cuenca. 1991. History of lysimeter design and effects of environmental disturbances, p. 10–19. *In* Lysimeters for evapotranspiration and environmental measurements, Proc. Int. Symp. Lysimetry, July 23–25, 1991, Honolulu, Hawaii. ASAE, New York.

Halldin, S., and A. Lindroth. 1992. Errors in net radiometry: Comparison and evaluation of six radiometer designs. J. Atmos. Ocean. Tech. 9:762–783.

Hanks, J., and J.T. Ritchie (eds.) 1991. Modeling plant and soil systems. Agronomy monograph 31. ASA, CSSA, SSSA, Madison, WI.

Hann, C.T., H.P. Johnson, and D.L. Brakensiek (eds.). 1982. Hydrologic modeling of small watersheds. ASAE monograph no. 5. ASAE, St. Joseph, MI.

Hatfield, J.L., and J.M. Baker (eds.). 2005. Micrometeorology in agricultural systems. Agronomy monograph no. 47. ASA, CSSA, SSSA. Madison, WI.

Hatfield, J.L., R.J. Reginato, and S.B. Idso. 1983. Comparison of long-wave radiation calculation methods over the United States. Water Resour. Res. 19:285–288.

Haverd, V., M. Cuntz, R. Leuning, and H. Keith. 2007. Air and biomass heat storage fluxes in a forest canopy: Calculation within a soil vegetation atmosphere transfer model. Agr. Forest Meteorol. 147:125–139.

Heilman, J.L., C.L. Brittin, and C.M.U. Neale. 1989. Fetch requirements for Bowen ratio measurements of latent and sensible heat fluxes. Agr. Forest Meteorol. 44:261–273.

Hillel, D. 1980. Fundamentals of soil physics. Academic Press, San Diego, CA.

Horton, R., P.J. Wierenga, and D.R. Nielsen. 1983. Evaluation of methods for determining the apparent thermal diffusivity of soil near the surface. Soil Sci. Soc. Am. J. 47:25–32.

Houser, P.R., C. Harlow, W.J. Shuttleworth, T.O. Keefer, W.E. Emmerich, and D.C. Goodrich. 1998. Evaluation of multiple flux measurement techniques using water balance information at a semi-arid site, p. 84–87. *In* Proc. Spec. Symp. Hydrol. American Meteorological Society, Phoenix, AZ.

Howell, T.A., A.D. Schneider, D.A. Dusek, T.H. Marek, and J.L. Steiner. 1995. Calibration and scale performance of Bushland weighing lysimeters. Trans. ASAE 38:1019–1024.

Howell, T.A., A.D. Schneider, and M.E. Jensen. 1991. History of lysimeter design and use for evapotranspiration measurements, p. 1–9. *In* Lysimeters evapotranspiration and environmental measurements. Proc. Int. Symp. Lysimetry, July 23–25, 1991, Honolulu, Hawaii. ASAE, New York.

Howell, T.A., J.L. Steiner, S.R. Evett, A.D. Schneider, K.S. Copeland, D.A. Dusek, and A. Tunick. 1993. Radiation balance and soil water evaporation of bare Pullman clay loam soil, p. 922–929. *In* R.G. Allen and C.M.U. Neale (eds.) Management of irrigation and drainage systems, integrated perspectives. Proc. Nat. Conf. Irrigat. Drain. Eng. Park City, UT, July 21–23, 1993. Am. Soc. Civil Eng., New York.

Howell, T.A., J.L. Steiner, A.D. Schneider, S.R. Evett, and J.A. Tolk. 1994. Evapotranspiration of irrigated winter wheat, sorghum and corn. ASAE paper no. 94-2081. ASAE, St. Joseph, MI.

Howell, T.A., and J.A. Tolk. 1990. Calibration of soil heat flux transducers. Theor. Appl. Climatol. 42:263–272.

Howell, T.A., S.R. Evett, A.D. Schneider, D.A. Dusek, and K.S. Copeland. 2000. Irrigated fescue grass ET compared with calculated reference grass ET, p. 228–242. *In* Robert G. Evans, Brian L. Benham, and Todd P. Trooien (eds.). Proceedings of the 4th Decennial National Irrigation Symposium, November 14–16, Phoenix, AZ. ASABE, St. Joseph, MO.

Idso, S.B. 1981. A set of equations for full spectrum and 8- to 14-μm and 10.5- to 12.5-μm thermal radiation from cloudless skies. Water Resour. Res. 17:295–304.

Idso, S.B., R.D. Jackson, R.J. Reginato, B.A. Kimball, and F.S. Nakayama. 1975. The dependence of bare soil albedo on soil water content. J. Appl. Meteorol. 14:109–113.

Idso, S.B., and R.J. Reginato. 1974. Assessing soil-water status via albedo measurement. Hydrol. Water Resour. Arizona SW 4:41–54.

Idso, S.B., R.J. Reginato, R.D. Jackson, B.A. Kimball, and F.S. Nakayama. 1974. The three stages of drying in a field soil. Soil Sci. Soc. Am. Proc. 38:831–837.

Irmak, S., A. Irmak, J.W. Jones, T.A. Howell, J.M. Jacobs, R.G. Allen et al. 2003. Predicting daily net radiation using minimum climatological data. J. Irrigat. Drain. Eng. 129:256–269.

Iwata, S., T. Tabuchi, and B.P. Warkentin. 1988. Soil–water interactions: Mechanisms and applications. Marcel Dekker, Inc., New York.

Jensen, M.E., R.D. Burman, and R.G. Allen (eds.). 1990. Evapotranspiration and irrigation water requirements. ASCE manuals and reports on engineering practices, no. 70. ASCE, New York.

Jones, H.G. 1992. Plants and microclimate: A quantitative approach to environmental plant physiology. 2nd edn. The Cambridge University Press, Cambridge, U.K.

Kaimal, J.C., and J.J. Finnigan. 1994. Atmospheric boundary layer flows: Their structure and measurement. Oxford University Press, New York.

Kimball, B.A., R.D. Jackson, R.J. Reginato, F.S. Nakayama, and S.B. Idso. 1976. Comparison of field measured and calculated soil heat fluxes. Soil Sci. Soc. Am. J. 40:18–24.

Kimball, B.A., R.L. LaMorte, P.J. Pinter, Jr., G.W. Wall, D.J. Hunsaker, F.J. Adamsen, S.W. Leavitt, T.L. Thompson, A.D. Matthias, and T.J. Brooks. 1999. Free-air CO_2 enrichment and soil nitrogen effects on energy balance and evapotranspiration of wheat. Water Resour. Res. 35:1179–1190.

Kluitenberg, G.J., K.L. Bristow, and B.S. Das. 1995. Error analysis of heat pulse method for measuring soil heat capacity, diffusivity, and conductivity. Soil Sci. Soc. Am. J. 59:719–726.

Kondo, J., N. Saigusa, and T. Sato. 1992. A model and experimental study of evaporation from bare-soil surfaces. J. Appl. Meteorol. 31:304–312.

Kreith, F., and W.D. Sellers. 1975. General principles of natural evaporation, p. 207–227. *In* D.A. de Vries and N.H. Afgan (eds.) Heat and mass transfer in the biosphere. Part 1. John Wiley & Sons, New York.

Kustas, W.P. 2002. Evaporation, p. 524–530. *In* R. Lal (ed.) Encyclopedia of soil science. Marcel Dekker, Inc., New York.

Kustas, W.P., N. Agam, M.C. Anderson, F. Li, and P.D. Colaizzi. 2007. Potential errors in the application of thermal-based energy balance models with coarse resolution data. Proc. SPIE Vol. 6742:674208.

Kustas, W.P., and J.M. Norman. 1999. Evaluation of soil and vegetation heat flux predictions using a simple two-source model with radiometric temperatures for partial canopy cover. Agr. Forest Meteorol. 94:13–29.

Kustas, W.P., J.H. Prueger, J.L. Hatfield, K. Ramalingam, and L.E. Hipps. 2000. Variability in soil heat flux from a mesquite dune site. Agr. Forest Meteorol. 103:249–264.

Kustas, W.P., J.H. Prueger, L.E. Hipps, J.L. Hatfield, and D. Meek. 1998. Inconsistencies in net radiation estimates from use of several models of instruments in a desert environment. Agr. Forest Meteorol. 90:257–263.

Lascano, R.J., and J.L. Hatfield. 1992. Spatial variability of evaporation along two transects of a bare soil. Soil Sci. Soc. Am. J. 56:341–346.

Lascano, R.J., and C.H.M. van Bavel. 1983. Experimental verification of a model to predict soil moisture and temperature profiles. Soil Sci. Soc. Am. J. 47:441–448.

Lascano, R.J., and C.H.M. van Bavel. 1986. Simulation and measurement of evaporation from a bare soil. Soil Sci. Soc. Am. J. 50:1127–1132.

Lascano, R.J., and C.H.M. van Bavel. 2007. Explicit and recursive calculation of potential and actual evapotranspiration. Agron. J. 99:585–590.

Lascano, R.J., C.H.M. van Bavel, J.L. Hatfield, and D.R. Upchurch. 1987. Energy and water balance of a sparse crop: Simulated and measured soil and crop evaporation. Soil Sci. Soc. Am. J. 51:1113–1121.

Lee, X., J. Finnigan, and K.T. Paw U. 2004. Coordinate systems and flux bias error. *In* X. Lee, W.J. Massman, and B. Law (eds.). Handbook of micrometeorology: A guide for surface flux measurement and analysis. Kluwer, Dordrecht, the Netherlands.

Legates, D.R., K.R. Nixon, T.D. Stockdale, and G. Quelch. 1996. Soil water management using a water resource decision support system and calibrated WSR-88D precipitation estimates. Proc. AWRA Symp. GIS and Water Resources. American Water Resources Association, Fort Lauderdale, FL.

Leuning, R. 2004. Measurements of trace gas fluxes in the atmosphere using eddy covariance: WPL corrections revisited. *In* X. Lee, W.J. Massman, and B. Law (eds.) Handbook of micrometeorology: A guide for surface flux measurement and analysis. Kluwer, Dordrecht, the Netherlands.

Leuning, R., and M.J. Judd. 1996. The relative merits of open- and closed-path analyzers for measurement of eddy fluxes. Global Change Biol. 2: 241–254.

List, R.J. 1971. Smithsonian meteorological tables. Smithsonian Institution Press, Washington, DC.

Lourence, F., and R. Moore. 1991. Prefabricated weighing lysimeter for remote research stations, p. 423–439. *In* Lysimeters for evapotranspiration and environmental measurements. Proceedings of the International Symposium on Lysimetry, July 23–25, 1991, Honolulu, Hawaii. ASAE, New York.

Marek, T., G. Piccinni, A. Schneider, T. Howell, M. Jett, and D. Dusek. 2006. Weighing lysimeters for the determination of crop water requirements and crop coefficients. Appl. Eng. Agr. 22:851–856.

Marek, T.H., A.D. Schneider, T.A. Howell, and L.L. Ebeling. 1988. Design and construction of large weighing monolithic lysimeters. Trans. ASAE 31:477–484.

Massman, W.J. 2000. A simple method for estimating frequency response corrections for eddy covariance systems. Agr. Forest Meteorol. 104:185–198.

Massman, W.J. 2004. Concerning the measurement of atmospheric trace gas fluxes with open- and closed-path eddy covariance system: The WPL terms and spectral attenuation. *In* X. Lee, W.J. Massman, and B. Law (eds.) Handbook of micrometeorology: A guide for surface flux measurement and analysis. Kluwer, Dordrecht, the Netherlands.

Massman, W.J., and R. Clement. 2004. Uncertainty in eddy covariance flux estimates resulting from spectral attenuation. *In* X. Lee, W.J. Massman, and B. Law (eds.) Handbook of micrometeorology: A guide for surface flux measurement and analysis. Kluwer, Dordrecht, the Netherlands.

Mazahrih, N.Th., N. Katbeh-Bader, S.R. Evett, J.E. Ayars, and T.J. Trout. 2008. Field calibration accuracy and utility of four down-hole water content sensors. Vadose Zone J. 7:992–1000.

McArthur, A.J. 1990. An accurate solution to the Penman equation. Agr. Forest Meteorol. 51:87–92.

McArthur, A.J. 1992. The Penman form equations and the value of delta: A small difference of opinion or a matter of fact? Agr. Forest Meteorol. 57:305–308.

McCullough, E.C., and W.P. Porter. 1971. Computing clear day solar radiation spectra for the terrestrial ecological environment. Ecology 52:1008–1015.

McNaughton, K.G. 1988. Surface temperature and the surface energy balance: Commentary, p. 154–159. *In* W. Steffan and O.T. Denmead (eds.) Flow and transport in the natural environment: Advances and applications. Springer, Berlin, Germany.

Meesters, A.G.C.A., and H.F. Vugts. 1996. Calculation of heat storage in stems. Agr. Forest Meteorol. 78:181–202.

Milly, P.C.D. 1991. A refinement on the combination equations for evaporation. Surv. Geophys. 12:145–154.

Moncrieff, J.B., J.M. Massheder, H.A.R. De Bruin, J. Elbers, T. Friborg, B. Heusinkveld, P. Kabat, S. Scott, H. Soegaard, and A. Verhoef. 1997. A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. J. Hydrol. 189:589–611.

Monteith, J.L. 1961. The reflection of short-wave radiation by vegetation. Quart. J. Roy. Meteorol. Soc. 85:386–392.

Monteith, J.L. 1965. Evaporation and the environment, p. 205–234. *In* The state and movement of water in living organisms. XIXth Symp. Soc. Exp. Biol. Cambridge University Press, Swansea, U.K.

Monteith, J.L., and G. Szeicz. 1961. The radiation balance of bare soil and vegetation. Quart. J. Roy. Meteorol. Soc. S7:159–170.

Monteith, J.L., and M.H. Unsworth. 1990. Principles of environmental physics. 2nd edn. Edward Arnold, London, U.K.

Moore, C.J. 1986. Frequency response corrections for eddy correlation systems. Bound. Layer Meteorol. 37:17–35.

Murray, F.W. 1967. On the computation of saturation vapor pressure. J. Appl. Meteorol. 6:203–204.

Nassar, I.N., and R. Horton. 1989. Determination of the apparent thermal diffusivity of a nonuniform soil. Soil Sci. 147:238–244.

Nassar, I.N., and R. Horton. 1990. Determination of soil apparent thermal diffusivity from multiharmonic temperature analysis for nonuniform soils. Soil Sci. 149:125–130.

Noborio, K., K.J. McInnes, and J.L. Heilman. 1996. Measurements of soil water content, heat capacity, and thermal conductivity with a single TDR probe. Soil Sci. 161:22–28.

Norman, J.M., M.C. Anderson, W.P. Kustas, A.N. French, J. Mecikalski, R. Torn, G.R. Diak, T.J. Schmugge, and B.C.W. Tanner. 2003. Remote sensing of surface energy fluxes at 101-m pixel resolutions. Water Resour. Res. 39:1221.

Norman, J.M., and J.M. Baker. 2002. Have eddy-flux measurements improved over 30 years? Presented in the symposium, "Physical measurements in the soil–plant–atmosphere system: Advances in measurements at and above the ground surface—A tribute to Dr. George Thurtell" at the 2002 Annual International Meeting of the American Society of Agronomy, November 10–14, 2002, Indianapolis, MN. Accessed March 4, 2009 at https://www.agronomy.org/divisions/a03/thurtell/Norman_GT02.pdf

Norman, J.M., W.P. Kustas, and K.S. Humes. 1995. A two-source approach for estimating soil and vegetation energy fluxes from observations of directional radiometric surface temperature. Agr. Forest Meteorol. 77:263–293.

Oke, T.R. 1978. Boundary layer climates. Methuen, New York.

Paw U, K.T., D. Baldocchi, T.P. Meyers, and K.B. Wilson. 2000. Correction of eddy covariance measurements incorporating both advective effects and density fluxes. Bound. Layer Meteorol. 97:487–511.

Paw U, K.T., and T. Meyers. 1989. Investigations with a higher-order canopy turbulence model into mean source-sink levels and bulk canopy resistances. Agr. Forest Meteorol. 47:259–271.

Peart, R.M., and R.B. Curry. 1998. Agricultural systems modeling and simulation. Marcel Dekker, Inc., New York.

Penman, H.L. 1948. Natural evapotranspiration from open water, bare soil and grass. Proc. Roy. Soc. Lond. Ser. A. 193:120–145.

Pereira, L.S., B.J. van den Broek, P. Kabat, and R.G. Allen (eds.) 1995. Crop-water-simulation models in practice. Wageningen Press, Wageningen, the Netherlands.

Qiu, G.Y., J. Ben-Asher, T. Yano, and K. Momii. 1999. Estimation of soil evaporation using the differential temperature method. Soil Sci. Soc. Am. J. 63:1608–1614.

Raupach, M., and J. Finnigan. 1988. Single-layer models of evaporation from plant canopies are incorrect but useful, whereas multilayer models are correct but useless. Aust. J. Plant Physiol. 15:705–716.

Richter, J. 1987. The soil as a reactor: Modeling processes in the soil. Catena Verlag, Cremlingen, West Germany.

Riha, S.J., K.J. McKinnes, S.W. Childs, and G.S. Campbell. 1980. A finite element calculation for determining thermal conductivity. Soil Sci. Soc. Am. J. 44:1323–1325.

Ritchie, J.T., T.A. Howell, W.S. Meyer, and J.L. Wright. 1996. Sources of biased errors in evaluating evapotranspiration equations, p. 147–157. *In* C.R. Camp, E.J. Sadler, and R.E. Yoder (eds.) Evapotranspiration and irrigation scheduling. Proc. Int. Conf. Nov. 3–6, 1996, San Antonio, TX.

Rochette, P., E. Pattey, R.L. Desjardins, L.M. Dwyer, D.W. Stewart, and P.A. Dube. 1991. Estimation of maize (*Zea mays* L.) canopy conductance by scaling up leaf stomatal conductance. Agr. Forest Meteorol. 54:241–261.

Rosenberg, N.J. 1969. Seasonal patterns of evapotranspiration by irrigated alfalfa in the Central Great Plains. Agron. J. 61(6):879–886.

Rosenberg, N.J., B.L. Blad, and S.B. Verma. 1983. Microclimate, the biological environment. John Wiley & Sons, New York.

Rudolf, J., F. Rothfuss, and R. Conrad. 1996. Flux between soil and atmosphere, vertical concentration profiles in soil, and turnover of nitric oxide: 1. Measurements on a model soil core. J. Atmos. Chem. 23:253–273.

Russell, G., B. Marshall, and P.G. Jarvis (eds.). 1989. Plant canopies: Their growth, form and function. Cambridge University Press, Cambridge, U.K.

Savage, M.J., K.J. McInnes, and J.L. Heilman. 1995. Placement height of eddy correlation sensors above a short turfgrass surface. Agr. Forest Meteorol. 74:195–204.

Savage, M.J., K.J. McInnes, and J.L. Heilman. 1996. The "footprints" of eddy correlation sensible heat flux density, and other micrometeorological measurements. S. Afr. J. Sci. 92:137–142.

Savitsky, A., and M.J.E. Golay. 1964. Smoothing and differentiation of data by simplified least squares. Anal. Chem. 36:1627–1639.

Schmid, H.P. 1997. Experimental design for flux measurements: Matching scales of observations and fluxes. Agr. Forest Meteorol. 87:179–200.

Schneider, A.D., T.A. Howell, T.A. Moustafa, S.R. Evett, and W.S. Abou-Zeid. 1996. A simplified weighing lysimeter for developing countries, p. 289–294. *In* C.R. Camp, E.J. Sadler, and R.E. Yoder (eds.) Proceedings of the International Conference on Evapotranspiration and Irrigation Scheduling. November 3–6, 1996, San Antonio, TX.

Schneider, A.D., T.A. Howell, A.T.A. Moustafa, S.R. Evett, and W.S. Abou-Zeid. 1998. A simplified weighing lysimeter for monolithic or reconstructed soils. Appl. Eng. Agr. 14:267–273.

Schuepp, P.J., M.Y. LeClerc, J.I. MacPherson, and R.L. Desjardins. 1990. Footprint prediction of scalar fluxes from analytical solutions of the diffusion equation. Bound. Layer Meteorol. 50:355–373.

Sepaskhah, A.R., and L. Boersma. 1979. Thermal conductivity of soils as a function of temperature and water content. Soil Sci. Soc. Am. J. 43:439–444.

Shanan, L., and N.H. Tadmor. 1979. Micro-catchment systems for arid zone development. Center of International Agricultural Cooperation, Ministry of Agriculture, Rehovot, Israel.

Shuttleworth, W.J. 2006. Towards one-step estimation of crop water requirements. Trans. ASABE 49:925–935.

Skidmore, E.L., J.D. Dickerson, and H. Schimmelpfennig. 1975. Evaluating surface-soil water content by measuring reflection. Soil Sci. Soc. Am. J. 39:238–242.

Snyder, R.L., K.T. Paw U, D. Spano, and P. Duce. 1997. Surface renewal estimates of evapotranspiration, p. 49–55. *In* K.S. Chartzoulakis (ed.) Proc. 2nd Int. Sym. on Irrigation of Hort. Crops. ISHS Acta Hortic., Leuven, Belgium.

Spitters, C.J.T., H.A.J.M. Toussaint, and J. Goudriaan. 1986. Separating the diffuse and direct component of global radiation and its implications for modeling canopy photosynthesis. Part I. Components of incoming radiation. Agr. Forest Meteorol. 38:217–229.

Stannard, D.I. 1997. A theoretically based determination of Bowen-ratio fetch requirements. Bound. Layer Meteorol. 83:375–406.

Swinbank, W.C. 1951. The measurement of vertical transfer of heat and water vapor by eddies in the lower atmosphere. J. Meteorol. 8:135–145.

Tanner, C.B., and G.W. Thurtell. 1969. Anemoclinometer measurements of Reynolds stress and heat transport in the atmospheric surface layer. ECOM 66-G22-F, ECOM, United States Army Electronics Command, Research and Development. U.S. Army Electronics Command, Atmospheric Sciences Laboratory, Fort Huachuca, AZ.

Todd, R.W., S.R. Evett, and T.A. Howell. 2000. The Bowen ratio-energy balance method for estimating latent heat flux of irrigated alfalfa evaluated in a semi-arid, advective environment. Agr. Forest Meteorol. 103:335–348.

Tolk, J.A., S.R. Evett, and T.A. Howell. 2006. Advection influences on evapotranspiration of alfalfa in a semiarid climate. Agron. J. 98:1646–1654.

Tracy, C.R., F.H. van Berkum, J.S. Tsuji, R.D. Stevenson, J.A. Nelson, B.M. Barnes, and R.B. Huey. 1984. Errors resulting from linear approximations in energy balance equations. J. Therm. Biol. 9:261–264.

Turner, N.C. 1991. Measurement and influence of environmental and plant factors on stomatal conductance in the field. Agr. Forest Meteorol. 54:137–154.

Twine, T.E., W.P. Kustas, J.M. Norman, D.R. Cook, P.R. Houser, T.P. Meyers, J.H. Prueger, P.J. Starks, and M.L. Wesely. 2000. Correcting eddy-covariance flux underestimates over a grassland. Agr. Forest Meteorol. 103:279–300.

Unland, H.E., P.R. Houser, W.J. Shuttleworth, and Z.-L. Zang. 1996. Surface flux measurement and modeling at a semi-arid Sonoran desert site. Agr. Forest Meteorol. 82:119–153.

van Bavel, C.H.M. 1966. Potential evaporation: The combination concept and its experimental verification. Water Resour. Res. 2:455–467.

van Bavel, C.H.M., and D.I. Hillel. 1976. Calculating potential and actual evaporation from a bare soil surface by simulation of concurrent flow of water and heat. Agr. Meteorol. 17:453–476.

van Genuchten, M.Th., F.J. Leij, and S.R. Yates. 1991. The RETC code for quantifying the hydraulic functions of unsaturated soils. EPA/600/2-91/065. R.S. Kerr Environ. Res. Lab., USEPA, ADA, OK.

Van Toai, T., D. Major, M. McDonald, J. Schepers, and L. Tarpley (eds.). 2003. Digital imaging and spectral techniques: Applications to precision agriculture and crop physiology. ASA, CSSA, SSSA, Madison, WI.

Van Wijk, W.R., and D.W. Scholte Ubing. 1963. Radiation, p. 62–101. *In* W.R. Van Wijk (ed.) Physics of plant environment. North-Holland Publishing Company, Amsterdam, the Netherlands.

Verma, S.B., N.J. Rosenberg, and B.L. Blad. 1978. Turbulent exchange coefficients for sensible heat and water vapor under advective conditions. J. Appl. Meteorol. 17:330–338.

Vieux, B.E., and N.S. Farajalla. 1996. Temporal and spatial aggregation of NEXRAD rainfall estimates on distributed storm runoff simulation. Third Intl. Conf. GIS and Environmental Modeling, January 21–25, 1996, Santa Fe, NM.

Watts, D.B., E.T. Kanemasu, and C.B. Tanner. 1990. Modified heat-meter method for determining soil heat flux. Agr. Forest Meteorol. 49:311–330.

Weast, R.C. (ed.). 1982. Handbook of chemistry and physics. CRC Press, Boca Raton, FL.

Webb, E.K. 1982. On the correction of flux measurements for effects of heat and water vapour transfer. Bound. Layer Meteorol. 23:251–254.

Webb, E.K., G.I. Pearman, and R. Leuning. 1980. Correction of the flux measurements for density effects due to heat and water vapour transfer. Quart. J. Roy. Meteorol. Soc. 106:85–100.

Weiss, A. 1982. An experimental study of net radiation, its components and prediction. Agron. J. 74:871–874.

Widmoser, P. 2009. A discussion on and alternative to the Penman–Monteith equation. Agr. Water Manage. 96:711–721.

Wierenga, P.J., D.R. Nielson, and R.J. Hagan. 1969. Thermal properties of a soil, based upon field and laboratory measurements. Soil Sci. Soc. Am. J. 44:354–360.

Wilczak, J.M., S.P. Oncley, and S.A. Stage. 2001. Sonic anemometer tilt correction algorithms. Bound. Layer Meteorol. 99:127–150.

Wraith, J.M., and J.M. Baker. 1991. High-resolution measurement of root water uptake using automated time-domain reflectometry. Soil Sci. Soc. Am. J. 55:928–932.

Wright, J.L. 1982. New evapotranspiration crop coefficients. J. Irrigat. Drain. Div. ASCE 108:57–74.

Wright, J.L. 1990. Comparison of ET measured with neutron moisture meters and weighing lysimeters, p. 202–209. *In* Irrigation and drainage: Proceedings of the National Conference. Durango, Colorado, July 11–13, 1990. ASCE, New York.

Wright, J.L. 1991. Using weighing lysimeters to develop evapotranspiration crop coefficients. *In* R.G. Allen, T.A. Howell, W.O. Pruitt, I.A. Walter, and M.E. Jensen (eds.) Lysimeters for evapotranspiration and environmental measurements. Proc. Int. Symp. Lysimetry, July 23–25, Honolulu, Hawaii. ASCE, New York.

Wright, J.L., and M.E. Jensen. 1972. Peak water requirements of crops in Southern Idaho. J. Irrigat. Drain. Div. ASCE 96:193–201.

Young, M.H., P.J. Wierenga, and C.F. Mancino. 1997. Monitoring near-surface soil water storage in turfgrass using time domain reflectometry and weighing lysimetry. Soil Sci. Soc. Am. J. 61:1138–1146.

Zapata, N., and A. Martinez-Cob. 2002. Evaluation of the surface renewal method to estimate wheat evapotranspiration. Agr. Water Manage. 55:141–157.

7
Solute Transport

Feike J. Leij
*California State University,
Long Beach*

Antonella Sciortino
*California State University,
Long Beach*

7.1 Introduction

Soil scientists and agricultural engineers have traditionally been interested in the behavior and effectiveness of agricultural chemicals (fertilizers and pesticides) applied to soils for enhancing crop growth, as well as in the effect of salts and other dissolved substances in the soil profile on plant growth. More recently, concern for the quality of the vadose zone and possible contamination of groundwater has provided a major impetus for studying solute transport in soils.

The movement and fate of solutes in the subsurface are affected by a large number of physical, chemical, and microbiological processes requiring a broad array of mathematical and physical sciences to study and describe solute transport. A range of experimental and mathematical procedures may be employed to quantify transport in soils. Transport of a dissolved substance (solute) depends on the magnitude and direction of the solvent (water) flux; considerable experimental and numerical effort may be needed to determine the transient flow regime in unsaturated soils. Furthermore, the determination of solute concentrations is not always straightforward, particularly if the solute is involved in partitioning between different phases or subject to transformations. Solute transport in porous media is studied in many scientific disciplines such as soil science, civil and environmental engineering, geology, and chemistry. A vast body of literature exists on the subject, and this chapter provides a somewhat subjective introductory treatment of the solute transport in soils. First, the standard transport mechanisms pertaining to the fundamental advection–dispersion equation (ADE) will be introduced in Section 7.2. This equation, also known as the convection–dispersion equation, is most often used to model solute transport in porous media. The movement of a solute that undergoes adsorption by the soil requires modifications of the ADE, particularly if several solute species are present that may participate in a number of different reactions. Section 7.3 is devoted to analytical and numerical methods for quantifying solute concentrations as a function of time and space. The traditional advection–dispersion concept is not always adequate to describe solute transport in field soils. Section 7.4 describes the stream tube model as an example of an alternative transport model that may be better suited to model transport in real-world situations. Finally, Section 7.5 deals with the dissolution of nonaqueous phase contaminants into the aqueous phase.

7.2 Advection–Dispersion Equation

Consider the transport of a chemical species in a three-phase soil–air–water system, and assume that the chemical species (the solute) is completely miscible with water (the solvent). At the macroscopic level and for one-dimensional flow, the mass balance equation for a solute species subject to arbitrary reactions is given as

$$\frac{\partial(\theta C)}{\partial t} = -\frac{\partial J_s}{\partial x} + \theta R_s,$$
(7.1)

where

  θ is the volumetric water content ($L^3 L^{-3}$)

  C is the solute concentration expressed as solute mass-per-solvent volume ($M L^{-3}$)

  t is time (T)

  x is position (L)

  J_s is the solute flux expressed in solute mass-per-cross-sectional area of soil-per-unit time ($M L^{-2} T^{-1}$)

  R_s denotes arbitrary solute sinks (<0) or sources (>0) due to zero- or higher-order rate processes in the aqueous phase or transfer to another phase ($M L^{-3} T^{-1}$)

For a conservative or nonreactive solute, R_s equals 0. Similar equations may be derived for multidimensional flow and transport. The solute flux is composed of an advective and a dispersive component according to

$$J_s = J_w C + J_D, \tag{7.2}$$

where

  J_w is the volumetric water flux ($L T^{-1}$), that is, the Darcian velocity expressed as volume of water-per-cross-sectional area of soil-per-unit time

  J_D is a solute flux to quantify transport caused by a gradient in the solute concentration ($M L^{-2} T^{-1}$), also per unit area of soil

7.2.1 Solute Spreading

The movement of a solute with flowing water, described by the solute flux ($J_w C$), is referred to as advection or convection. Because dissolved substances move in a passive fashion, advective transport can be readily quantified when the solvent flux (J_w) is known. The water flux is generally a function of time and position. However, for transport in laboratory soil columns, it is often constant while simplifications may also be possible for field studies.

Even if the macroscopic water flux is known or can be measured, the velocity at smaller pore scales is not easily determined. Variations in the microscopic velocity will lead to unequal solute movement in the direction of flow. This movement is quantified by means of the dispersive solute flux. If, during steady water flow, the solute concentration of the solution at the inlet of a water-saturated soil column is changed abruptly at time $t = 0$, the observed breakthrough of a solute at the column outlet at times $t > 0$ will not exhibit an equivalent abrupt change (Nielsen and Biggar, 1961). The solute concentration will change more gradually with time as a result of (hydrodynamic) dispersion, which quantifies the effects of both mechanical dispersion and diffusion. Molecular diffusion and mechanical dispersion will be discussed first for "free" solutions and then for soil solutions.

7.2.1.1 Diffusion

Molecular or ionic diffusion is an important mechanism for solute transport in soils in directions where there is little or no water flow. A net transfer of molecules of a solute species usually occurs from regions with higher to lower concentrations as

the result of diffusion as described by Fick's first law. Diffusive transport in the free or bulk solution is given by

$$J_{dif} = -D_o \frac{\partial C}{\partial x}, \tag{7.3}$$

where

  J_{dif} is the one-dimensional mass flux ($M L^2 T^{-1}$)

  D_o is the coefficient of molecular diffusion for a free or bulk solution ($L^2 T^{-1}$)

Many publications exist that provide data on D_o (Kemper, 1986; Lide, 1995).

The experimentally observed proportionality between J_{dif} and the concentration gradient can be described at the molecular or ionic level with a balance of forces. The driving force for particle movement from higher to lower concentrations is the gradient in the chemical potential. For mixtures with ideal behavior, the chemical potential can be expressed in terms of its mole fraction. For solutions with nonideal mixing behavior, the activity coefficient of the solute needs to be determined. Ionic activity coefficients can be estimated with the extended Debye–Hückel equation or the Davies extension for solutions up to 0.5 M. Activity coefficients for a greater concentration range up to 16 M can be estimated with the Pitzer virial equations (Pitzer, 1979; Harvie and Weare, 1980). Codiffusion or counterdiffusion occurs in systems with multiple ion species. Diffusion rates for individual species as predicted solely by Fick's model would violate the electroneutrality principle. The ionic diffusion flux consists then of a term for ordinary Fickian diffusion and a term accounting for electric transference of ions. The corresponding diffusion coefficient is related to the ionic mobility using the Nernst–Planck equation (Helfferich, 1962).

To characterize diffusion in soils, the coefficient for diffusion of a substance in a free solution is typically adjusted with terms accounting for a reduced solution phase (a smaller cross-sectional area available for diffusion) and an increased path length. A general treatment of diffusion in soils can be found in Olsen and Kemper (1968) and Nye (1979). The macroscopic diffusive flux per unit area of soil can be written as

$$J_{dif} = -\theta D_{dif} \frac{\partial C}{\partial x}, \tag{7.4}$$

with D_{dif} as the coefficient of molecular or ionic diffusion for the liquid phase of the soil. Diffusion coefficients for the soil liquid and a free liquid are related by (Epstein, 1989)

$$D_{dif} = \frac{D_o}{(L_{dif}/L)^2} \equiv \frac{D_o}{\tau^2} \equiv D_o \tau_o, \tag{7.5}$$

where

  L_{dif} and L are the actual and the shortest path lengths for diffusion (L)

  $\tau = L_{dif}/L$ is known as the tortuosity

  τ^2 is the tortuosity factor, while $\tau_a = (L/L_{dif})^2$ is often designated as an apparent tortuosity factor

The tortuosity L_{dif}/L is squared in Equation 7.5 because the concentration gradient along the streamline diminishes and the travel distance increases in the soil solution compared with diffusion along a straight path with length L in a free solution (Olsen and Kemper, 1968). It should be noted that the terms tortuosity and tortuosity factor have not been used consistently in the literature. Furthermore, some authors include the water content in their definition of tortuosity (Dykhuizen and Casey, 1989) or solute adsorption (retardation) in the expression for either τ or D_{dif} (Nye, 1979; Robin et al., 1987).

For unsaturated conditions, it is convenient to quantify the dependency of the diffusion coefficient on water content. Assuming that the tortuosity affects diffusion in the liquid phase in the same way as in the gaseous phase, the tortuosity term previously derived for gaseous diffusion in soils by Millington (1959) and Millington and Quirk (1961) can be adapted to describe aqueous diffusion in variably saturated soils. The following expressions then result:

$$D_{dif} = \frac{\theta^{10/3}}{\varepsilon^2} D_o, \tag{7.6a}$$

$$D_{dif} = \frac{\theta^2}{\varepsilon^{2/3}} D_o, \tag{7.6b}$$

where ε is the soil porosity ($L^3 L^{-3}$). An equivalent of the first expression, using a volumetric air content instead of θ, has been used frequently to describe gaseous diffusion in soils although Jin and Jury (1996) reported that the lesser known second version provided a better description.

Diffusion coefficients in soil systems are usually determined by mathematically analyzing solute concentration profiles in the soil as a function of time or position. Results and procedures for diffusion in soils have been summarized by Flury and Gimmi (2002). van Rees et al. (1991) measured diffusivities by allowing diffusion from a spiked solution or soil sample into a solute-free medium. Typically, two blocks of soil with different concentrations are brought together at time $t = 0$ (Kemper, 1986; Oscarson et al., 1992). After sufficient time has elapsed for solute diffusion to occur from the block with the higher to the lower concentration, the joined soil blocks are sectioned. The solute concentration of each section is determined, for example, by using extraction, centrifugation, and chemical analysis of the supernatant liquid. This approach yields a concentration profile versus distance from which the diffusion coefficient may be estimated using an appropriate analytical solution of the governing solute diffusion equation. Consider the diffusion equation,

$$\frac{\partial C}{\partial t} = D_{dif} \frac{\partial^2 C}{\partial x^2}, \tag{7.7}$$

subject to the initial condition,

$$C(x,0) = \begin{cases} C_o & -\infty < x < 0 \\ C_i & 0 < x < \infty \end{cases}, \tag{7.8}$$

and the approximate boundary conditions

$$C(-\infty,t) = C_o, \quad C(\infty,t) = C_i. \tag{7.9}$$

The solution for this problem is given by (Crank, 1975)

$$C(x,t) = C_i + 0.5(C_o - C_i)\operatorname{erfc}\left(\frac{x}{\sqrt{4D_{dif}t}}\right), \tag{7.10}$$

where erfc is the complementary error function (Gautschi, 1964).

7.2.1.2 Dispersion

Local variations in water flow in a porous medium will lead to mechanical dispersion. Several mechanisms that are commonly used to contribute to mechanical dispersion are illustrated with hypothetical tracer particles in Figure 7.1. Dispersion may occur because of (a) the development of a velocity profile within an individual pore such that the highest velocity occurs in the center of the pore and presumably little or no flow at the pore walls; (b) different mean flow velocities in pores of different sizes; (c) the mean water flow direction in the porous medium being different from the actual streamlines within individual pores, which differ in shape, size, and orientation; and (d) solute particles converging to or diverging from the same pore. All these processes contribute to solute spreading, in which initially steep concentration fronts become smoother during movement along the main flow direction.

Taylor (1953) analyzes dispersion of a solute injected in a cylindrical tube with laminar flow; this study is often used to elucidate dispersion in porous media. For laminar flow in a cylindrical tube filled with water, the parabolic velocity profile is

$$q(r) = q_o\left(1 - \frac{r^2}{r_o^2}\right), \tag{7.11}$$

where

q_o is the maximum velocity at the center of the tube ($L\ T^{-1}$)
r is the radial distance from the axis (L)
r_o is the tube radius (L)

The mean velocity over a cross section of the tube equals $q_o/2$. If a solute is injected in the tube, the velocity profile tends to stretch its concentration profile while diffusion will mitigate the large (transverse) concentration gradients caused by the velocity distribution. Taylor skillfully analyzed the governing two-dimensional transport problem. He obtained the following dispersion coefficient assuming proportionality between the longitudinal solute flux and the concentration gradient (i.e., Fick's law is applicable):

$$D_{dis} = \frac{r_o^2 v_o^2}{192 D_o}. \tag{7.12}$$
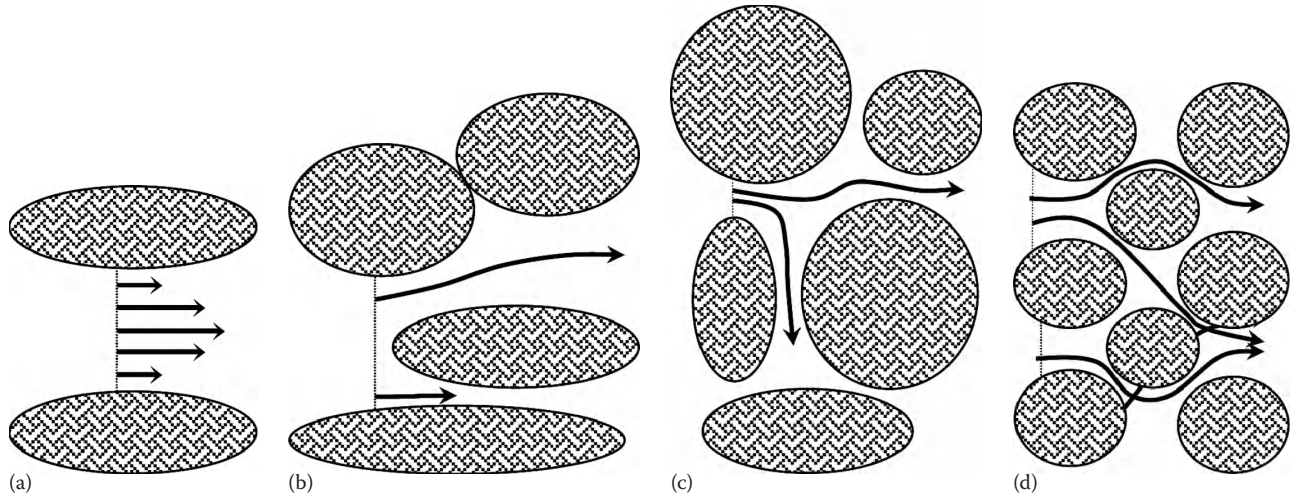
FIGURE 7.1   Schematic concepts contributing to mechanical dispersion.

Note that D_{dis} (L^2 T^{-1}) is inversely proportional to the coefficient of molecular diffusion D_o after sufficient time has elapsed for equilibrium to be established. The results by Taylor were later derived for more general conditions by Aris (1956) using moment analysis.

The Taylor–Aris model has been widely applied to describe dispersion in porous media. It was adapted early on in soil physics by Nielsen and Biggar (1962). Research to further elucidate dispersion phenomena in porous media has continued along the lines of the original theory (e.g., Gupta and Bhattacharya, 1983). At the early stage, dispersion in soils cannot be described as a Fickian process, and other models may need to be employed (Jury and Roth, 1990).

The one-dimensional solute flux due to mechanical dispersion in a uniform isotropic soil may be approximated using Fick's law:

$$J_{dis} = -\theta D_{dis} \frac{\partial C}{\partial x}, \tag{7.13}$$

where J_{dis} is the dispersive solute flux (M L^{-2} T^{-1}). Because of the similarity between Equations 7.4 and 7.13, solute spreading during transport in soils is typically described with a single coefficient of hydrodynamic dispersion, that is, $D = D_{dis} + D_{dif}$. As is often noted, this is done for convenience in spite of the conceptual differences between the diffusion and dispersion mechanisms (Scheidegger, 1974). The practice is consistent with results from laboratory and field experiments, which do not permit a distinction between mechanical dispersion and molecular diffusion. The hydrodynamic dispersive flux (J_D) (Equation 7.2) consists of contributions from molecular diffusion (Equation 7.4) and mechanical dispersion (Equation 7.13):

$$J_D = J_{dif} + J_{dis}. \tag{7.14}$$

Since diffusion is independent of flow, the contribution of diffusion to hydrodynamic dispersion diminishes if the soil water flow

rate increases. Hydrodynamic dispersion is often simply referred to as dispersion, as will be done in the remainder of this chapter.

The effects of dispersion can be illustrated with a hypothetical laboratory experiment in which water and a dissolved tracer are applied to an initially tracer-free, uniformly packed soil column of length $L = 1$ m (Figure 7.2a). The column is subjected to steady-state water flow with uniform water content. The average pore water or interstitial velocity $v = 0.5$ m day^{-1}. Note that v is the actual travel-time velocity, which is given per unit area of fluid. It is determined by the ratio of the Darcian water flux density (J_w) and the volumetric water content (θ), where J_w is defined per unit area of (bulk) soil. There is a step change in solute concentration at the inlet (Figure 7.2b), and the solute front becomes smoother because of solute spreading during passage of the soil column (Figure 7.2c). A smooth effluent curve can be monitored at the column exit as shown in Figure 7.2d. In the absence of diffusion and dispersion ($D = 0$), the front of a perfectly inert tracer will travel as a square wave through the column (a process often called piston flow) to reach the bottom of the column at time $t = L/v$. For piston flow, a conservative tracer reaches the column exit exactly after one pore volume of tracer solution has been injected (or collected at the column exit). Pore volume is defined as the volume of water stored in that column.

The degree of spreading is usually related to the solute travel time, although some constraints do exist on the amount of spreading. Dispersion, as sketched in Figure 7.1a, is limited because of transverse molecular diffusion, which causes solutes to move from the center of a tube to areas near the pore walls, or vice versa, in response to local concentration gradients. Such transverse diffusion counteracts spreading caused by variations in the longitudinal flow velocity. Dispersion is also limited since capillaries in a porous medium generally are not independent cylindrical tubes, but branch and join or rejoin each other at distances characteristic of the pore or particle size distribution of the medium. This branching and rejoining promote lateral mixing of solutes from different pores as sketched in Figure 7.1d.
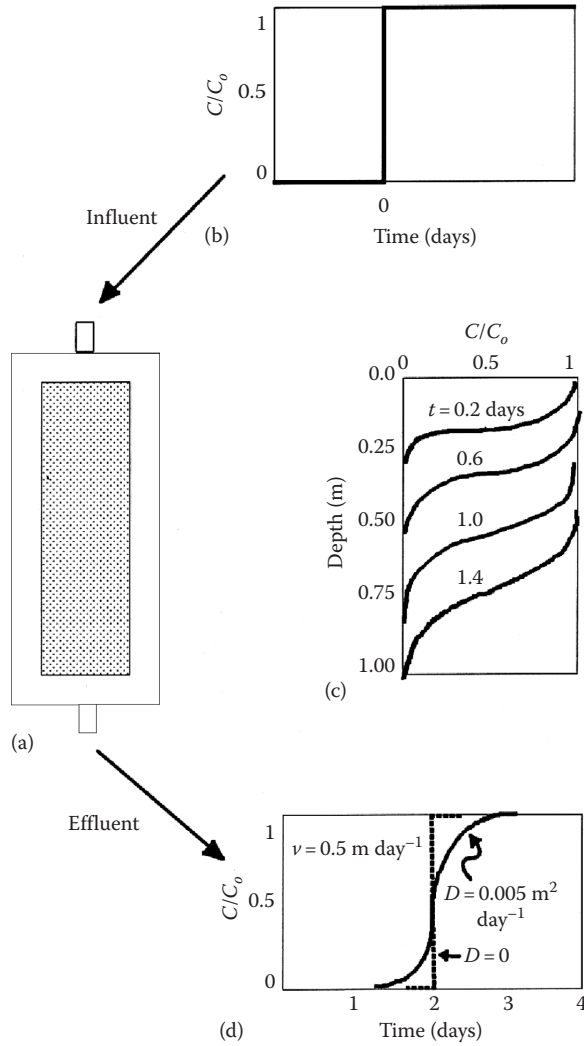
FIGURE 7.2 Hypothetical laboratory tracer experiment: (a) column of soil, (b) influent curve, (c) concentration distributions inside the column, and (d) breakthrough curves with and without dispersion. (Modified after van Genuchten, M.Th. 1988. Solute transport, p. 360–362. *In* S.P. Parker (ed.) McGraw-Hill yearbook of science and technology. McGraw-Hill Book Co., New York.)

Three-dimensional dispersion is quantified with a dispersion tensor. The components of the symmetric dispersion tensor for an isotropic soil are given as (Bear and Verruijt, 1987)

$$D_{ij} = \delta_{ij}\alpha_T \, |\vec{v}| + \frac{(\alpha_L + \alpha_T)v_i v_j}{|\vec{v}|}, \qquad (7.15)$$

where
  $|\vec{v}|$ denotes the magnitude of the pore water velocity with v_i as the ith component (L T^{-1})
  δ_{ij} is the Kronecker delta ($\delta_{ij} = 1$, if $i = j$ and $\delta_{ij} = 0$, if $i \neq j$)
  α_L and α_T are, respectively, the longitudinal and transverse dispersivity (L)

For a one-dimensional system, Equation 7.15 reduces to $D_{dis} = \alpha_L|v|$. Mechanical dispersion, as sketched in an idealized fashion in Figure 7.1a, can be reversed by changing the flow direction to make a smooth front steep again. In soils, however, dispersion is not reversible since mixing erases antecedent concentration distributions, as illustrated in Figure 7.1d. Absolute values are, therefore, used for v in Equation 7.15. In case of uniform water flow parallel to the x-axis of a Cartesian coordinate system, only the following three main components of Equation 7.15 need to be considered:

$$D_{xx} = \alpha_L v, \quad D_{yy} = \alpha_T v, \quad D_{zz} = \alpha_T v, \qquad (7.16)$$

where
  D_{xx} is the coefficient of longitudinal (mechanical) dispersion
  D_{yy} and D_{zz} are the coefficients of transverse dispersion

Dispersion coefficients may be determined by fitting a mathematical solution to observed concentration (Toride et al., 1995). Additional procedures to determine dispersion coefficients are given by Fried and Combarnous (1971) and van Genuchten and Wierenga (1986). Values of the longitudinal dispersivity (α_L) for laboratory experiments typically vary between 0.1 and 10 cm with 6–20 times smaller values for α_T (Klotz et al., 1980). Dispersivities for field soils are generally much higher. Gelhar et al. (1992) reviewed published field-scale dispersivities determined in aquifer materials that are typically one or two orders of magnitudes larger, even more so for relatively large experimental scales.

7.2.2 Advection–Dispersion Equation

The expressions for the advective and dispersive solute fluxes can be substituted in mass balance Equation 7.1. The one-dimensional ADE for solute transport in a homogeneous soil becomes

$$\frac{\partial(\theta C)}{\partial t} = -\frac{\partial}{\partial x}\left(J_w C - \theta D\frac{\partial C}{\partial x}\right) + \theta R_s. \qquad (7.17)$$

In the case where the water content is invariant with time and space, the ADE may be simplified to ($v = J_w/\theta$),

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x} + R_s. \qquad (7.18)$$

This is a parabolic differential equation similar to the diffusion equation. To complete the mathematical formulation of the transport, several concentration types and mathematical conditions will be reviewed in Section 7.3.1.

A variety of solute source or sink terms may be substituted for R_s. The most common source/sink term is due to adsorption/desorption and ion exchange stemming from chemical and physical interactions between the solute and the soil solid phase.

Many other processes such as radioactive decay, aerobic and anaerobic transformations, volatilization, photolysis, precipitation/dissolution, reduction/oxidation, and complexation may also affect the solute concentration. A further refinement of the transport model is necessary in the case of nonuniform interactions between the solute and the soil, or if there is adsorption on moving particles and colloids. In the following, only interactions at the solid–liquid interface will be considered.

7.2.3 Adsorption

Dissolved substances in the liquid phase can interact with several soil constituents such as primary minerals, oxides, and inorganic or organic colloids. Dissolved ions in the soil solution counterbalance the surface charge of soil particles caused by isomorphous substitution of one element for another in the crystal lattice of clay minerals, by the presence of hydronium or hydroxyl ions at the solid surface, or other mechanisms. The net surface charge of an assemblage of soil particles produces an electric field that affects the distribution of cations and anions within water films surrounding the soil particles.

Adsorption of solute (adsorbate) by the soil (adsorbent) is an important phenomenon affecting the fate and movement of solutes. The ADE for one-dimensional transport of an adsorbed solute may be written as

$$\frac{\partial C}{\partial t} + \frac{\rho_b}{\theta}\frac{\partial S}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x}, \tag{7.19}$$

where S is the adsorbed concentration, defined as mass of solute per mass of dry soil ($M\ M^{-1}$). The above equation can be expressed in terms of one dependent variable by assuming a suitable relationship between the adsorbed and liquid concentrations. This is typically done with a simple adsorption isotherm to quantify the adsorbed concentration as a function of the liquid concentration at a constant temperature. In addition to temperature, the adsorption isotherm is generally also affected by the solution composition, total concentration, the pH of the bulk solution, and sometimes the method used for measuring the isotherm. A mathematically pertinent distinction is often made between linear and nonlinear adsorption. Although most adsorption isotherms are nonlinear, the adsorption process may often be assumed linear for low solute concentrations or narrow concentration ranges.

7.2.3.1 Linear Adsorption

Consider the general case of nonequilibrium adsorption, where a change in C is accompanied by a delayed change in S. The adsorption rate can be described assuming first-order kinetics:

$$\frac{\partial S}{\partial t} = kh(C,S), \tag{7.20}$$

where
  k is a rate parameter (T^{-1})
  h is a function to quantify how far the adsorption or desorption process is removed from equilibrium

A single-valued isotherm for equilibrium adsorption $\Gamma(C)$ is used to define $h(C, S)$ according to

$$h(C,S) = \Gamma(C) - S. \tag{7.21}$$

For equilibrium adsorption $k \to \infty$, and hence $h(C, S) \to 0$, which implies that $S = \Gamma(X)$. For a linear adsorption isotherm, the relation between Γ and C can simply be given as

$$\Gamma = K_d C, \tag{7.22}$$

where K_d is a partition coefficient, often referred to as the distribution coefficient, expressed in volume of solvent per mass of soil ($L^3\ M^{-1}$). For $S = \Gamma$, substitution of Equation 7.22 into Equation 7.19 leads to the following ADE commonly used to describe transport of a solute that undergoes linear equilibrium exchange:

$$R\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x}, \tag{7.23}$$

in which the retardation factor R is given by

$$R = 1 + \frac{\rho_b}{\theta}K_d, \tag{7.24}$$

with ρ_b as soil bulk density. The advective and dispersive fluxes are reduced by a factor R as a result of adsorption. The movement of the solute is said to be retarded with respect to the average solvent movement. If there is no interaction between the solute and the soil ($K_d = 0$), the value for R is equal to unity. The value for R may be calculated from K_d as obtained from chemical analyses of the solution and adsorbed phases. Frequently, R is obtained by optimizing a mathematical solution to observed concentrations. For step application of a solute to a soil column (cf. Figure 7.2), the net solute flux into the column should be equal to the change in the amount of solute in the soil column. The following mass balance can hence be formulated to estimate R:

$$v\big[g(t) - C_e\big] = R\int_0^L \big[C(x,t) - f(x)\big]\,dx, \tag{7.25}$$

where
  C_e is the effluent concentration
  $f(x)$ and $g(t)$ are the initial and influent concentrations, respectively

The effects of linear adsorption on solute transport in a homogeneous soil profile are shown in Figure 7.3. Analytically predicted solution and adsorbed concentrations are plotted 4 days after the start of a 1-day application of influent with a unit solute concentration (units may be selected arbitrarily, e.g., g m^{-3}) to an initially solute-free soil profile subject to steady saturated water flow. Other parameters for this example are $J_w = 10$ cm day^{-1}, $\theta = 0.40$ cm^3 cm^{-3}, and $D = 62.5$ cm^2 day^{-1}. The pore water velocity ($v = J_w/\theta$) is hence 25 cm day^{-1} and $\alpha_L = 2.5$ cm. Solute distributions are plotted for three values of the retardation factor, R (Figure 7.3).
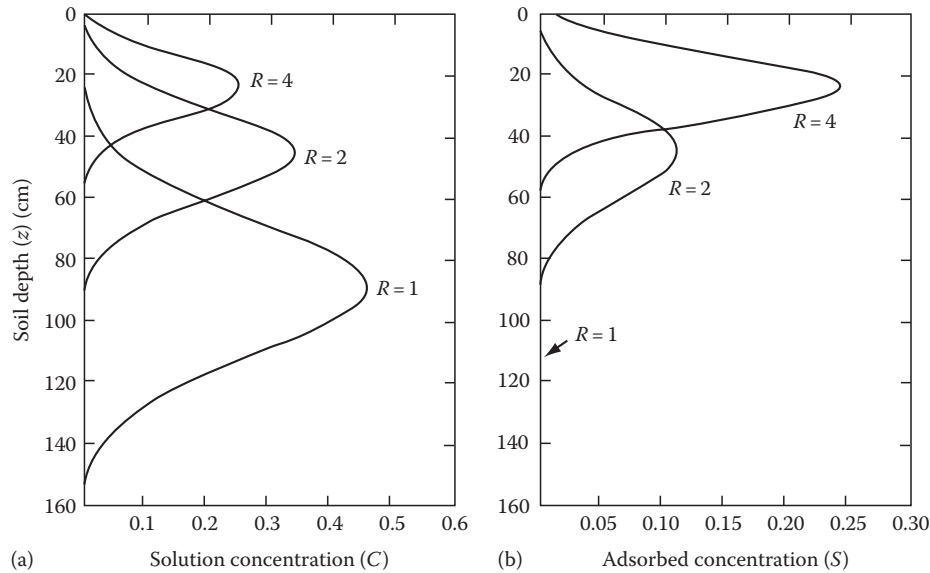
FIGURE 7.3 Effect of adsorption, as accounted for by the retardation factor R, on (a) solution (C) and (b) adsorbed (S) concentration distributions in a homogeneous soil profile.

When R is increased from 1.0 to 2.0, the apparent solute velocity (v/R) is reduced by one-half (Figure 7.3a), causing a shallower penetration of the solute pulse into the profile. At the same time, the area under the curve in Figure 7.3a is also reduced by one-half. When $R = 4$, the apparent solute velocity and the area under the curve are again reduced by half. The distribution for the adsorbed concentration S, which may be expressed as mg kg^{-1} soil, increases from 0 (no adsorption) when $R = 1$ to a maximum when $R = 4$ are similar (Figure 7.3b). Assuming a soil bulk density (ρ_b) of 1.25 g cm^{-3} and the same water content as before ($\theta = 0.40$ cm^3 cm^{-3}), one may calculate, using Equation 7.24, that the distribution coefficient $K_d = 0$, 0.32, and 0.96 cm^3 g^{-1} for $R = 1$, 2, and 4, respectively.

Anion exclusion occurs when negatively charged surfaces of clays and ionizable organic matter are present; anions are repelled from such surfaces and accumulate in the center of pores. Because water flow velocities are 0 at pore walls and maximum in the center of pores (Figure 7.1a), the average anion movement will be faster than the average water movement. Many displacement experiments also suggest faster anion than water movement simply because the apparent displacement volume is smaller for anions than water. The quantity $(1 - R)$ is the relative anion exclusion volume. The exclusion volume-per-unit mass of soil can also be estimated as

$$V_{ex} = \int \left(1 - \frac{c}{C_o} \right) dV \qquad (7.26)$$

where
V_{ex} is the exclusion volume (L^3 M^{-1})
c is the local concentration of the anion (M L^{-3})
C_o its bulk concentration (M L^{-3})
V is the entire volume encompassing the liquid phase per unit mass of soil

Instead of using $R < 1$, anion transport may be modeled with $R = 1$, which restricts the accessible liquid volume (Krupp et al., 1972).

Anions are also adsorbed by the soil through surface complexation and adsorption onto positively charged areas of the solid matrix. If the effect of adsorption exceeds exclusion, the anion will be retarded. The retardation factor should be viewed as an effective parameter since it quantifies a variety of adsorption and exclusion processes to which the solute (anion) is subjected.

Breakthrough curves typical for the transport of an excluded anion (Cl$^-$), a nonreactive solute (tritiated water, ^{3}H$_2$O), and an adsorbed cation (Ca^{2+}) are presented in Figure 7.4. The first two tracers pertain to transport through 30 cm columns containing disturbed Glendale clay loam soil (P.J. Wierenga, personal
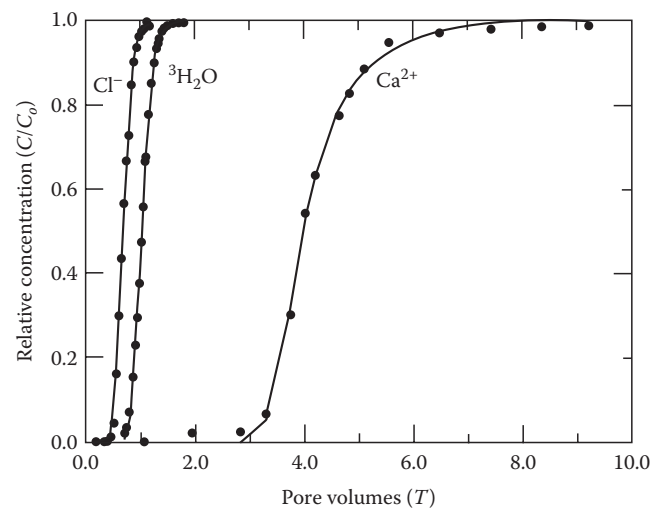
FIGURE 7.4 Observed and fitted ADE breakthrough curves for three tracers typifying the transport of anions (Cl$^-$), a (nearly) nonreactive solute (^{3}H$_2$O), and an adsorbing solute (Ca^{2+}).

communication; van Genuchten and Cleary, 1982), while the Ca^{2+} data are for transport through a 30 cm long column containing a Troup loam and a Savannah fine loam (Leij and Dane, 1989). Optimizing the solutions of the ADE to the three breakthrough curves yielded R values of 0.681, 1.027, and 4.120 for Cl^-, ^3H_2O, and Ca^{2+}. Hence, the Cl^- curve was strongly affected by anion exclusion, while ^3H_2O transport was subject to very minor adsorption/exchange.

7.2.3.2 Nonlinear Adsorption

In many cases, adsorption, and hence, the retardation factor, cannot be described using a simple K_d approach. For nonlinear equilibrium adsorption, R is given as

$$R(C) = 1 + \frac{\rho_b}{\theta}\frac{\partial \Gamma}{\partial C}. \tag{7.27}$$

Two common nonlinear adsorption isotherms are the Langmuir and Freundlich equations:

$$\Gamma = \frac{k_1 C}{1 + k_2 C} \quad \text{(Langmuir)}, \tag{7.28}$$

$$\Gamma = k_3 C^n \quad \text{(Freundlich)}, \tag{7.29}$$

where k_1, k_2, k_3, and n are empirical constants. Many other equations for adsorption exist, including some that account for differences between adsorption and desorption isotherms (van Genuchten and Sudicky, 1999).

The Freundlich isotherm will be used in the following to illustrate the effects of nonlinear equilibrium adsorption on solute transport. In order to keep the calculations simple, the value of k_3 in Equation 7.29 is taken to be 0.64. Three different values of the exponent n are used, viz., 0.5, 1.0, and 1.5, to demonstrate favorable, linear, and unfavorable adsorption isotherms (Figure 7.5). Calculated distributions of the solution (C) and adsorbed (S) concentrations versus soil depth (z) 8 days after application of a 4-day long solute pulse to the soil surface are shown in Figure 7.6. The same pore water velocity is used as for the example illustrated in Figure 7.4 but with a smaller dispersion coefficient of $D = 25\,cm^2\,day^{-1}$ ($\alpha_L = 1$ cm). Notice that, as in Figure 7.4, the solution concentration distribution for $n = 1$ (linear adsorption) has a nearly symmetrical shape versus depth. The other two n values yield nonsymmetric profiles.

If $n = 0.5$, a very sharp concentration front develops, while the curve near the soil surface becomes more dispersed. The sharp front can be explained by considering the retardation factor (R) for nonlinear adsorption (Equation 7.27), which for $n = 0.5$, $\rho_b = 1.25\,g\,cm^{-3}$, $\theta = 0.40$, and $k_3 = 0.64$ leads to $R = 1 + 1/C$. This shows that R increases rapidly when C decreases with the extreme $R \to \infty$, if $C = 0$. Consequently, the apparent solute velocity $v_a = v/R$ is very small at the lower liquid concentrations but increases at higher values. Of course, higher concentrations
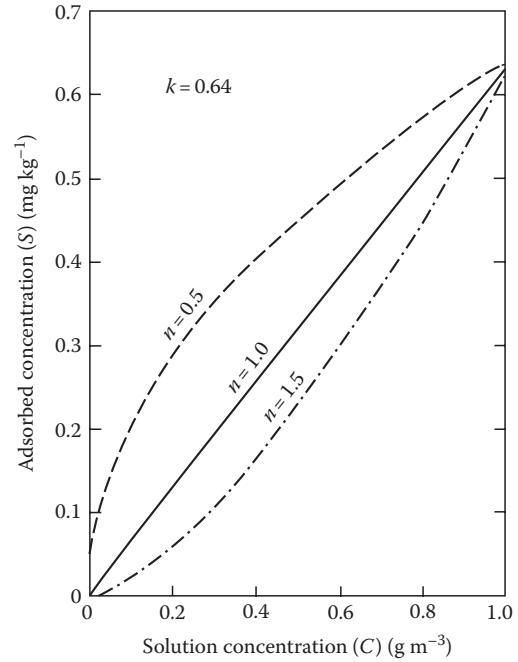
FIGURE 7.5 Freundlich equilibrium plots for $k_3 = 0.64$ and three values of the exponent n.
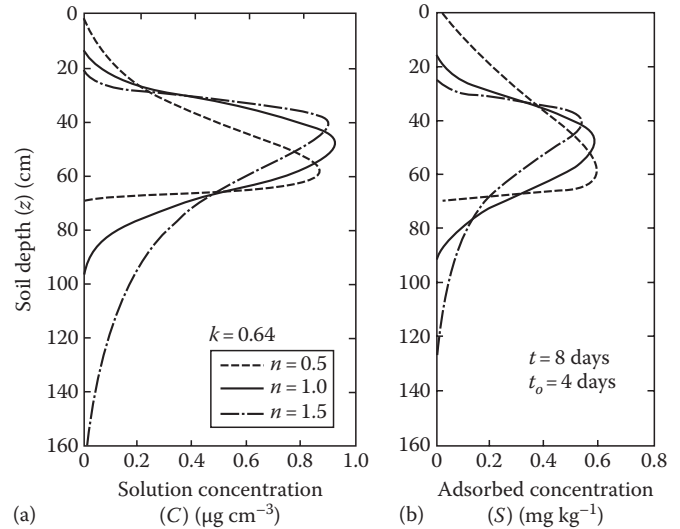
FIGURE 7.6 Effect of nonlinear adsorption on (a) solution (C) and (b) adsorbed (S) concentrations in a deep homogeneous soil profile during steady-state flow. The distributions were obtained for the three isotherms shown in Figure 7.5.

cannot move faster than lower concentrations; front sharpening will lead to a steep solute front. This front never becomes a step function because the large concentration gradient across the front will create a large diffusion/dispersion flux. If $n \ll 1$, the front may be determined from the average slope of the isotherm between the initial and the maximum aqueous concentration. In this example, those concentrations are zero and approximately $1\,g\,m^{-3}$. The average slope of the isotherm is therefore exactly the same as the linear distribution coefficient ($d\Gamma/dC = 0.64$).

Substituting this value into Equation 7.27 yields $R = 3$. Hence, the apparent solute velocity (v_a) equals 25/3 or 8.33 cm day^{-1}, and the solute front after 8 days is located at a depth of about 67 cm (Figure 7.6). Transport of favorably adsorbed solutes is frequently modeled with traveling wave solutions (van der Zee, 1990; Simon et al., 1997).

A reverse scenario occurs if $n > 1$ (unfavorable exchange). Adsorption at the lower concentrations is now relatively small and, as displayed in Figure 7.6, the toe of the front moves through the profile at a velocity nearly equal to that of an inert solute. Adsorption at the higher concentrations, on the other hand, is much more extensive, resulting in a lower apparent solute velocity in the higher range of concentrations. As a result, the concentration front becomes increasingly dispersed over time. Ignoring dispersion, the velocity of the solute front ($v_a = v/R$) at any given value of C is given by

$$v_a = \frac{v}{1 + 3\sqrt{C}}, \tag{7.30}$$

while the depth of the solute front can be approximated by

$$z(C,t) = z(C,0) + v_a t, \tag{7.31}$$

where $z(C, 0)$ is the location of a solute concentration with value C at $t = 0$.

Cation exchange processes during solute transport will involve at least two species. The simplest case arises when two cations of the same valency and total concentration such as Ca^{2+} and Mg^{2+} are considered. The resulting exchange process is then approximately linear for relatively small changes in the composition of the soil solution, that is, a change in aqueous concentration will involve a commensurate change in adsorbed concentration. Exchange between Na^+ and Ca^{2+}, on the other hand, is considerably more nonlinear. Equations that quantify the exchange reaction have been proposed by Gapon, Kerr, Vanselow, Eriksson, and others. Typically, the soil solution will contain many different species, and their interactions with the solid soil phase will have to be simulated numerically.

7.2.4 Nonequilibrium Transport

Solute breakthrough curves for aggregated soils will exhibit asymmetrical distributions or nonsigmoidal concentration fronts. The difference in water flow in the inter- and intra-aggregate regions leads to nonequilibrium conditions for solute transport due to (transversal) gradients in the solute concentration between the regions. Depending upon the exact pore structure of the medium, asymmetry is sometimes enhanced by desaturation when the relative fraction of water residing in the marginally continuous immobile region increases. Physical nonequilibrium transport due to different flow regimes is often modeled by partitioning the liquid phase into two flow domains with diffusive solute transfer between them. Similarly, differences in sorption

of solutes by various domains of the solid phase may lead to chemical nonequilibrium transport. This may also be described with a two-domain model; sorption occurs instantaneous in the equilibrium domain and is a rate-limited process for the remainder of the sorption sites of the solid phase. It is not always possible to differentiate between physical and chemical nonequilibrium since many sorption sites are only accessible after diffusion through the immobile region of the liquid phase.

In the following, nonequilibrium transport will be described with a two-domain model with a separate liquid or sorbed concentration for each domain, that is, the two-region model for physical and the two-site model for chemical nonequilibrium. The equilibrium ADE (Equation 7.19) can be readily modified for this purpose. The same dimensionless mathematical formulation can be used for physical and chemical nonequilibrium models. If necessary, the ADE can be modified to incorporate additional nonequilibrium processes and continua.

7.2.4.1 Physical Nonequilibrium

Consider one-dimensional solute movement in an isotropic soil with uniform flow and transport properties during steady flow and assume that the solute is subject to linear retardation, that is, equilibrium sorption can be described with a linear exchange isotherm. The physical nonequilibrium approach is based on a partitioning of the liquid phase into a mobile or flowing region and an immobile or stagnant region. Solute movement in the mobile region occurs by both advection and dispersion, whereas solute exchange between the two regions occurs by first-order diffusion (Coats and Smith, 1964). Following van Genuchten and Wierenga (1976), the governing equations for the two region model are as follows:

$$(\theta_m + f\rho_b K_d)\frac{\partial C_m}{\partial t} = \theta_m D_m \frac{\partial^2 C_m}{\partial x^2} - \theta_m v_m \frac{\partial C_m}{\partial x} - \alpha(C_m - C_{im}), \tag{7.32}$$

$$\left[\theta_m + (1-f)\rho_b K_d\right]\frac{\partial C_{im}}{\partial t} = \alpha(C_m - C_{im}), \tag{7.33}$$

where
f represents the fraction of sorption sites in equilibrium with the fluid of the mobile region
α is first-order mass transfer coefficient (T^{-1})
subscripts m and im, respectively, refer to the mobile and immobile liquid regions (with $\theta = \theta_m + \theta_{im}$), while ρ_b and K_d are the soil bulk density and distribution coefficient for linear sorption

Transport Equation 7.32 follows directly from addition of a source/sink term (R_s) to Equation 7.19.

Anion exclusion can be viewed as a particular example of physical nonequilibrium, the exclusion volume roughly corresponds to the immobile region (Krupp et al., 1972). The physical

nonequilibrium concept may, therefore, be adapted to describe transport of excluded anions (van Genuchten, 1981) instead of using a retardation factor of less than 1.

7.2.4.2 Chemical Nonequilibrium

Sorption of solute, especially for organic chemicals, has often been described with a combined equilibrium and kinetic sorption expression so as to better simulate transport in soils with a wide variety of soil constituents (clay minerals, organic matter, and oxides). The lack of an instantaneous equilibrium for the sorption process is sometimes referred to as chemical nonequilibrium. This terminology is somewhat misleading since the rate of adsorption or exchange is usually determined mostly by physical phenomena such as diffusion through the liquid film around soil particles and inside the aggregates (Boyd et al., 1947; Sparks, 1989).

The simplest and by far most popular approach distinguishes between type 1 sites, with instantaneous adsorption, and type 2 sites, where adsorption obeys a kinetic rate law (Selim et al., 1976). In the case of first-order kinetics, the general adsorption rates can be given with a model similar to Equations 7.20 and 7.21 as

$$\frac{\partial S_1}{\partial t} = \alpha_1 \left[\Gamma_1(C) - S_1 \right], \tag{7.34}$$

$$\frac{\partial S_2}{\partial t} = \alpha_2 \left[\Gamma_2(C) - S_2 \right], \tag{7.35}$$

where

α is again a rate constant (T^{-1})

S is the actual adsorbed concentration (M M^{-1})

Γ is the final adsorbed concentration at equilibrium as prescribed by the adsorption isotherm

subscripts 1 and 2 refer to the type of adsorption site $\Gamma_1 + \Gamma_2 = \Gamma$

Because type 1 sites are always at equilibrium, $S_1 = \Gamma_1$ and Equation 7.34 can further be ignored. The transport equation becomes

$$\frac{\partial C}{\partial t} + \frac{\rho_b}{\theta}\frac{\partial \Gamma}{\partial t} + \frac{\alpha_2 \rho_b}{\theta}(\Gamma_2 - S_2) = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x}. \tag{7.36}$$

If the fraction of exchange sites that is at equilibrium (type 1) equals f and if equilibrium adsorption is governed by the same linear isotherm for both types 1 and 2 ($\Gamma_1 = \Gamma_2$) then

$$\Gamma_1 + \Gamma_2 = fK_dC + (1-f)K_dC. \tag{7.37}$$

Of course, nonlinear equilibrium isotherms may also be used in nonequilibrium transport models. The complete transport problem can now be written as

$$\left(1 + \frac{\rho_b fK_d}{\theta}\right)\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x} - \frac{\alpha\rho_b}{\theta}\left[(1-f)K_dC - S_2\right], \tag{7.38}$$

$$\frac{\partial S_2}{\partial t} = \alpha\left[(1-f)K_dC - S_2\right], \tag{7.39}$$

where the subscript for α has been dropped. This two-site chemical nonequilibrium model reduces to a one-site kinetic nonequilibrium model by setting $f = 0$. The two-site chemical nonequilibrium model was applied successfully to describe solute breakthrough curves by Selim et al. (1976), van Genuchten (1981), and Nkedi-Kizza et al. (1983), among others.

7.2.4.3 General Nonequilibrium Formulation

The two-site and the two-region nonequilibrium models can be cast in the same (dimensionless) model according to Nkedi-Kizza et al. (1984):

$$\beta R\frac{\partial C_1}{\partial T} = \frac{1}{P}\frac{\partial^2 C_1}{\partial X^2} - \frac{\partial C_1}{\partial X} + \omega(C_2 - C_1), \tag{7.40}$$

$$(1-\beta)R\frac{\partial C_2}{\partial T} = \omega(C_1 - C_2), \tag{7.41}$$

where

β is a partition coefficient

R is a retardation factor

C_1 and C_2 are dimensionless equilibrium and nonequilibrium concentrations

T is time

X is distance

P is the column Peclet number

ω is a mass transfer coefficient

subscripts 1 and 2 refer to the equilibrium and nonequilibrium phases, respectively

The common dimensionless parameters are defined using an arbitrary characteristic concentration (C_o) and length (L):

$$T = \frac{vt}{L}, \quad X = \frac{x}{L}, \quad P = \frac{vL}{D}, \quad R = 1 + \frac{\rho_b K_d}{\theta}. \tag{7.42}$$

For the physical nonequilibrium model, the remaining dimensionless parameters are

$$\beta = \frac{\theta_m + f\rho_b K_d}{\theta + \rho_b K_d}, \quad \omega = \frac{\alpha L}{\theta v}, \quad C_1 = \frac{C_m}{C_o}, \quad C_2 = \frac{C_{im}}{C_o}, \tag{7.43}$$

whereas for the chemical nonequilibrium model:

$$\beta = \frac{\theta_m + f\rho_b K_d}{\theta + \rho_b K_d}, \quad \omega = \frac{\alpha(1-\beta)RL}{v}, \quad C_1 = \frac{C}{C_o}, \quad C_2 = \frac{S_2}{(1-f)K_dC_o}. \tag{7.44}$$

In the chemical engineering literature, $\alpha L/v$ is known as the Damköhler number; it quantifies the rate of the reaction or exchange relative to advective transport.

7.3 Solutions of the Advection–Dispersion Equation

The research and management of solute behavior in soils almost invariably require that the temporal and spatial solute distribution be known. Solute distributions as a function of time and/or space can be estimated with a variety of analytical and numerical solutions of the ADE, some of which will be briefly reviewed in the following.

7.3.1 Basic Concepts

A complete mathematical formulation of the transport problem requires that the pertinent-dependent variable or concentration type is used and that the proper auxiliary conditions are specified.

7.3.1.1 Concentration Types

Concentration is conventionally defined as the amount of solute-per-unit volume of the liquid. Since microscopic concentrations are based on a relatively small scale, a concentration at a larger scale needs to be introduced to allow use of the ADE, which is based on larger macroscopic variables and parameters. For this purpose, a macroscopic resident or volume-averaged concentration (C_R) is defined as

$$C_R = \frac{1}{\Delta V} \iiint c \, dV, \tag{7.45}$$

where

c is the variable local-scale (microscopic) concentration (M L^{-3})
ΔV is some representative elementary volume (Bear and Verruijt, 1987)

A different concentration type may be encountered at soil boundaries where the concentration is implicitly determined or defined as the ratio of the solute flux (J_s) and water flux (J_w) densities,

$$C_F = \frac{J_s}{J_w}, \tag{7.46}$$

with C_F as the flux-averaged concentration. This concentration represents the mass of solute-per-unit volume of fluid passing through a soil cross section during an elementary time interval (Kreft and Zuber, 1978). For a one-dimensional solute flux consisting of an advective and a dispersive component, the flux-averaged concentration can be derived from the resident concentration according to the transformation:

$$C_F = C_R - \frac{D}{v} \frac{\partial C_R}{\partial x}. \tag{7.47}$$

Conversely, the resident concentration may also be determined from the flux-averaged concentration. For solute application to an initially solute-free medium, the transformation is (van Genuchten et al., 1984)

$$C_R(x,t) = \frac{v}{D} \exp\left(\frac{vx}{D}\right) \int_x^\infty \exp\left(-\frac{v\xi}{D}\right) C_F(\xi,t) d\xi. \tag{7.48}$$

The difference between C_R and C_F is usually small, except when the second term on the right-hand side of Equation 7.46 is relatively large. It should be noted that a distinction between flux and resident type can be made for both the application and the detection of solutes (Kreft and Zuber, 1978, 1986). Resident concentrations are used for solute detection with, for example, time domain reflectometry, and to specify most initial conditions. Flux-averaged concentrations, on the other hand, are used for effluent samples, and to specify the influent concentration in most boundary value problems. Unless stated otherwise, it is assumed that solute concentrations are of the resident type.

Averaged concentrations can also be defined in terms of the observation scale, the latter exceeding the macroscopic scale associated with using the ADE. A time-averaged concentration (C_t) is obtained by averaging over a time interval (Δt) about a discrete time (t_o) (Fischer et al., 1979):

$$C_t(x,t_o) = \frac{1}{\Delta t} \int_{t_o - \Delta t/2}^{t_o + \Delta t/2} C(x,t) dt, \tag{7.49}$$

where C is a continuous solution of the solute concentration, which can be obtained by solving the ADE. This type of concentration occurs if solute breakthrough curves are measured using, for example, fraction collectors or gamma-ray attenuation. Similarly, a one-dimensional spatial average can be defined as

$$C_x(x_o,t) = \frac{1}{\Delta x} \int_{x_o - \Delta x/2}^{x_o + \Delta x/2} C(x,t) dx. \tag{7.50}$$

This concentration may be used to describe experimental results obtained for samples with centroid (x_o) and length (Δx). This situation occurs, for example, when the measured concentration of a large core sample is to be modeled as a point value (Leij and Toride, 1995).

7.3.1.2 Boundary and Initial Conditions

Initial and boundary conditions need to be specified in order to obtain a meaningful solution of the ADE. For a finite or semi-infinite soil, the initial condition can be formulated as

$$C(x,0) = f(x) \quad x \geq 0, \tag{7.51}$$

where $f(x)$ is an arbitrary function. Initial concentrations are almost invariably of the resident type.

The selection of the most appropriate boundary conditions for a transport problem is a somewhat esoteric topic that has received considerable attention in the literature. This is partly due to a lack of detailed experimental information for evaluating and applying boundary conditions, and inherent shortcomings of the transport equation itself at boundaries. Many transport problems involve the application to the soil of a solute, whose influent concentration may be described by a function $g(t)$. The application method may be pumping, ponding, or sprinkling. Two different types of inlet conditions are used, which assume either continuity in solute concentration or solute flux density. Simultaneous use of both conditions is seldom possible. It is generally more desirable to ensure mass conservation in the whole system than a continuous concentration at the inlet. The solute fluxes at the inlet boundary are, therefore, equated to obtain the following third or flux-type inlet condition:

$$\left(vC - D\frac{\partial C}{\partial x} \right)_{x=0^+} = vg(t) \quad t \geq 0, \tag{7.52}$$

where 0^+ indicates a position just inside the soil. It is assumed that there is no dispersion outside the soil. The alternative condition requires the concentration to be continuous across the interface at all times. At smaller scales, such continuity will likely exist. However, at the scale of the ADE, it appears difficult to maintain a constant concentration at the interface, particularly during the initial stages of solute displacement for low influent fluxes and high dispersive fluxes in the soil. Mathematically, the first or concentration-type condition is expressed as

$$C(0,t) = g(t) \quad t \geq 0. \tag{7.53}$$

The outlet condition can be defined as a zero gradient at a finite or infinite distance from the inlet. The infinite outlet condition,

$$\frac{\partial C}{\partial x}(\infty,t) = 0 \quad t \geq 0, \tag{7.54}$$

is more convenient for mathematical solutions than the finite condition,

$$\frac{\partial C}{\partial x}(L^-,t) = 0 \quad t \geq 0. \tag{7.55}$$

The use of Equation 7.54 implies that there is a semi-infinite fictitious soil layer beyond $x = L$, with identical properties as the actual soil. Such a layer does not affect the movement of the solute in the actual soil upstream of the exit boundary. Since the transport at the outlet cannot be precisely described, the intuitive contradiction of an infinite mathematical condition to describe a finite physical system is often more acceptable than using Equation 7.55, which precludes dispersion inside the soil near the outlet.

The formulation of the boundary and inlet conditions should account for the injection and detection modes in order to arrive at a mathematically consistent formulation of the problem with the same concentration type as independent variable (Parker and van Genuchten, 1984; Leij and Toride, 1995). Normally, observed or predicted resident (volume-averaged) concentrations require the use of a third-type inlet condition (also in terms of the resident concentration) whereas a first-type inlet condition is used in conjunction with flux-type concentrations.

7.3.2 Analytical Solutions

Analytical solutions can formally be obtained only for linear transport problems. It would appear that analytical solutions are not very useful for transport in field soils where there is (1) spatial and temporal variability of flow and transport parameters, (2) transient flow, especially for unsaturated soils, and (3) nonuniformity in the boundary and initial conditions. However, analytical solutions can still be quite valuable. A nonlinear transport problem may be linearized through a suitable transformation to obtain a problem for which an analytical solution is available. Also, analytical solutions provide quick estimates of solute behavior over large temporal and spatial scales while they may offer insight into the underlying transport processes. Moreover, there is usually a lack of input parameters for field problems, which diminishes the advantage of numerical over analytical model results. Analytical solutions are also routinely used to evaluate the performance of numerical schemes. Finally, the mathematical and physical conditions tend to be well defined for laboratory settings and an analytical solution can often be used, especially to estimate transport parameters by fitting analytical solutions to experimental data (Parker and van Genuchten, 1984; van Genuchten and Parker, 1987).

7.3.2.1 Variable Transformation

One straightforward way to obtain an analytical solution is to transform the ADE to an equation for which a solution already exists. As an example, consider transport in an infinite system given by

$$R\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x}, \tag{7.56}$$

$$C(x,0) = \begin{cases} C_o & x < 0 \\ 0 & x > 0 \end{cases}. \tag{7.57}$$

The new coordinates

$$\xi = x - vt, \tag{7.58a}$$

$$\tau = t, \tag{7.58b}$$

transform the ADE into a heat or solute diffusion problem given by

$$R \frac{\partial C}{\partial \tau} = D \frac{\partial^2 C}{\partial \xi^2}, \qquad (7.59)$$

$$C(\xi, 0) = \begin{cases} C_o & \xi < 0 \\ 0 & \xi < 0 \end{cases}. \qquad (7.60)$$

The solution for this and many other cases can be readily found in the literature on diffusion problems (Carslaw and Jaeger, 1959; Crank, 1975):

$$C(x, t) = \frac{C_o}{2} \text{erfc}\left(\frac{Rx - vt}{\sqrt{4DRt}} \right). \qquad (7.61)$$

Warrick (2003) reviews the use of a transformation similar to Equation 7.58a for solute transport in a three-dimensional porous medium of infinite extent. Other transformations to the diffusion problem have been employed as well (Brenner, 1962; Selim and Mansell, 1976; Zwillinger, 1989). Transformation of time to a time-integrated flow variable sometimes also allows one to derive an analytical solution of the nonlinear ADE for transient flow (Wierenga, 1977; Parker and van Genuchten, 1984; Huang and van Genuchten, 1995).

7.3.2.2 Laplace Transformation

The ADE is commonly solved directly with the method of Laplace transforms. The solution procedure will be illustrated here for an initially solute-free semi-infinite soil with a constant solute flux (vC_o) or concentration (C_o) at the inlet boundary. The mathematical problem consists of solving the ADE given by Equation 7.56 subject to

$$C - \delta \frac{D}{v} \frac{\partial C}{\partial x} = vC_o \quad \delta = \begin{cases} 0 & \text{first type} \\ 1 & \text{third type} \end{cases}, \qquad (7.62)$$

$$\frac{\partial C}{\partial x}(\infty, t) = 0, \qquad (7.63)$$

with δ as a coefficient depending on the type of inlet condition. The Laplace transform of the solute concentration with respect to time is defined as (Spiegel, 1965)

$$\overline{C}(x, s) = \mathcal{L}\left[C(x, t) \right] = \int_0^\infty C(x, t) \exp(-st) dt, \qquad (7.64)$$

where s is the (complex) transformation variable (T^{-1}). This transformation changes the transport equation from a partial to an ordinary differential equation:

$$\frac{d^2 \overline{C}}{dx^2} - \frac{v}{D} \frac{d\overline{C}}{dx} - \frac{sR}{D} \overline{C} = 0, \qquad (7.65)$$

subject to

$$\overline{C} - \delta \frac{D}{v} \frac{d\overline{C}}{dx} = \frac{C_o}{s}, \qquad (7.66)$$

$$\frac{d\overline{C}}{dx}(\infty, s) = 0, \qquad (7.67)$$

where the bar denotes a transformed variable. The following solution for the concentration in the Laplace domain is obtained with the help of the inlet condition:

$$\overline{C}(x, s) = \frac{v}{v - \delta \lambda^- D} \frac{C_o}{s} \exp(\lambda^- s), \quad \lambda^- = \frac{v}{2D} - \left[\left(\frac{v}{2D} \right)^2 + \frac{sR}{D} \right]^{1/2}. \qquad (7.68)$$

Inversion of this solution may be done with a table of Laplace transforms, by applying the inversion theorem or by using a numerical inversion program. The solution in the Laplace domain is also convenient to obtain transformations between resident and flux-averaged concentrations such as given by Equation 7.48. Finally, it should be noted that the solution for a finite outlet condition is also possible with the Laplace transform, although a bit more cumbersome (Brenner, 1962; Leij and van Genuchten, 1995).

7.3.2.3 Equilibrium Transport

van Genuchten and Alves (1982) provided a compendium of available analytical one-dimensional solutions for a variety of mathematical conditions and physical processes. Four common analytical solutions for a zero initial condition involving a first- or third-type inlet condition and an infinite or finite outlet condition are listed in Table 7.1. The solutions may be expressed in terms of the dimensionless variables P, T, and X (Equation 7.42). Typically, L is equal to the position of the outlet (the column length) for a finite system, whereas for a semi-infinite soil system an arbitrary length may be used.

Figure 7.7 contains solute profiles $C/C_o(X)$ according to the solutions listed in Table 7.1 using $R = 1$, $P = 5$ or 20, and $T = 0.25$ or 1. Curves are for a first- (A1) or a third-type (A2) inlet condition assuming an infinite outlet condition, or a first- (A3) or third-type (A4) inlet condition in case of finite outlet condition. The predicted profile for a first-type condition (A1, A3) for the lower Peclet number ($P = 5$) lies considerably above the line predicted for a third-type condition (A2, A4). The effect of the outlet condition is initially minor, but when the solute front reaches the outlet (L), a clear difference between a finite and an infinite outlet condition can be observed for both a first- (A1, A3) and a third-type (A2, A4) inlet condition. The simpler solution for a semi-infinite system can, in many cases, be used to approximate the solution for a finite condition; van Genuchten and Alves (1982) formulated the empirical restriction $X < 0.9 - 8/P$ on the position

TABLE 7.1 Analytical Solutions of the ADE for Different Boundary Conditions

Case	Inlet Condition	Outlet Condition	Analytical Solution $C(x,t)/C_o$
A1	$C(0,t) = C_o$	$\dfrac{\partial C}{\partial x}(\infty, t) = 0$	$\dfrac{1}{2}\,\mathrm{erfc}\!\left(\dfrac{Rx - vt}{\sqrt{4DRt}}\right) + \dfrac{1}{2}\exp\!\left(\dfrac{vx}{D}\right)\mathrm{erfc}\!\left(\dfrac{Rx + vt}{\sqrt{4DRt}}\right)$
A2	$\left(vC - D\dfrac{\partial C}{\partial x}\right)_{x=0^+} = vC_o$	$\dfrac{\partial C}{\partial x}(\infty, t) = 0$	$\dfrac{1}{2}\,\mathrm{erfc}\!\left(\dfrac{Rx - vt}{\sqrt{4DRt}}\right) + \sqrt{\dfrac{v^2 t}{\pi DR}}\exp\!\left(-\dfrac{(Rx - vt)^2}{4DRt}\right) - \dfrac{1}{2}\left(1 + \dfrac{vx}{D} + \dfrac{v^2 t}{DR}\right)\exp\!\left(\dfrac{vx}{D}\right)\mathrm{erfc}\!\left(\dfrac{Rx + vt}{\sqrt{4DRt}}\right)$
A3	$C(0,t) = C_o$	$\dfrac{\partial C}{\partial x}(L, t) = 0$	$1 - \displaystyle\sum_{n=1}^{\infty} \dfrac{2\beta_n \sin\!\left(\dfrac{\beta_n x}{L}\right)\exp\!\left(\dfrac{vx}{D} - \dfrac{v^2 t}{4DR} - \dfrac{\beta_n^2 Dt}{RL^2}\right)}{\beta_n^2 + \left(\dfrac{vL}{DR}\right)^2 + \dfrac{vL}{2D}} \qquad \beta_n \cot(\beta_n) + \dfrac{vL}{2D} = 0$
A4	$\left(vC - D\dfrac{\partial C}{\partial x}\right)_{x=0^+} = vC_o$	$\dfrac{\partial C}{\partial x}(L, t) = 0$	$1 - \displaystyle\sum_{n=1}^{\infty} \dfrac{\dfrac{2vL}{D}\beta_n\left[\beta_n \cos\!\left(\dfrac{\beta_n x}{L}\right) + \dfrac{vL}{2D}\sin\!\left(\dfrac{\beta_n x}{L}\right)\right]\exp\!\left(\dfrac{vx}{D} - \dfrac{v^2 t}{4DR} - \dfrac{\beta_n^2 Dt}{RL^2}\right)}{\left[\beta_n^2 + \left(\dfrac{vL}{DR}\right)^2 + \dfrac{vL}{2D}\right]\left[\beta_n^2 + \left(\dfrac{vL}{2D}\right)^2\right]} \qquad \beta_n \cot(\beta_n) - \dfrac{\beta_n^2 D}{vL} + \dfrac{vL}{2D} = 0$
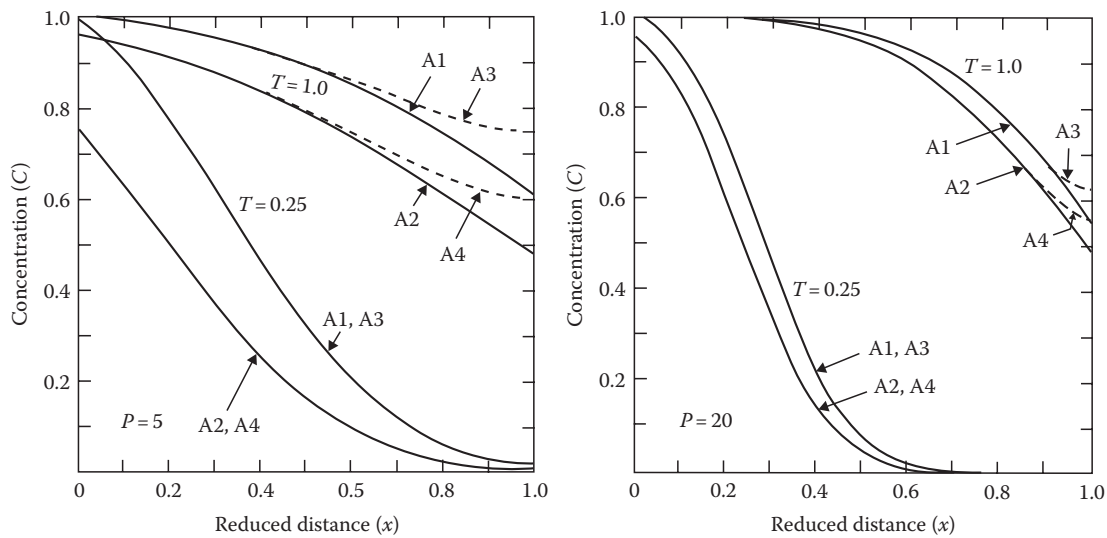
FIGURE 7.7 Plot of the concentration (C/C_o) as a function of distance (X) calculated for four different combinations of boundary conditions according to the solutions in Table 7.1 at two different times for a Peclet number of 5 and 20. (After van Genuchten, M.Th., and W.J. Alves. 1982. Analytical solutions of the one-dimensional convective–dispersive solute transport equation. USDA Tech. Bull. 1661. USDA, Washington, DC, 1982.)

for which such an approximation is reasonable. For smaller times when the solute has not reached the outlet, the finite and infinite outlet conditions obviously lead to a similar solution.

For a third-type inlet condition, the concentration at $X = 0$ just inside the soil is not equal to the influent concentration, even at time $T = 1$. Although the jump in concentration is physically odd, mass conservation is ensured. For the higher Peclet number $(P = 20)$, deviations between a first- and a third-type inlet condition are considerably reduced.

Large differences in the predicted concentration may occur if the solute front approaches the outlet at $X = 1$ or $x = L$. Calculated concentrations according to solutions A1, A2, A3, and A4 versus the Peclet number at the outlet for $T = 1$ are illustrated in Figure 7.8.

The greatest difference occurs for small Peclet numbers, namely, when hydrodynamic dispersion is relatively important. The nature of hydrodynamic dispersion suggests that C/C_o should be approximately 0.5 for $X = T = R = 1$, the average of the zero initial concentration and the influent concentration (C_o). Because of the effect of the boundary conditions, this only happens when the Peclet number exceeds 10 or more, depending on the type of solution. For the first-type inlet condition (A1, A3), C/C_o exceeds 0.5 at low Peclet numbers since a considerable amount of solute is forced to diffuse into the column to establish a constant inlet concentration.

Finally, for displacement experiments involving finite columns, it may be of interest to quantify the amount of solute
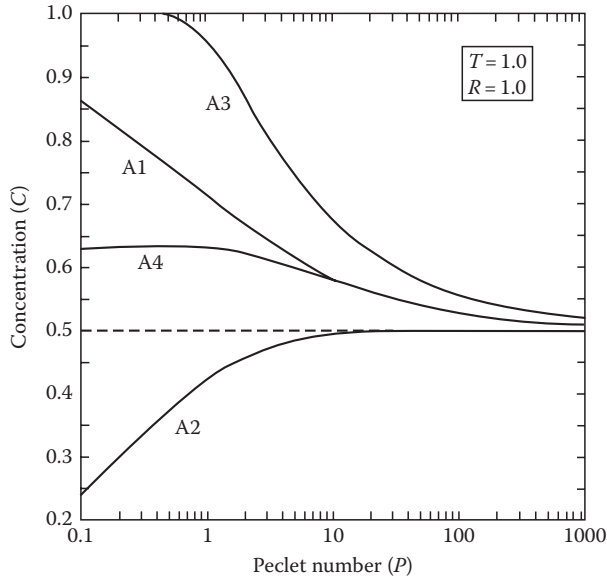
FIGURE 7.8 Solute concentration predicted with solutions A1, A2, A3, and A4 as a function of the Peclet number at $T = 1$ and $X = 1$. (After van Genuchten, M.Th., and W.J. Alves. 1982. Analytical solutions of the one-dimensional convective–dispersive solute transport equation. USDA Tech. Bull. 1661. USDA, Washington, DC, 1982.)

that can be stored in the liquid and sorbed phases of the soil (Equation 7.24). When, beginning at $t = 0$, a solution with concentration (C_o) is applied to a soil column, holdup can be defined as (Nauman and Buffham, 1983)

$$H = \frac{v}{L} \int_0^\infty \left[1 - \frac{C(L,t)}{C_o} \right] dt. \qquad (7.69)$$

The holdup is determined by the area above the breakthrough curves, that is, by integrating the "complementary" solute concentration versus dimensionless time. van Genuchten and Parker (1984) showed that $H = R$ for solutions A1 and A4, $H = R[1 + (1/P)]$ for solution A2, and $H = R\{1 - (1/P) [1 - \exp(-P)]\}$ for solution A3. In case of anion exclusion, the relative exclusion volume equals $1 - R$ and the column holdup will be less than 1.

7.3.2.4 Nonequilibrium Transport

Analytical solutions for two-domain nonequilibrium transport have been presented by, among others, Lindstrom and Narasimhan (1973), Lindstrom and Stone (1974), Lassey (1988), and Toride et al. (1993). The boundary value problem involving solute application with a constant concentration may be specified by Equations 7.40 and 7.41 subject to the following conditions:

$$C_1(X,0) = C_2(X,0) = 0, \qquad (7.70)$$

$$\left(C_1 - \frac{1}{P} \frac{\partial C_1}{\partial X} \right)_{X=0^+} = 1, \qquad (7.71)$$

$$\frac{\partial C_1}{\partial X}(\infty, T) = 0. \qquad (7.72)$$

Solutions for the equilibrium and nonequilibrium concentrations are

$$C_1 = \int_0^T J(a,b)G(X,\tau)d\tau, \qquad (7.73)$$

$$C_2 = \int_0^T \left[1 - J(b,a) \right] G(X,\tau)d\tau, \qquad (7.74)$$

where the auxiliary (equilibrium) function $G(X, \tau)$ for resident concentrations is defined as

$$G(X,\tau) = \sqrt{\frac{P}{\pi\beta R\tau}} \exp\left(-\frac{(\beta RX - \tau)^2}{4\beta R\tau/P} \right) - \frac{P}{2\beta R} \text{erfc}\left(-\frac{\beta RX + \tau}{\sqrt{4\beta R\tau/P}} \right), \qquad (7.75)$$

and for flux-averaged concentrations by

$$G(X,\tau) = \sqrt{\frac{\beta RPX^2}{4\pi\tau^3}} \exp\left(-\frac{(\beta RX - \tau)^2}{4\beta R\tau/P} \right). \qquad (7.76)$$

Furthermore, J denotes Goldstein's J function (Goldstein, 1953), which is given by

$$J(a,b) = 1 - \exp(-b) \int_0^a \exp(-x)I_0(2\sqrt{ab})d\xi, \qquad (7.77)$$

with I_0 as the zero-order modified Bessel function. The variables a and b are given by

$$a = \frac{\omega\tau}{\beta R}, \qquad (7.78a)$$

$$b = \frac{\omega(T - \tau)}{(1 - \beta)R}. \qquad (7.78b)$$

The above solution for a flux-averaged concentration was used to describe breakthrough data for the pesticide 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) as observed from a 30 cm long column containing a Glendale clay loam with soil aggregates with a diameter of up to 6 mm (van Genuchten and Parker, 1987). The two-region model (TRM) three adjustable parameters (P, R, ω) provided an excellent description of the data (Figure 7.9b). The retardation factor (R) was estimated independently, using
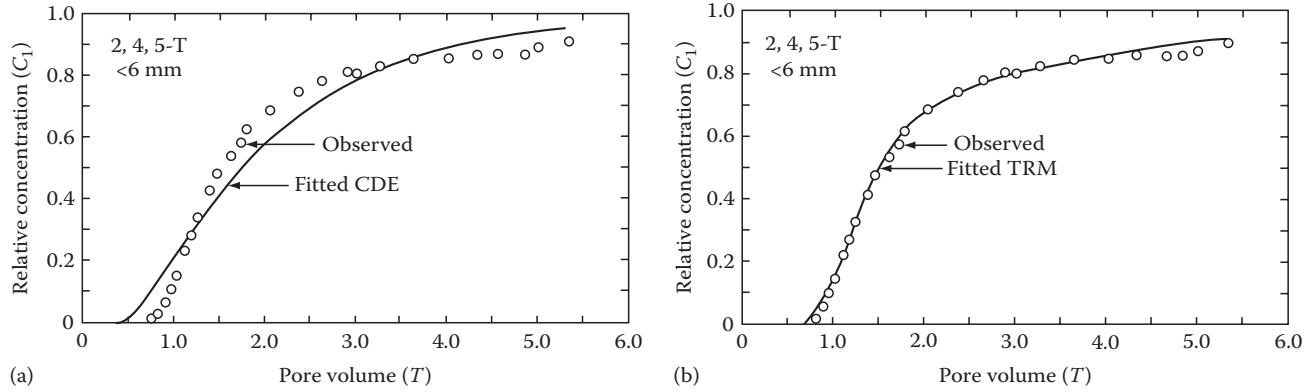
FIGURE 7.9 Observed and fitted effluent curves for 2,4,5-T movement through Glendale clay loam. The fitted curves were based on (a) the equilibrium ADE or CDE and (b) the nonequilibrium ADE or TRM.

the distribution coefficient obtained from batch experiments, according to Equation 7.23. Using the ADE fit with P (Figure 7.9a), or P and R (not shown), as adjustable parameters resulted in a poor description of the data.

As pointed out by van Genuchten and Dalton (1986), the main disadvantage of the first-order physical nonequilibrium approach is the obscure dependency of the transfer coefficient (α or ω) on the actual diffusion process in the aggregate, particularly the value for the diffusion coefficient and the aggregate geometry. The diffusion process may be modeled for well-defined structured or aggregated porous media to allow a better description of the concentration distribution inside the aggregate. Analytical solutions are available for several aggregate geometries. The simplified immobile concentration, which is used in the general nonequilibrium formulation, can always be obtained by averaging the more detailed solution over the aggregate volume.

7.3.2.5 Time Moments

Moments are frequently used to characterize statistical distributions such as those of solute particles (concentrations) versus time or position. Analytical expressions for lower order moments are sometimes derived for deterministic transport models, especially when a direct analytical solution may be difficult to obtain. Moment analysis is more widely employed in stochastic than deterministic transport models since the uncertainty in both model parameters and predicted results is most conveniently quantified with moments. Only time moments will be considered here.

The pth (time) moment of the breakthrough curve, as obtained, say, from effluent samples collected from a soil column (with length $x = L$) to which a solute pulse is applied during steady water flow, is defined as

$$m_p(L) = \int_0^\infty t^p C(L,t)dt \quad (p = 0,1,2,\ldots). \tag{7.79}$$

The zero moment is proportional to the total solute mass, the first moment quantifies the mean displacement, the second moment is indicative of the variance (dispersion), whereas the

third moment quantifies the asymmetry or skewness of the breakthrough curve. Normalized moments (μ_p) are obtained as follows:

$$\mu_p = \frac{m_p}{m_0}. \tag{7.80}$$

The mean breakthrough time is given by μ_1. Central moments are defined with respect to this mean according to

$$\mu'_p(L) = \frac{1}{m_0}\int_0^\infty (t - \mu_1)^p C(L,t)dt. \tag{7.81}$$

The variance of a breakthrough curve, which can be used to assess solute dispersion, is given by the second central moment μ'_2. The degree of asymmetry of the breakthrough curve is indicated by its skewness $\mu'_3/(\mu'_2)^{3/2}$.

The previous definitions are employed to obtain numerical values for moments from experimental results. Substitution of an analytical solution for the solute concentration into the definitions allows the derivation of algebraic expressions for time moments. Values for transport parameters can be obtained by equating numerical and algebraic moments (Jacobsen et al., 1992; Leij and Dane, 1992). This procedure is not reliable if experimental moments of higher order ($p > 2$) are needed since even small deviations, at larger times, between experimental and modeled concentrations will greatly bias such moments.

Algebraic moments are normally obtained using the solution for the concentration in the Laplace domain. The following equality can be established from properties of the Laplace transform (cf. Aris, 1958; van der Laan, 1958):

$$m_p(x) = (-1)^p \lim_{s \to 0} \frac{d^p \overline{C}(x,s)}{ds^p}, \tag{7.82}$$

where, as before,
$\overline{C}(x, s)$ is the concentration in the Laplace domain
s is the (complex) transformation variable

Expressions for moments can hence be obtained by differentiating the solution in the Laplace domain and evaluating the result for $s = 0$. This task is conveniently handled by mathematical software.

Time moments will now be considered for three different transport problems. First, the mathematical problem for physical nonequilibrium transport can be written as

$$\theta_m R \frac{\partial C_m}{\partial t} + \theta_{im} R \frac{\partial C_{im}}{\partial t} = \theta_m D_m \frac{\partial^2 C_m}{\partial x^2} - \theta v \frac{\partial C_m}{\partial x}, \quad (7.83)$$

$$\theta_{im} R \frac{\partial C_{im}}{\partial t} = \alpha(C_m - C_{im}). \quad (7.84)$$

The conditions for instantaneous solute application to a soil are

$$C(x,0) = 0 \quad 0 \le x < \infty, \quad (7.85)$$

$$C(0,t) = \frac{m_0}{v}\delta(t), \quad (7.86)$$

$$\frac{\partial C}{\partial X}(\infty, t) = 0, \quad (7.87)$$

where

m_0 is the solute mass that is applied per unit area of soil solution at $t = 0$

$\delta(t)$ is the Dirac delta function (T^{-1})

The first-type inlet condition is used to describe flux-averaged concentrations such as effluent samples from column displacement experiments. Second, the equilibrium problem is defined by the same set of equations by setting $\theta_m = \theta$, $\theta_{im} = 0$, and $\alpha \to 0$. Third, the chemical nonequilibrium transport equations are as follows:

$$\frac{\partial C}{\partial t} + \frac{\rho_b}{\theta}\frac{\partial S}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x}, \quad (7.88)$$

$$\frac{\partial S}{\partial t} = \alpha(K_d C - S). \quad (7.89)$$

These equations are also subject to boundary and initial conditions in Equations 7.85 through 7.87.

Formulas for the mean breakthrough time (μ_1) and the variance (μ_2) of the breakthrough curve predicted for these three models are presented in Table 7.2 for a first-type inlet condition, which should be employed if effluent samples are analyzed (i.e., flux-averaged concentrations). The expressions suggest that nonequilibrium conditions do not affect the mean travel time but they do increase solute spreading. Since only the solution in the Laplace domain is needed, moment analysis is particularly useful for more complex transport problems to study the general behavior of the breakthrough curve.

TABLE 7.2 Mean Breakthrough Time (μ_1) and Variance (μ_2') for the Equilibrium and Nonequilibrium Solution of the ADE at a Distance x from the Inlet as a Result of a Dirac Delta Input Described with a First-Type Inlet Condition

ADE Model	Mean Breakthrough Time, μ_1	Variance, μ_2'
Equilibrium	$\dfrac{Rx}{v}$	$\dfrac{2DR^2 x}{v^3}$
Nonequilibrium		
Two-region	$\dfrac{Rx}{v}$	$\dfrac{2\theta_m D_m R^2 x}{\theta v^3} + \dfrac{2\theta(1-\beta)^2 R^2 x}{\alpha v}$
Two-site	$\dfrac{Rx}{v}$	$\dfrac{2DR^2 x}{v^3} + \dfrac{2(1-\beta)Rx}{\alpha v}$

7.3.3 Numerical Solutions

The solution of many practical transport problems requires the use of numerical methods because of changes in water saturation (as the result of irrigation, evaporation, and drainage), spatial and temporal variability of soil properties, or complicated boundary and initial conditions. Numerical methods are based on a discretization of the spatial and temporal solution domain and subsequent calculation of the concentration at discrete nodes in the domain. This approach is in contrast with analytical methods, which offer a continuous description of the concentration. In some cases, a combination of analytical and numerical techniques may be employed (Sudicky, 1989; Moridis and Reddell, 1991; Li et al., 1992).

The accuracy of the numerical results depends on the input parameters, the approximation of the governing partial differential equation, the discretization, and implementation of the numerical solution in a computer code solving the simulated problem. Numerous texts exist on the numerical modeling of flow and transport in porous media (Pinder and Gray, 1977; Huyakorn and Pinder, 1983; Campbell, 1985; van der Heijde et al., 1985; Istok, 1989).

The many numerical methods for solving the ADE can be classified into three groups (Neuman, 1984): (1) Eulerian, (2) Lagrangian, and (3) mixed Lagrangian–Eulerian. In the Eulerian approach, the transport equation is discretized by the method of finite differences or finite elements using a fixed mesh. For the Lagrangian approach, the mesh deforms along with the flow while it is stable in a moving coordinate system. A two-step procedure is followed for a mixed approach. First, advective transport is solved using a Lagrangian approach and concentrations are obtained from particle trajectories. Subsequently, all other processes including sinks and sources are modeled with an Eulerian approach using finite elements, finite differences, etc.

The method of finite differences (Bresler and Hanks, 1969; Bresler, 1973) and the Galerkin method of finite elements (Gray and Pinder, 1976; van Genuchten, 1978) belong to the first group as do the previously mentioned combination of analytical and numerical techniques. The finite difference and finite element

methods were the first numerical methods used for solute transport problems and, in spite of their problems discussed below, are still the most often utilized methods. Numerical experiments have shown that both methods give very good results if significant dispersion exists as quantified with the Peclet number. If advection is dominant, however, numerical oscillations may occur for both methods and small spatial increments should be used. It may not always be possible to decrease the spatial step size due to the associated increase in computations, and a variety of approaches have been developed to overcome the oscillations (Chaudhari, 1971; van Genuchten and Gray, 1978; Donea, 1991).

Lagrangian solution methods will result in very few numerical oscillations (Varoglu and Finn, 1982). However, Lagrangian methods, which are based on the method of characteristics, suffer from inherent diffusion and do not conserve mass. They are difficult to implement for 2D and three-dimensional problems. Instabilities resulting from inadequate spatial discretization may occur during longer simulations due to deformation of the stream function, especially if the solute is subject to sorption, precipitation, and other reactions.

The mixed approach has been applied by several authors (Konikov and Bredehoeft, 1978; Yeh, 1990). In view of the different behaviors of the diffusive (parabolic) and advective (hyperbolic) terms of the ADE, the problem is decomposed into an advection and a diffusion problem. Advective transport is solved with the Lagrangian approach while all other terms are solved with the Eulerian approach. The trajectories of flowing particles are obtained using continuous forward particle tracking (to follow a set of particles as they move through the flow domain), single-step reverse particle tracking (the initial position of particles arriving at nodal points was calculated for each time step), and a combination of both approaches. Ahlstrom et al. (1977), among others, used the random walk model to describe the movement of individual solute particles by viewing the travel distance for a particular time step as the sum of a deterministic and a stochastic velocity component.

In the following, only a brief introduction to the finite difference and finite element methods to solve transport problems is provided. Both methods encompass a wide variety of more specific numerical approaches. As a rule of thumb, the finite difference method is attractive because of its simplicity and the availability, at least early on, of handbooks and computer programs simulating flow and transport in porous media. On the other hand, the finite element method has proven to be more suitable for problems involving irregular geometries of the flow and transport problem, such as situations involving flow to drainage pipes and along sloping soil surfaces.

7.3.3.1 Finite Difference Methods

For one-dimensional transient solute transport, the dependent variable can be discretized according to

$$C(x,t) = C(i\Delta x, j\Delta t) = C_{ij} \quad (i = 0,1,2,\ldots,n; \; j = 0,1,2,\ldots,m). \quad (7.90)$$

Consider the simulation of the one-dimensional ADE for steady flow of a conservative tracer as given by Equation 7.56 with $R = 1$. Temporal and spatial derivatives are approximated with Taylor series expansions. Assume that the concentrations are known at the current time ($t = j\Delta t$) and that the objective is to calculate the concentration distribution $\{C[i\Delta x, (j + 1)t]\}$ at the next time. A forward-in-time finite difference scheme where the unknown concentration is given explicitly in terms of known concentrations is written as

$$\frac{C_{i,j+1} - C_{i,j}}{\Delta t} \approx D \frac{C_{i+1,j} - 2C_{i,j} + C_{i-1,j}}{(\Delta x)^2} - v \frac{C_{i+1,j} - C_{i-1,j}}{2(\Delta x)}. \quad (7.91)$$

On the other hand, a backward-in-time or implicit scheme is given by

$$\frac{C_{i,j+1} - C_{i,j}}{\Delta t} \approx D \frac{C_{i+1,j+1} - 2C_{i,j+1} + C_{i-1,j+1}}{(\Delta x)^2} - v \frac{C_{i+1,j+1} - C_{i-1,j+1}}{2(\Delta x)},$$

$$(7.92)$$

which contains several concentrations at the next time level. The problem is solved implicitly after combining the difference schemes of all nodes, and subsequently using a matrix equation solver. Finally, a weighted scheme can be defined as

$$\frac{C_{i,j+1} - C_{i,j}}{\Delta t}$$

$$\approx D \frac{w(C_{i+1,j+1} - 2C_{i,j+1} + C_{i-1,j+1}) + (1-w)(C_{i+1,j} - 2C_{i,j} + C_{i-1,j})}{(\Delta x)^2}$$

$$- v \frac{w(C_{i+1,j+1} - C_{i-1,j+1}) + (1-w)(C_{i+1,j} - C_{i-1,j})}{2(\Delta x)}, \quad (7.93)$$

with w being a weighting constant between 0 and 1. The scheme is said to be fully explicit for $w = 0$ and fully implicit for $w = 1$, while a Crank–Nicolson central-in-time scheme is derived if $w = 0.5$. For transient flow, the scheme becomes more complicated; the velocity (v) and water content (θ) are usually obtained by solving the Richards equation prior to solving the transport problem.

Errors associated with the discretization and the solution procedure can be evaluated by comparison with analytical results, provided that the problem can be sufficiently idealized (linearized) to permit the use of such solutions. This is true for other numerical solutions as well. For a convergent scheme, the difference between the numerical and analytical solutions should decrease if smaller space and time steps are used in the numerical solution; the difference should become zero if $\Delta x \to 0$ and $\Delta t \to 0$, barring round off and computational errors.

An even more important question relates to the stability of the finite difference approximation, namely, the degree to which the numerical solution is affected by errors that occur during the simulation. Such errors can usually not be eliminated completely; they depend on the implemented discretization,

the values of the input parameters, and the type of numerical scheme used for approximation of the governing transport equation. Errors are damped during the course of a simulation when a stable scheme is used, while unstable schemes allow such errors to grow without bounds.

Implicit systems are unconditionally stable, but their results may not be as amenable to changes in grid sizes as is the case with explicit systems. The Crank–Nicolson method offers an attractive compromise of being unconditionally stable and having a truncation error of order $O[(\Delta x)^2 + (\Delta t)^2]$. These properties imply, among other things, that a variable time step can be used independently of the spatial step to effectively balance the needs of an accurate approximation and a limited number of computations. However, oscillations near the concentration front may develop even for unconditionally stable methods due to the hyperbolic (convection) term in the solute transport equation. To avoid these oscillations, stability criteria in terms of the grid Peclet (Pe) and Courant (Cr) numbers are frequently formulated. Huyakorn and Pinder (1983) provided the following conservative guidelines for one-dimensional transport of a nonreactive solute:

$$Pe = \frac{v\Delta x}{D} < 2, \tag{7.94a}$$

$$Cr = \frac{v\Delta t}{\Delta x} < 1. \tag{7.94b}$$

For transport of a reactive solute, the retardation factor R must be included in the denominator of the Courant number.

7.3.3.2 Finite Element Methods

The application of the finite element methods for modeling solute transport in soils involves several steps, which will be briefly reviewed in a qualitative manner. The specifics of the method can be found in, among others, Huyakorn and Pinder (1983) and Istok (1989). The (spatial) solution domain should first be discretized. The finite element mesh, consisting of nodal points marking the elements, is usually tailored to the problem at hand. The selection of mesh size is a somewhat subjective process that considers the required degree of accuracy, the geometry of the problem, the ease of mesh generation, and the mathematical complexity associated with the use of a particular mesh. The shape of the finite elements is determined by the dimensionality and the geometry of the transport problem. One-dimensional elements consist of lines between nodal points along the coordinate where (one-dimensional) transport occurs. Examples of two-dimensional elements are triangles, rectangles, and parallelograms, while many different three-dimensional elements can be constructed.

A second step in implementing the finite element method is the description of the transport problem for each element. Generally, the method of weighted residuals is used to arrive at an integral formulation for the governing transport equation.

An approximate or trial solution for the concentration in each element is formulated as the weighted sum of the unknown concentration at the nodes of that element. A wide variety of so-called basis functions can be selected to assign weights to the nodes; the functions depend on the type of element being selected. The approximation for the element concentration will not be exact; substitution of the element concentration in the ADE yields an expression for this error (residual). The objective is to minimize the residual over the entire domain with some kind of weighting procedure for the elements. The method of weighted residuals forces the weighted average of the residuals at the nodes to be 0 and gives an expression for the residual of each element. After applying Green's theorem for each element, the residual can be written in matrix form.

The third step is the assembly of all element matrices into a global matrix system that contains the nodal concentrations and its temporal derivative as unknown variables.

The last step involves the solution of this global matrix equation. Since there are, in general, no obvious advantages for also using a finite element approximation for the temporal derivative, the latter is normally dealt with using the (simpler) finite difference method. The time step may be constant or variable. As noted earlier, the explicit or forward scheme ($w = 0$) is conditionally stable whereas the implicit or backward ($w = 1$) and the centered or Crank–Nicolson ($w = 0.5$) schemes are unconditionally stable. The nodal concentration can then be solved using standard numerical procedures.

7.3.3.3 Application

Numerical oscillations and dispersion are illustrated for several numerical schemes using results by Huang et al. (1997). The numerical solution of the ADE is particularly difficult for relatively sharp concentration fronts with advection-dominated transport characterized by small dispersivities (i.e., large Pe). As mentioned earlier, undesired oscillations can often be prevented with judicious space and time discretizations. A smaller grid size with $Pe < 2$ will virtually eliminate numerical oscillations. However, acceptable results may still be obtained for local Peclet numbers as high as 10 (Huyakorn and Pinder, 1983). The time step is selected based upon the Courant number.

Consider one-dimensional, steady-state flow in a soil column with $v = 4\,cm\,day^{-1}$. The column was initially free of any solute; a nonreactive solute ($C_o = 1$) was subsequently applied for 20 days. For negligible molecular diffusion and a longitudinal dispersivity ($\alpha_L = 0.02\,cm$), the grid $Pe = 100$ for a spatial increment $\Delta x = 2\,cm$. This grid Peclet number indicates that advection is the dominant mechanism for solute transport. In all calculations, the Courant number was less than 1, which is the stability condition for Eulerian methods.

The concentration predicted with four different finite element methods and with the analytical solution is shown in Figure 7.10. The first numerical method is based on the central Crank–Nicolson scheme for time with weighting constant ($w = 0.5$). Second, the analytical solution is shown as a solid line. Third, an implicit scheme ($w = 1.0$) is used. The last two numerical methods
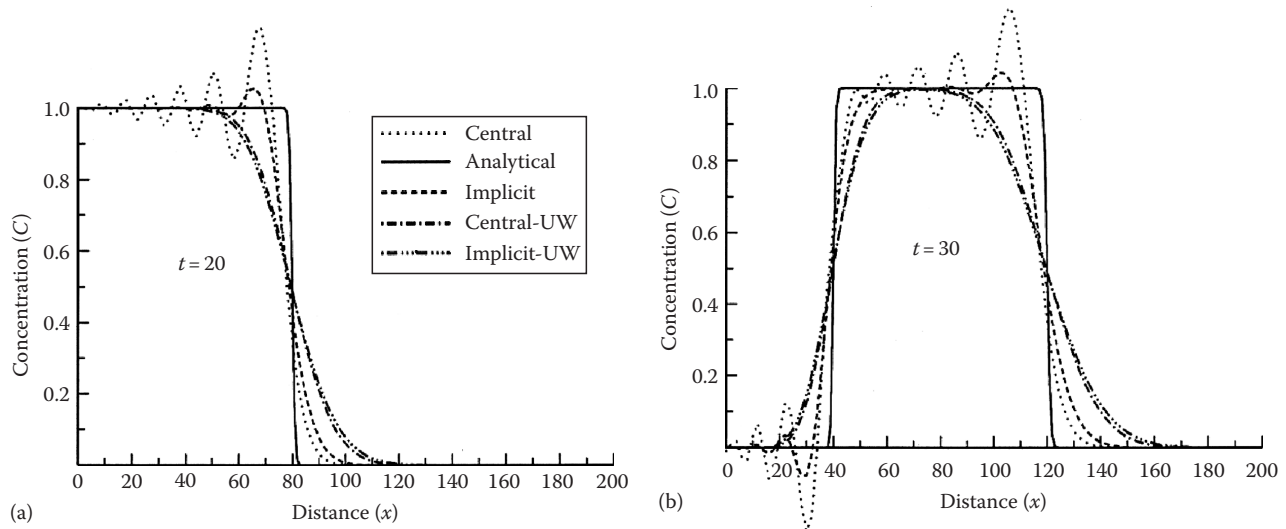
FIGURE 7.10 Concentration profiles predicted with different numerical schemes for (a) solute infiltration ($t = 20$ days) and (b) solute leaching ($t = 30$ days).

implement upstream weighting (Huyakorn and Pinder, 1983) in the central and the implicit schemes, respectively. The results obtained with the Crank–Nicolson method have significant numerical oscillations (both overshoot and undershoot) Figure 7.10). This is not surprising given the large value for *Pe*. The oscillations are reduced by using an implicit scheme. Upstream weighting leads to very similar results for the central and implicit schemes. The oscillations are virtually eliminated but the solute profile exhibits more numerical dispersion; there is a greater discrepancy with the steep analytical profile than for the regular implicit scheme.

7.4 Stream Tube Models

Considerable errors may be made when applying deterministic methods to the field since model parameters are actually stochastic due to spatial and temporal variability, measurement errors, and different averaging scales. Usually, it is not possible or important to obtain discrete values for parameters in deterministic models of field-scale transport. Instead, transport properties and model results are described with statistical functions. Stochastic modeling is no substitute for data collection or model development but merely a method to deal with uncertainty of model parameters and complexity of flow and transport processes. The scale at which solute movement is observed or modeled is important. The averaging process associated with larger field-scale descriptions tends to filter out the variability at smaller scales.

The stochastic modeling of actual field problems is seldom possible. The stream tube model provides a possible exception (Dagan and Bresler, 1979; Amoozegar-Fard et al., 1982; Rubin and Or, 1993). The field is conceptualized as a system of parallel tubes illustrated in Figure 7.11. A process-based model is used to describe a one-dimensional, autonomous transport in each tube as a function of time and depth. Transport parameters are
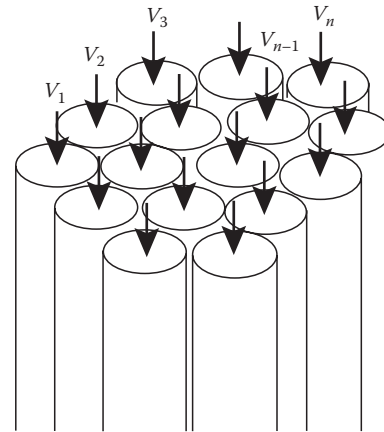
FIGURE 7.11 Schematic of the stream tube model.

either deterministic or stochastic. The problem may be solved analytically at the scale of the tube and the field. The stream tube model is suitable for inverse procedures to estimate transport and statistical properties (Toride et al., 1995).

7.4.1 Model Formulation

For one-dimensional transport, the solute concentration at the outlet of a stream tube of length (L) may be written with the transfer function approach as (Jury and Roth, 1990)

$$C(L,t) = \int_0^t C(0, t - \tau) f(L, \tau) d\tau. \qquad (7.95)$$

The outlet concentration is a convolution of the input signal and the residence time distribution $f(L, \tau)$. The latter is, in effect, a probability density function (pdf) of the time a solute particle resides in the soil between $x = 0$ and L. The pdf, which has

dimension of inverse time, can be determined from experimental results or it may be a theoretical expression derived from a process-based model such as the ADE. The pdf depends mostly on the transport properties of the soil and, to a lesser extent, on the mode of solute application and detection, but not on the input and initial concentrations. For the equilibrium problem involving instantaneous solute application to a solute-free soil, the flux-averaged concentration and hence the pdf is given by

$$f(x,t) = \frac{x}{\sqrt{4\pi Dt^2}} \exp\left(-\frac{(x-vt)^2}{4Dt} \right).\qquad(7.96)$$

Note that this is a Gaussian pdf. Expressions for $f(x, t)$ derived at a particular depth can be readily used for predicting concentrations at other depths if the soil is homogeneous.

The stochastic approach is implemented by considering the (constant) column parameters as realizations of a random distribution. The (horizontal) field-averaged solute concentration is considered the ensemble average over the probability distribution. It is assumed that each random parameter obeys a distribution function that is independent of location and that the ensemble of possible concentrations may be estimated from a sufficient number of samples taken at different locations. The average concentration across the field is then identical to the ensemble average:

$$<C(x,t)> = \frac{1}{A}\iint C(x,t)dA = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}C_i(x,t),\qquad(7.97)$$

where

  A denotes the area of the field
  n is the number of samples
  $<\ >$ indicates an ensemble average

7.4.2 Application

Solute transport in a local-scale stream tube may be described with the one-dimensional ADE. Following Toride and Leij (1996a), the effects of heterogeneity are studied using pairs of random parameters (v and K_d). Note that the water content (θ) and bulk density (ρ_b) are the same for all stream tubes. The field-scale mean concentration (denoted by ^) is equal to the ensemble average:

$$\hat{C}(x,t) = <C(x,t)> = \iint_0^{\infty\ \infty} C(x,t;v,K_d)f(v,K_d)dv\,dK_d,\qquad(7.98)$$

where the bivariate lognormal joint probability density function $f(v, K_d)$ is given by

$$f(v,K_d) = \frac{1}{2\pi\sigma_v\sigma_{K_d}vK_d\sqrt{1-\rho_{vK_d}^2}}\exp\left(-\frac{Y_v^2 - 2\rho_{vK_d}Y_vY_{K_d} + Y_{K_d}^2}{2\left(1-\rho_{vK_d}^2\right)}\right),$$
$$(7.99)$$

with

$$\rho_{vK_d} = <Y_vY_{K_d}> = \iint_0^{\infty\ \infty} Y_vY_{K_d}f(v,K_d)dv\,dK_d,\qquad(7.100)$$

$$Y_v = \frac{\ln(v) - \mu_v}{\sigma_v},\qquad(7.101a)$$

$$Y_{K_d} = \frac{\ln(K_d) - \mu_{K_d}}{\sigma_{K_d}},\qquad(7.101b)$$

where

  μ and σ are the mean and standard deviation of the log-transformed variable, respectively

  ρ_{vK_d} is the correlation coefficient between Y_v and Y_{K_d} (K_d tends to increase with v for positive ρ_{vK_d} and to decrease with v for negative correlation)

Ensemble averages of v and K_d are given by (Aitcheson and Brown, 1963),

$$<v> = \exp\left(\mu_v + \frac{1}{2}\sigma_v^2\right),\qquad(7.102a)$$

$$<K_d> = \exp\left(\mu_{K_d} + \frac{1}{2}\sigma_{K_d}^2\right),\qquad(7.102b)$$

with coefficients of variation (CV):

$$CV(v) = \sqrt{\exp(\sigma_v^2) - 1},\qquad(7.103a)$$

$$CV(K_d) = \sqrt{\exp\left(\sigma_{K_d}^2\right) - 1}.\qquad(7.103b)$$

Based upon the detection mode of the local-scale concentration, three types of field-scale concentrations can be defined: (1) the ensemble average of the flux-averaged concentration ($<C_F>$); (2) the field-scale resident concentration $\hat{C}_R$, which is equal to the ensemble average of the resident concentration ($<C_R>$); and (3) the field-scale flux-averaged concentration ($\hat{C}_F$), which is defined as $<vC_F>/<v>$. The second type of concentration is obtained from averaging values of the resident concentration at a particular depth across the field. The third type is defined as the ratio of ensemble solute and water fluxes in a similar manner as for deterministic transport. It should be noted that $<vC_F>/<v> \neq <C_F>$ because v is a stochastic variable.

The use of the stream tube model for field-scale transport is illustrated in Figure 7.12. The local-scale concentration only depends on the particular realizations of the two stochastic parameters (v and K_d) after the independent variables (t and x) have been specified. The solution for the ADE, the resident concentration
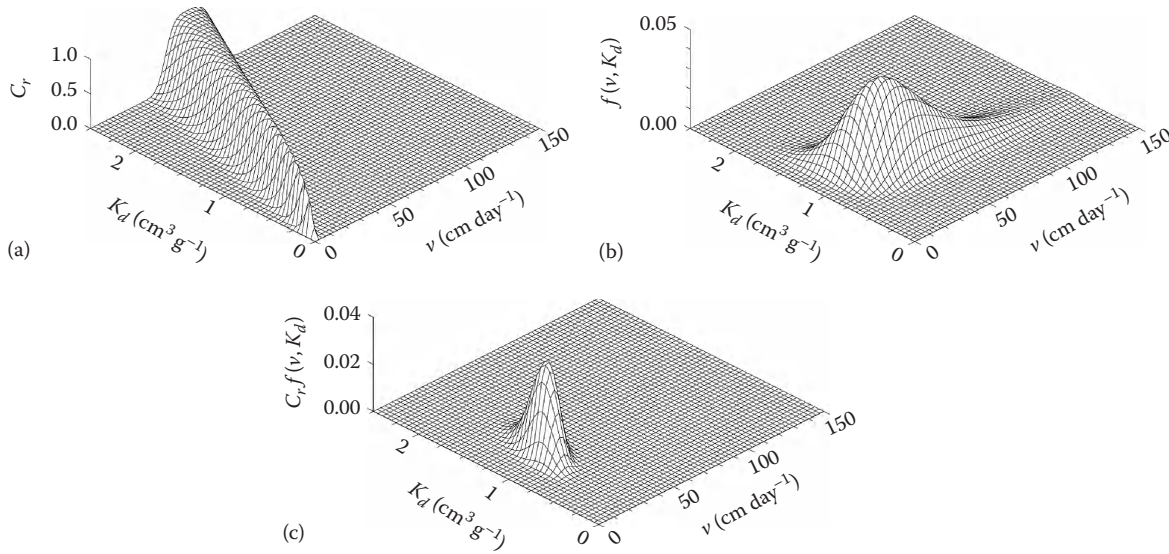
FIGURE 7.12 Illustration of the stream tube model for field-scale transport for a two-dimensional solute application with $D = 20\,cm^2\,day^{-1}$ and $\rho_b/\theta = 4\,g\,cm^{-3}$: (a) the local-scale resident concentration as a function of v and K_d at $x = 100\,cm$ and $t = 5$ days; (b) a bivariate lognormal pdf for $\rho_{vK_d} = 0.5$, $<v> = 50\,cm\,day^{-1}$, $\sigma_v = 0.2$, $<K_d> = 1\,cm^3\,g^{-1}$, and $\sigma K_d = 0.2$; and (c) the expected concentration at $x = 100\,cm$ and $t = 5$ days.

at $x = 100\,cm$ and $t = 5$ days as a function of v and K_d is shown in Figure 7.12a and the bivariate lognormal pdf for v and K_d in Figure 7.12b; the distribution is skewed with respect to v since σ_v is fairly high; the smaller σ_{Kd} results in a more symmetric distribution for K_d. The negative ρ_{vK_d} results in an increasing v with a decreasing K_d, and vice versa. The expected concentration—again a resident or volume-average—is shown in Figure 7.12c. It is obtained by weighting the local concentration (Figure 7.12a) by multiplying it with the joint pdf (Figure 7.12b). The peak in Figure 7.12c suggests that stream tubes with approximately $v = 25\,cm\,day^{-1}$ and $K_d = 1\,cm^3\,g^{-1}$ contribute the most to solute breakthrough when $x = 100\,cm$ and $t = 5$ days. The volume of the distribution in Figure 7.12c corresponds to the ensemble average ($<C_R>$).

Variations in the local-scale concentration between stream tubes, at a particular depth and time, can be characterized by its variance. The variance across the horizontal plane is given by (Bresler and Dagan, 1981; Toride and Leij, 1996b):

$$\mathrm{Var}\left[C(x,t)\right] = \int_0^\infty \int_0^\infty \left[C(x,t) - <C(x,t)>\right]^2 f(v, K_d)\, dv\, dK_d$$

$$= <C^2(x,t)> - <C(x,t)>^2. \tag{7.104}$$

For a deterministic distribution coefficient (K_d), Equation 7.99 reduces to

$$<C(x,t)> = \int_0^\infty C(x,t;v) f(v)\, dv, \tag{7.105}$$

where the lognormal pdf for the single stochastic variable (v) is given by

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v v} \exp\left(-\frac{\left[\ln(v) - \mu_v\right]^2}{2\sigma_v^2}\right). \tag{7.106}$$

The mean (Figure 7.13a) ($\hat{C}_R = <C_R>$) and the variance (Figure 7.13b) are shown as a function of depth for three values of σ_v at $t = 3$ days for a nonreactive solute. More solute spreading occurs in the $\hat{C}_R$ profile when σ_v increases (Figure 7.13a). The variation in the local-scale C_R also increases with σ_v, indicating a more heterogeneous solute distribution in the horizontal plane (Figure 7.13b). Because flow and transport become more heterogeneous as σ_v increases, more observations are needed to estimate $\hat{C}_R$ for $\sigma_v = 0.5$ than for $\sigma_v = 0.1$. The double peak in the variance profiles of Figure 7.13b is due to the relative minimum for this pulse application at around $x = 30\,cm$, where the highest concentration occurs.

For reactive solutes, the variability in the distribution coefficient (K_d) must also be considered. The field-scale resident concentration ($\hat{C}_R$) at $t = 5$ days is plotted versus depth in Figure 7.14 for perfect and no correlation between v and K_d. The figure shows that the negative correlation between v and K_d causes additional spreading. Such a negative correlation seems plausible since coarse-textured soils generally have a relatively high v, due to their higher hydraulic conductivity, and a small K_d, due to their lower exchange capacity. The opposite may hold for fine-textured soils. Robin et al. (1991) reported a weak negative correlation between K_d and the saturated hydraulic conductivity for a sandy aquifer.
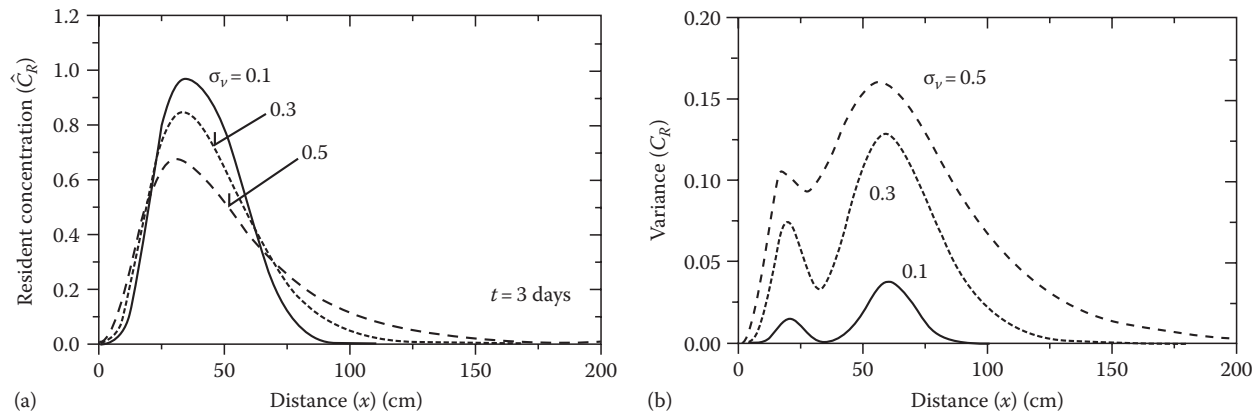
FIGURE 7.13 The effect of variability in the pore-water velocity, v, on (a) the field-scale resident concentration profile, and (b) the distribution of the variance for $\hat{C}_R$ in the horizontal plane; for a two-dimensional application of a nonreactive solute ($R = 1$) assuming $<v> = 20\,$cm day^{-1} and $D = 20\,$cm^2 day^{-1}.
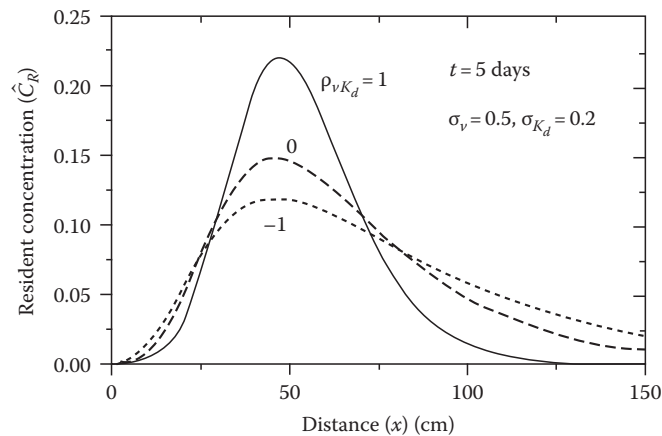
FIGURE 7.14 The field-scale resident concentration profile, $\hat{C}_R(x)$, for three different correlations between v and K_d with $<v> = 50\,$cm day^{-1}, $D = 20\,$cm^2 day^{-1}, $<K_d> = 1\,$g^{-1} cm^3, $\sigma_{Kd} = 0.2$, $<R> = 5$, and $\rho_b/\theta = 4\,$g cm^{-3}.

7.5 Nonaqueous Phase Liquids

Contamination of soil and water resources by nonaqueous phase liquids (NAPLs) has received considerable attention over the past few decades. These organic liquids have a low aqueous solubility. They are classified according to density as LNAPLs and DNAPLs, that is, lighter and denser than water. Oil and gasoline represent LNAPLs while many DNAPLs are chlorinated solvents used as cleaning (degreasing) agents or in manufacturing processes. NAPLs may be accidentally released to the soil through spillage during transport and handling operations or leakage from underground storage tanks, or they could have been disposed in the past when there was less awareness about any risks. NAPLs move downward through the vadose zone due to gravity as a separate liquid phase. A portion of the liquid will remain entrapped in the soil pores due to capillary forces forming isolated ganglia known as residual NAPL. The contaminant may partition among the solid, water, and air phases. The remaining liquid will eventually reach the water table. LNAPLs will accumulate on the water table and in the capillary fringe and, if a sufficient amount is spilled, the contaminant will spread on the water table. DNAPLs, on the other hand, will penetrate across the groundwater table and through the saturated porous medium if the spill volume is sufficient to feed the flow. Further vertical migration will be halted if the penetrating liquid encounters layers of fine-grained material such as silt or clays. In this case, the DNAPL will deflect horizontally until it reaches a more permeable vertical path or it forms a pool. Most NAPLs are slightly soluble in water and, either as entrapped in the soils or as a pool, will constitute a long-term source of groundwater pollution. The slow degradation and the extreme toxicity of many NAPLs found in groundwater make them a persistent and serious threat to human health.

Transport of NAPLs in porous media is a complex multiphase, multidimensional, and multicomponent problem (Willson et al., 2006). Further details pertaining to the multiphase transport of NAPL compounds can be found in Bear and Verruijt (1987), Miller et al. (1998), and Charbeneau (2000). Several researchers have developed mathematical models for multiphase flow and transfer between nonaqueous and aqueous phase (e.g., Kueper and Frind, 1988; Huyakorn et al., 1994; Imhoff et al., 2003; Willson et al., 2006; Fagerlund et al., 2008). Numerical models such as TOUGH2 (Pruess et al., 1999) and STOMP (White and Oostrom, 2000) are available to describe the multiphase transport of NAPLs in saturated and unsaturated media. The concern here is dissolution of static NAPL from both residual ganglia and pools and subsequent transport of the dissolved constituents in the aqueous phase by the same mechanisms of a (generic) solute that were previously reviewed.

7.5.1 Dissolution Process

Schwarzenbach et al. (1993) define dissolution as a thermodynamic process by which the molecules of an organic solute in its pure liquid phase are transferred into water. When an organic liquid is in contact with pure water, some of the organic

molecules will transfer from the organic into the water phase while some water molecules will transfer from the water into the organic phase. Upon equilibrium, the concentration of the organic molecules in the water phase equals the aqueous solubility of the compound. Partitioning of NAPL pollutants in the water phase has been described according to the local equilibrium assumption. Hence, if the organic concentration is known for one phase, it may be inferred for other phases by equilibrium partitioning relationships.

Several early laboratory experiments (van der Waarden et al., 1971; Dietz, 1978; Duffy et al., 1980) suggested the validity of the local equilibrium assumption in groundwater systems. Other studies however (Feenstra and Coburn, 1986; Sitar et al., 1987; Mercer and Cohen, 1990) note that observed concentrations can be several orders of magnitude lower than the aqueous solubility suggesting that no equilibrium conditions exist. Powers et al. (1991) found that nonequilibrium mass transfer conditions must be assumed for small contact times, large NAPL blobs, and low NAPL saturations. In terms of groundwater contamination, this implies that the duration of contamination and the volume of contaminated groundwater are much larger than that predicted by the local equilibrium assumption (Mackay et al., 1985). Powers et al. (1992) pointed out that the lower observed concentrations may be due to rate-limited mass transfer between the NAPL and the aqueous phase or to a bypassing of the water flow around the low permeability area contaminated with NAPL ganglia (Kueper and Frind, 1988). Other causes could be NAPL morphology and available surface contact area (Cho and Annable, 2005; Dobson et al., 2006; Zhao and Ioannidis, 2007), heterogeneities in the aquifer (Miller et al., 1998), or multicomponent dissolution effects (Geller and Hunt, 1993; Garg and Rixey, 1999; Brahma and Harmon, 2003). Several studies have attempted to quantify the dissolution kinetics and the NAPL–water interfacial area using photography of frozen or polymerized NAPL blobs (Powers et al., 1992; Kennedy and Lennox, 1997; Jia et al., 1999), gamma radiation (Brusseau et al., 2002), magnetic resonance (Baldwin and Gladden, 1996; Zhang et al., 2007), and x-ray microtomography (Brusseau et al., 2008).

7.5.1.1 Entrapped NAPL

In view of the inadequacy of the local equilibrium assumption, several researchers have further examined rate-limited dissolution of NAPL in saturated porous media. Most of the studies dealt with NAPL at residual saturation (Miller et al., 1990; Powers et al., 1991, 1992; Geller and Hunt, 1993; Sahloul and Chatzis, 2002). A detailed discussion of the residual NAPL dissolution mechanisms in porous media can be found in Powers et al. (1991, 1992) and in Khachikian and Harmon (2000). Dissolution from a single-component source may be described with a linear driving force model, and dissolution of an organic species from a multicomponent source will be more complicated to model. For a single component, the mass flux from the organic to the aqueous phase is proportional to the difference between the solubility of the compound and its actual aqueous concentration (Powers et al. 1992). The proportionality factor k_f is often referred to as the mass transfer coefficient, that is:

$$J = k_f(C_s - C), \qquad (7.107)$$

where
J is the mass flux [M L^{-2}T^{-1}]
C_s [M L^{-3}] is the solubility of the compound
C [M L^{-3}] is the concentration in the aqueous phase
k_f is the mass transfer coefficient [L T^{-1}]

Imhoff et al. (2003) pointed out that the mass transfer coefficient depends on the morphology of the entrapped NAPL as determined by pore structure (Powers et al., 1992) and wettability (Bradford et al., 2003), the velocity in the aqueous phase, viscosity, density (Miller et al., 1990), and NAPL saturation (Imhoff et al. 1994). The mass transfer coefficient may be predicted from regression equations, and interfacial area may be inferred or measured (Powers et al., 1991; Geller and Hunt, 1993; Bradford and Leij, 1997; Dobson et al., 2006). A lumped mass transfer rate constant equal to the product of the specific interfacial area and the mass transfer coefficient may also be used (Miller et al., 1990; Powers et al., 1992; Imhoff et al., 1994). Various correlations have been developed, which relate dimensionless groups, such as the Reynolds, Schmidt, and Peclet numbers, and residual NAPL saturation to the Sherwood number, which represents the dimensionless concentration gradient at the interface. Once the total mass transfer flux is evaluated, the term is incorporated into the classical ADE.

7.5.1.2 NAPL Pools

The description of the dissolution of NAPLs from a planar source (pools) poses special challenges. The low cross-sectional area with respect to pool volume implies slow dissolution. The merits of different conditions at the interface of pool and porous medium have been reviewed for mass transfer into pure fluids (Sherwood et al., 1975). A comparison of boundary conditions for dissolution in porous media has been lacking, in part because of a paucity of experimental data (Schwille, 1988; Voudrias and Yeh, 1994; Whelan et al., 1994; Dela Barre, 1999).

For a NAPL pool located at the water table (LNAPL) or an effectively impermeable layer (DNAPL), a concentration boundary layer develops at the organic–water interface. The thickness of the concentration boundary layer depends on the interstitial groundwater velocity, hydrodynamic dispersion, and pool size (Dela Barre, 1999). After a hydrodynamic boundary layer is established for water flow, a concentration gradient develops. In general, the NAPL–water interface and the geometry of the porous medium are difficult to determine accurately, and the dissolution process at the macroscopic scale is described via mass transfer coefficients. For flow over a NAPL pool, the local mass transfer coefficient decreases with distance from the front end of the pool and has a maximum value at the upstream edge (Chrysikopoulos and Lee, 1998). However, due to the complexity

of the subsurface environment and associated scale dependency of modeling parameters, the NAPL dissolution process is commonly described with an average mass transfer coefficient or rate constant.

Johnson and Pankow (1992) used experimental results from Schwille (1988) and the analytical solution from Hunt et al. (1988) to calculate a surface-area-averaged mass transfer rate whose value decreases as the length of the pool increases to reproduce the decreasing dissolution rate across the top of the pool. The mass transfer term was then included in the steady-state ADE, which was solved analytically for a homogeneous semi-infinite medium with a square pool of negligible thickness. Local equilibrium was assumed by selecting the aqueous solubility as boundary condition at the pool–water interface while the remaining boundary conditions away from the pool where of the Dirichlet type with fixed zero concentration.

Holman and Javandel (1996) also used the local equilibrium assumption in the transient two-dimensional ADE that describes the dissolution and transport of a mixture of slightly soluble compounds from a LNAPL pool into the groundwater. Their model describes the dissolution and transport of the contaminant as a heterogeneous reaction involving three steps: molecular diffusion of the compound inside the pool toward the LNAPL–water interface, dissolution of the compound at the interface, and transport of the dissolved compound from the interface into the aquifer.

Chrysikopoulos et al. (1994) developed a mathematical model for dissolution of a single-component rectangular DNAPL pool of negligible thickness in a two-dimensional semi-infinite homogeneous porous medium with steady-state flow, equilibrium sorption, and first-order decay. The boundary condition at the DNAPL–water interface was a second-type boundary condition. It is represented by a mass transfer relationship in which the fixed diffusive flux is related to the saturated aqueous concentration through a time-varying mass transfer coefficient. Lee and Chrysikopoulos (2002) applied the same boundary conditions to the analysis of the dissolution of a circular trichloroethylene (TCE) plume. The authors developed correlations between the time invariant overall Sherwood number and the Peclet number, which were found in good agreement with the experimental data.

Leij et al. (2000) used the ADE to model the three-dimensional equilibrium solute transport in a homogeneous porous medium generated from the dissolution of planar and parallelepipedal sources of persistent contamination. Analytical solutions were formulated based upon Green's functions for a first-, second-, and third-type boundary condition in the case of a planar source. A third-type or Cauchy-type boundary condition, which imposes a fixed total flux (diffusive and advective) at the DNAPL–water interface, more realistically predicts the concentration profiles compared with either the first-type or fixed concentration or a second-type boundary condition. Leij and van Genuchten (2000) extended Green's function method to obtain analytical solutions for nonequilibrium transport from a rectangular pool. The authors observed that the solute flux from the

pool, and therefore the concentration in the medium, depends strongly on the mass transfer coefficient.

7.5.2 Numerical Modeling

Numerical modeling of NAPL dissolution in saturated porous media has been widely investigated in the past decades. Several numerical techniques have been applied in order to enhance the performance of computer codes designed to simulate contaminant transport in the subsurface. A distinction is again made between dissolution from NAPL blobs entrapped in the pores of a saturated porous medium and from a DNAPL pool at the bottom of a saturated porous medium.

In the first case, the NAPL blobs are incorporated in the domain of interest. Therefore, the dissolution term, expressed by a first-order rate-limited relationship used by Geller and Hunt (1993) and Powers et al. (1991), among others, will be added to the ADE as a point source term. In the second case, the pool is just outside the solution domain, and NAPL dissolution is accounted for by formulating the appropriate boundary conditions.

Examples of numerical models developed to describe groundwater plumes emanating from entrapped NAPL in the subsurface can be found in Guiguer (1991) and Clement et al. (2004). Christ et al. (2009) simulated dissolution of a DNAPL (tetrachloroethylene or PCE) and mass recovery in two- and three-dimensional nonuniform permeability fields using a modified version of the popular MT3D numerical code to assess the influence of model dimensionality on the mass recovery predictions.

The numerical simulation of NAPL blobs dissolving in a saturated porous medium does not present particular difficulties when the mass transfer coefficient is considered independent of the aqueous concentration. In this case, the governing partial differential equation is linear and its solution can be achieved by using one of the numerical techniques available to solve the classical ADE. For dispersion–diffusion-dominated problems, the classical Eulerian approach either by using finite difference techniques or using Galerkin finite element methods on a fixed grid in space has performed quite satisfactorily. For advection-dominated transport problems and for pure advection problems, however, the Eulerian methods produce inaccurate solutions due to oscillation and numerical dispersion. As mentioned in Section 7.3, a way to obviate this problem is to use refined grids or higher order finite difference and finite element methods or, alternatively, to use the upstream weighted finite difference or finite element method (Settari and Aziz, 1975; Dalen, 1979). Lagrangian and mixed Lagrangian–Eulerian approaches have also been popular to solve advection-dominated transport problems (Huyakorn and Pinder, 1983). The method of characteristics and the random walk method belong to these classes of techniques. Yeh (1990) proposed an Eulerian–Lagrangian method with a zoomable hidden fine-mesh approach, which eliminates the numerical oscillations and reduces the numerical dispersion. A variety of approaches exist to combine numerical methods with analytical solutions (e.g., Hwang et al., 1985).

When the contamination in the aquifer is produced by the dissolution of NAPL pools, the process is described by the classical two or three-dimensional ADE where the presence of the pool is accounted for in the boundary conditions. This problem is also known as the patch-source problem. According to Segol (1994), the term "refers to the transport of a solute from a boundary source of finite extent into a rectangular domain subjected to a uniform, unidirectional velocity field."

The boundary conditions applied to simulate the source term are the Dirichlet or first-type boundary condition, the Neuman- or second-type boundary condition, or the Cauchy- or third-type boundary condition. Numerically, the problem of modeling solute transport with a dissolving source on the boundary is somewhat different than dissolution from a source inside the domain of integration. Segol (1994) reviewed several studies where the patch-source problem has been used to illustrate numerical techniques developed to solve the convection–dispersion equation. Daus and Frind (1985) formulated the patch-source problem using either Dirichlet or Cauchy boundary conditions. The governing equation is then solved numerically by the alternating direction Galerkin technique, which consists in splitting the governing equation into two one-dimensional ADEs coupled with cross derivatives terms. For both boundaries, the technique produces accurate results, and it is more efficient than the classical finite-element method due to the decoupling procedure. The cell analytical–numerical method was successfully applied by Elnawawy et al. (1990) to the patch-source problem with a Dirichlet boundary condition at the source. The results showed to be accurate but were obtained at the expenses of geometrical flexibility. Leismann and Frind (1989) solved the patch-source problem by applying the symmetric-matrix time integration technique. They used the Dirichlet-type boundary condition at the source. The method applied by the authors uses an integration technique for the time derivative, which produces an algebraic system of equations whose coefficient matrix is symmetric and positive definite. In this case, the resulting algebraic system can be efficiently solved. Sudicky (1989) used the patch-source problem with the source described by Dirichlet-type boundary conditions to illustrate the Laplace transform Galerkin method. In this method, the Laplace transform is first applied to the ADE to eliminate the temporal derivatives. The transformed equation is then solved for a discretized spatial domain using the standard Galerkin finite element procedure. Finally, the inversion of the transformed concentration values is obtained using Crump's method. According to the author, the method has the advantage of producing accurate and robust solutions, but it is restricted to linear mass transport in a steady flow field.

In the four studies reviewed by Segol (1994), the source term is perpendicular to the flow direction. Lee and Chrysikopoulos (1995) used a Neuman or second-type boundary condition for dissolution from a multicomponent DNAPL pool at the bottom of an aquifer with steady flow. The boundary condition corresponding to the pool location is described by a relationship in which the diffusive flux is specified using a time- and position-dependent mass transfer coefficient. The numerical scheme adopted by the authors to solve the governing equation is the classical implicit finite difference approximation, which yielded a band diagonal system of algebraic equations.

7.5.3 Remediation

Over the past decades, researchers have aimed to develop efficient techniques to remediate soil and groundwater. The simplest and commonly applied technique is known as pump-and-treat, where contaminated groundwater is extracted and treated above ground. This technique promotes a decline in contaminant concentrations over time, but it requires the treatment of large volumes of water. A combination of optimization methods and simulation models has been proposed to determine more efficient cleanup design strategies (Gorelick et al., 1993). In particular, these optimization techniques have been developed to determine the location and number of pumping wells, the optimal pumping rate (which will maximize the contaminant removal at the lowest energy cost) and treatment strategies.

Among the remediation techniques are hot-water flushing (Imhoff et al., 1995; O'Carroll and Sleep, 2007), air sparging (Adams and Reddy, 2000), cosolvent flushing (Imhoff et al., 1995), the use of surfactants (Childs et al., 2006), and in situ bioremediation (Zoller and Rubin, 2001). The effectiveness of the remediation may be substantially improved if the location and extent of the NAPL source are known.

7.5.4 Inverse Modeling to Locate Contaminant Sources

The problem of defining the location of the DNAPL pool and the release history of the contaminant (the time period and the mass at which the contaminant was released) is important to design of the most efficient cleanup strategy but also from a legal and regulatory point of view. Remediating a contaminated soil or aquifer is typically costly and cumbersome. Therefore, as Skaggs and Kabala (1994) noted, a model capable of backtracking a contaminant plume in space and time would help to individuate the responsible parties and to distribute the cleanup costs accordingly. Unfortunately, the dispersive nature of the contaminant transport problem makes it an irreversible problem. Modeling the contaminant release history by simply reversing time will result in an ill-posed problem with a nonunique or unstable solution (Skaggs and Kabala, 1994).

Another obstacle in the development of a model to identify contaminant release history is the scarcity of historical data. Due to this lack of data, reconstruction must be based on observations of the existing contaminant plume. Usually, an extensive number of observations are required in order to obtain a detailed description of the concentration profiles. Recently, several parameter identification techniques coupled with numerical models have been developed to minimize the number of locations (wells) where the NAPL concentration needs to be observed (Yeh and Sun, 1984; McCarthy and Yeh, 1990).

Several researchers have dealt in the past with the problem of contaminant source characterization (Gorelick et al., 1983; Skaggs and Kabala, 1994). Much emphasis has been given to the temporal reconstruction of the source release while the problem of locating existing DNAPL contaminant sources has been neglected, especially for those cases involving heterogeneous flow and transport fields. Hwang and Koerner (1983) explored the performance of two different approaches: nonsequential and sequential optimization. In the former the optimization process is applied to the partial differential equation describing solute transport. In the sequential optimization approach, the partial differential equation is discretized and a linear dynamic system is developed that may subsequently be optimized. Hwang et al. (1988) applied their methods to two field cases with pollution due to surface ponding of organic waste and a landfill. In both cases the estimated source location was relatively close to the actual location but the procedure was limited to two-dimensional geometries.

Gorelick et al. (1983) combined least-square regression and linear programming to obtain the least absolute error to identify the location, magnitude, and disposal period of underground pollutant point sources. The authors examined the case of steady and transient scenarios. The solution procedure was based on the response matrix approach, which assumes a linear system response. Their results showed that at steady state both the linear programming and the multiple regression formulations produced accurate results only when data were relatively noise free. In the transient case, both procedures were successful in estimating the location and magnitude of pollutant sources due to the abundance of concentration data.

Uffink (1989) and Bagtzoglou et al. (1992) applied the Random-walk method, which is based on reversing the advective part of the transport model while the dispersive part is left positive. The results provide a probability distribution of the contaminant initial position, but the approach is not able to define a contaminant release history. Simultaneous estimation of aquifer parameters and a distributed pollutant source term was attempted by Wagner (1992) by combining nonlinear maximum likelihood estimation with groundwater flow and transport models.

Sidauruk et al. (1998) presented an inverse method based on analytical solutions of contaminant transport problems to estimate mass transport parameters such as dispersion coefficients and flow velocities as well as the amount of pollutant, its initial location, and time origin. The inverse procedure is based upon the method of correlation coefficient optimization. By minimizing the correlation coefficient of the linear regression, the parameters of interest are determined by explicit formulas similar to linear regression equations. The method was applied to the two-dimensional cases of a plume generated by instantaneous and continuous point sources. As the authors pointed out, the method is restricted to simple problems involving a homogeneous porous medium, and simple geometries and flow fields.

Mahar and Datta (1997) combined optimal groundwater quality monitoring network design and point-source identification algorithms to estimate the magnitude, location, and active duration of a pollutant source using a three-step procedure.

In the first step, a nonlinear optimal source identification model with embedded flow and transport simulation constraints is employed to identify source fluxes using concentration data for arbitrary locations. The results from this preliminary step are then perturbed statistically, to account for uncertainties in measurements, and incorporated into a groundwater monitoring network design model with an integer-programming-based optimization model. The concentration data generated by the monitoring network are then fed back to the source identification model to obtain more accurate predictions.

Skaggs and Kabala (1994) applied Tikonov regularization in one-dimensional numerical experiments to recover the release history of a contaminant originated from a known site. Their method was found to be very sensitive to observation errors. In a subsequent paper, Skaggs and Kabala (1995) attempted to solve the same problem by applying the method of quasi-reversibility to the convection–dispersion equation. The method was preferable to Tikonov regularization due to its simple implementation and its capacity to include nonconstant flow and transport parameters, although it was not as accurate.

Snodgrass and Kitanidis (1997) proposed Bayesian analysis and a geostatistical approach to the problem. Their method provides a measure of the error associated with the best estimate of the release history and helps to identify sources of uncertainties. One limitation is that the location of the contaminant source has to be known a priori. The same assumption was made by Kondisetty et al. (1996) in an inverse optimization approach to analytically track instantaneous point-source releases in the saturated zone. The inverse approach, based on the minimization of the error in the difference between calculated data and observations, was coupled with a two-dimensional transport model using the method of characteristics.

An attempt to extend the release history problem to three-dimensional domains and spatially distributed sources was presented by Woodbury et al. (1998). The authors applied the minimum entropy method to recover the source-release history of a three-dimensional plume generated by a rectangular first-type patch source. Their work is carried out in an analytical framework, and it is limited to steady flow in a uniform porous medium. Liu and Ball (1999) applied inverse modeling to estimate the contaminant plume history from observed concentrations at a field site that was being remediated. The authors estimated the contaminant boundary concentration by assuming a simple mass-conserving two-layer diffusion model. They used least-square minimization where the objective function was regularized by adding a term to smooth the boundary concentration.

Sciortino et al. (2000) developed an inverse model procedure to determine the size and the location of a single-component DNAPL pool in a saturated homogeneous porous medium for steady flow. The inverse problem was solved using least-squares minimization with a search procedure based on the Levenberg–Marquardt method. A three-dimensional analytical model was used to describe the transport of solute from a dissolving DNAPL pool. The algorithm was tested to predict the location and size of a DNAPL pool placed in a controlled three-dimensional

bench-scale experiment. The Levenberg–Marquardt algorithm was used to minimize (1) ordinary residuals, (2) residuals with weights equal to the square of the inverse of the observations, and (3) residuals with weights obtained by adding a constant term to the observed concentrations. The results were found to be sensitive to the location of the observation wells and to the type of residuals minimized. The inverse problem is nonunique and nonconvex even in the absence of observation errors.

Neupauer and Wilson (1999, 2001) developed a backward location and travel-time probabilistic model to characterize source of groundwater contamination and their prior location. The authors derived the governing equation of the backward probabilities using the adjoints of the ADE modified with the addition of a load term that depends on both the type of probability that it describes (location or travel time) and the type of well (pumping or monitoring) used to detect pollution. Michalak and Kitanidis (2004) proposed a geostatistical approach combined with the adjoint state method that was able to reconstruct the historical multidimensional contaminant distribution in porous media. Recently, Yeh et al. (2007) presented the SATS-GWT approach, which consists of a combination of simulated annealing (SA), tabu search (TS), and a three-dimensional groundwater flow and solute transport model (MODFLOW-GWT). Source location, release concentration, and release period of groundwater contaminants may be estimated using the proposed model. The authors applied their model to different contamination scenarios, and they claim that the approach yields accurate estimates, even in the presence of measurement errors.

References

Adams, J.A., and K.R. Reddy. 2000. Removal of dissolved- and free-phase benzene pools from ground water using in situ air sparging. J. Environ. Eng. 126:697–707.

Ahlstrom, S.W., H.P. Foote, R.C. Arnett, C.R. Cole, and R.J. Serne. 1977. Multicomponent mass transport model theory and numerical implementation. Battelle Pacific Northwest Laboratory Report BNWL-2127, Richland, WA.

Aitcheson, J., and J.A.C. Brown. 1963. The lognormal distribution. Cambridge University Press, London, U.K.

Amoozegar-Fard, A., D.R. Nielsen, and A.W. Warrick. 1982. Soil solute concentration distribution for spatially varying pore water velocities and apparent diffusion coefficients. Soil Sci. Soc. Am. J. 46:3–9.

Aris, R. 1956. On the dispersion of a solute in a fluid flowing through a tube. Proc. Roy. Soc. Lond. Ser. 235:67–77.

Aris, R. 1958. On the dispersion of linear kinematic waves. Proc. Roy. Soc. Lond. Ser. 245:268–277.

Bagtzoglou, A.C., D.E. Dougherty, and A.F.B. Tompson. 1992. Application of particle methods to reliable identification of groundwater pollution sources. Water Resour. Manag. 6:15–23.

Baldwin, C.A., and L.F. Gladden. 1996. NMR imaging of non-aqueous-phase liquid dissolution in a porous medium. AIChE J. 42:1341–1349.

Bear, J., and A. Verruijt. 1987. Modeling groundwater flow and pollution. Kluwer Academic Publishers, Norwell, MA.

Boyd, G.E., A.W. Adamson, and L.S. Meyers. 1947. The exchange adsorption of ions from aqueous solutions by organic zeolites. II. Kinetics. J. Am. Chem. Soc. 69:2836–2848.

Bradford, S., and F.J. Leij. 1997. Estimating interfacial areas for multi-fluid soil systems. J. Contam. Hydrol. 27:83–105.

Bradford, S.A., K.M. Rathfelder, J. Lang, and L.M. Abriola. 2003. Entrapment and dissolution of DNAPLs in heterogeneous porous media. J. Contam. Hydrol. 67:133–157.

Brahma, P.P., and T.C. Harmon. 2003. The effect of multicomponent diffusion on NAPL dissolution from spherical ternary mixtures. J. Contam. Hydrol. 67:43–60.

Brenner, H. 1962. The diffusion model of longitudinal mixing in beds of finite length. Numerical values. Chem. Eng. Sci. 17:229–243.

Bresler, E. 1973. Simultaneous transport of solutes and water under transient unsaturated flow conditions. Water Resour. Res. 9:975–986.

Bresler, E., and G. Dagan. 1981. Convective and pore scale dispersive solute transport in unsaturated heterogeneous fields. Water Resour. Res. 17:1683–1693.

Bresler, E., and R.J. Hanks. 1969. Numerical method for estimating simultaneous flow of water and salt in saturated soils. Soil Sci. Soc. Am. Proc. 33:827–832.

Brusseau, M.L., H. Janousek, A. Murao, and G. Schnaar. 2008. Synchrotron X-ray microtomography and interfacial partitioning tracer test measurements of NAPL–water interfacial areas. Water Resour. Res. 43:3619–3625.

Brusseau, M.L., Z. Zhang, N.T. Nelson, R.B. Cain, G.R. Tick, and M. Oostrom. 2002. Dissolution of nonuniformly distributed immiscible liquid: Intermediate-scale experiments and mathematical modeling. Environ. Sci. Technol. 36:1033–1041.

Campbell, G.S. 1985. Soil physics with BASIC: Transport models for soil–plant systems. Elsevier, New York.

Carslaw, H.S., and J.C. Jaeger. 1959. Conduction of heat in solids. Clarendon Press, Oxford, U.K.

Charbeneau, R. 2000. Groundwater hydraulics and pollutant transport. Prentice Hall, Upper Saddle River, NJ.

Chaudhari, N.M. 1971. An improved numerical technique for solving multidimensional miscible displacement. Soc. Pet. Eng. J. 11:277–284.

Childs, J., E. Acosta, M.D. Annable, M.C. Brooks, C.G. Enfield, J.H. Harwell, M. Hasegawa, R.C. Knox, P.S.C. Rao, D.A. Sabatini, B. Shiau, E. Szekeres, and A.L. Wood. 2006. Field demonstration of surfactant-enhanced solubilization of DNAPL at Dover Air Force Base, Delaware. J. Contam. Hydrol. 82:1–22.

Cho, J., and M.D. Annable. 2005. Characterization of pore scale NAPL morphology in homogeneous sands as a function of grain size and NAPL dissolution. Chemosphere 61:899–908.

Christ, J.A., L.D. Lemke, and L.M. Abriola. 2009. The influence of dimensionality on simulations of mass recovery from non-uniform dense non-aqueous phase liquid (DNAPL) source zones. Adv. Water Resour. 32:401–412.

Chrysikopoulos, C.V., and K.Y. Lee. 1998. Contaminant transport resulting from multicomponent nonaqueous phase liquid pool dissolution in three-dimensional subsurface formations. J. Contam. Hydrol. 31:1–21.

Chrysikopoulos, C.V., E.A. Voudrias, and M.M. Fyrillas. 1994. Modeling of contaminant transport resulting from dissolution of nonaqueous phase liquid pools in saturated porous media. Transp. Porous Media 16:125–145.

Clement, T.P., Y.-C. Kim, T.R. Gautam, and K.-K. Lee. 2004. Experimental and numerical investigation of DNAPL dissolution processes in a laboratory aquifer model. Ground Water Monit. Remediat. 24:88–96.

Coats, K.H., and B.D. Smith. 1964. Dead end pore volume and dispersion in porous media. Soc. Pet. Eng. J. 4:73–84.

Crank, J. 1975. The mathematics of diffusion. Clarendon Press, Oxford, U.K.

Dagan, G., and E. Bresler. 1979. Solute dispersion in unsaturated heterogeneous soil at field-scale. I. Theory. Soil Sci. Soc. Am. J. 43:461–467.

Dalen, V. 1979. Simplified finite-element models for reservoir flow problems. Soc. Pet. Eng. J. 19:333–343.

Daus, A.D., and E.O. Frind. 1985. An alternating direction Galerkin technique for simulation of contaminant transport in complex groundwater systems. Water Resour. Res. 21:653–664.

Dela Barre, B.K. 1999. Mass transfer coefficient estimation for dense, nonaqueous phase liquid pool dissolution using a three-dimensional physical aquifer models. Ph.D. Dissertation. University of California. Los Angeles, CA.

Dietz, D.N. 1978. Large scale experiments on groundwater pollution by oil spills-interim results, p. 253–265. *In* International symposium on groundwater pollution by oil hydrocarbons. International Association of Hydrogeologists, Prague, Czechoslovakia.

Dobson, R., M.H. Schroth, M. Oostrom, and J. Zeyer. 2006. Determination of NAPL–water interfacial areas in well-characterized porous media. Environ. Sci. Technol. 40:815–822.

Donea, J. 1991. Generalized Galerkin methods for convection dominated transport phenomena. Appl. Mech. Rev. 44:205–214.

Duffy, J.J., E. Peake, and M.F. Mohtadi. 1980. Oil spills on land as potential source of groundwater contamination. Environ. Int. 3:107–120.

Dykhuizen, R.C., and W.H. Casey. 1989. An analysis of solute diffusion in rocks. Geochim. Cosmochim. Acta 53:2797–2805.

Elnawawy, O.A., A. Valocchi, and A.M. Ongonag. 1990. The cell analytical–numerical method for solution of the advection–dispersion equation: Two-dimensional problems. Water Resour. Res. 26:2705–2716.

Epstein, N. 1989. On tortuosity and the tortuosity factor in flow and diffusion through porous media. Chem. Eng. Sci. 44:777–780.

Fagerlund, F., A. Niemi, and T.H. Illangasekare. 2008. Modeling of nonaqueous phase liquid (NAPL) migration in heterogeneous saturated media: Effects of hysteresis and fluid immobility in constitutive relations. Water Resour. Res. W03409.

Feenstra, S., and J. Coburn. 1986. Subsurface contamination from spills of denser than water chlorinated solvents. Calif. WPCA Bull. 23:26–34.

Fischer, H.B., E. List, R.C.Y. Koh, J. Imberger, and N.H. Brooks. 1979. Mixing in inland and coastal waters. Academic Press, New York.

Flury, M., and T.F. Gimmi. 2002. Solute diffusion. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis: Physical methods. Soil Sci. Soc. Am., Madison, WI.

Fried, J.J., and M.A. Combarnous. 1971. Dispersion in porous media. Adv. Hydrosci. 9:169–282.

Garg, S., and W.G. Rixey. 1999. Dissolution of benzene, toluene, *m*-xylene and naphthalene from a residually trapped nonaqueous phase liquid under mass transfer limited conditions. J. Contam. Hydrol. 36:313–331.

Gautschi, W. 1964. Error function and Fresnel integrals, p. 295–329. *In* M. Abramowitz and I.A. Stegun handbook of mathematical functions. Appl. Math. Ser. 55. National Bureau of Standards, Washington, DC.

Gelhar, L.W., C. Welthy, and K.R. Rehfeldt. 1992. A critical review of data on field-scale dispersion in aquifers. Water Resour. Res. 28:1955–1974.

Geller, J.T., and J.R. Hunt. 1993. Mass transfer from nonaqueous phase organic liquids in water-saturated porous media. Water Resour. Res. 29:833–846.

Goldstein, S. 1953. On the mathematics of exchange processes in fixed columns. I. Mathematical solutions and asymptotic expansions. Proc. Roy. Soc. Lond. Ser. A 219:151–185.

Gorelick, S.M., B. Evans, and I. Remson. 1983. Identifying sources of groundwater pollution: An optimization approach. Water Resour. Res. 19:779–790.

Gorelick, S.M., R.A. Freeze, D. Donhoue, and J.F. Keely. 1993. Groundwater contamination: Optimal capture and containment. Lewis Publishers, Boca Raton, FL.

Gray, W.G., and G.F. Pinder. 1976. An analysis of the numerical solution of the transport equation. Water Resour. Res. 12:547–555.

Guiguer, N. 1991. Numerical modelling of the fate of residual immiscible fluids in saturated porous media. Water Sci. Technol. 24:261–270.

Gupta, V.K., and R.N. Bhattacharya. 1983. A new derivation in the Taylor–Aris theory of solute dispersion in a capillary. Water Resour. Res. 19:945–951.

Harvie, C.E., and J.H. Weare. 1980. The prediction of mineral solubilities in natural waters: The Na-K-Mg-Ca-Cl-SO$_4$-H$_2$O system from zero to high concentration at 25°C. Geochim. Cosmochim. Acta 44:981–997.

Helfferich, F. 1962. Ion exchange. McGraw-Hill, New York.

Holman, H.-Y.N., and I. Javandel. 1996. Evaluation of transient dissolution of slightly water-soluble compounds from a light nonaqueous phase liquid pool. Water Resour. Res. 32:915–923.

Huang, K., and M.Th. van Genuchten. 1995. An analytical solution for predicting solute transport during ponded infiltration. Soil Sci. 159:217–223.

Huang, K., J. Šimunek, and M.Th. van Genuchten. 1997. A third-order numerical scheme with upwind weighting for modeling solute transport. Int. J. Numer. Methods Eng. 40:1623–1637.

Hunt, J.R., N. Sitar, and K.S. Udell. 1988. Nonaqueous phase liquid transport and cleanup. 2. Experimental studies. Water Resour. Res. 24:1259–1269.

Huyakorn, P.S., S. Panday, and Y.S. Wu. 1994. Three-dimensional multiphase flow model for assessing NAPL contamination in porous and fractured media, 1. Formulation source. J. Contam. Hydrol. 16:109–130.

Huyakorn, P.S., and G.F. Pinder. 1983. Computational methods in subsurface flow. Academic Press, New York.

Hwang, J.C., A. Ayubcha, S.-H. Chien, and S. Richardson. 1988. Feasibility studies of groundwater pollution source identification from actual field monitoring well data. Contaminated groundwater control. Vol. III. HMCRI Hazardous Waste Technology Monographs, Silver Spring, MD.

Hwang, J.C., C.J. Chen, M. Sheikoslami, and B.K. Panigrahi. 1985. Finite analytical–numerical solution for two-dimensional groundwater solute transport. Water Resour. Res. 21:1354–1360.

Hwang, J.C., and R.M. Koerner. 1983. Groundwater pollution source identification from limited monitoring well data: Part 1—Theory and feasibility. J. Hazard. Mater. 8:105–119.

Imhoff, P.T., M.W. Farthing, and C.T. Miller. 2003. Modeling NAPL dissolution fingering with upscaled mass transfer coefficients. Adv. Water Res. 26:1097–1111.

Imhoff, P.T., S.N. Gleyzer, J.F. McBride, L.A. Vancho, I. Okuda, and C.T. Miller. 1995. Cosolvent-enhanced remediation of residual dense nonaqueous phase liquids: Experimental investigation. Environ. Sci. Technol. 29:1966–1976.

Imhoff, P.T., P.R. Jaffé, and G.F. Pinder. 1994. An experimental study of complete dissolution of a nonaqueous phase liquid in saturated porous media. Water Resour. Res. 30:307–320.

Istok, J. 1989. Groundwater modeling by the finite element method. American Geophysical Union, Washington, DC.

Jacobsen, O.H., F.J. Leij, and M.Th. van Genuchten. 1992. Parameter determination for chloride and tritium transport in undisturbed lysimeters during steady flow. Nord. Hydrol. 23:89–104.

Jia, C., K. Shing, and Y.C. Yortsos. 1999. Visualization and simulation of NAPL solubilization in pore networks. J. Contam. Hydrol. 35:363–387.

Jin, Y., and W.A. Jury. 1996. Characterizing the dependence of gas diffusion coefficient on soil properties. Soil Sci. Soc. Am. J. 60:66–71.

Johnson, R.L., and J.F. Pankow. 1992. Dissolution of dense chlorinated solvents into groundwater. 2. Source functions for pools and solvent. Environ. Sci. Technol. 26:896–901.

Jury, W.A., and K. Roth. 1990. Transfer functions and solute movement through soil. Theory and applications. Birkhäuser Verlag, Basel, Switzerland.

Kemper, W.D. 1986. Solute Diffusivity, p. 1007–1024. *In* A. Klute (ed.) Methods of soil analysis. I. Physical and mineralogical methods. Soil Science Society of America, Madison, WI.

Kennedy, C.A., and W.C. Lennox. 1997. A pore-scale investigation of mass transport from dissolving DNAPL droplets. J. Contam. Hydrol. 24:221–246.

Khachikian, C., and T.C. Harmon. 2000. Nonaqueous phase liquid dissolution in porous media: Current state of knowledge and research needs. Transp. Porous Media 38:3–28.

Klotz, D., K.P. Seiler, H. Moser, and F. Neumaier. 1980. Dispersivity and velocity relationship from laboratory and field experiments. J. Hydrol. 45:169–184.

Kondisetty, S.R., P.W. Jayawickrama, and K.A. Rainwater. 1996. Source characterization at contaminated sites, p. 927–943. Geotech. Spec. Publ. Vol. 58/2. ASCE, New York.

Konikov, L.F., and J.D. Bredehoeft. 1978. Computer model of two-dimensional solute transport and dispersion in groundwater, p. 2–94. *In* Techniques of water resources investigation, Book 7. U.S. Geological Survey, Reston, VA.

Kreft, A., and A. Zuber. 1978. On the physical meaning of the dispersion equations and its solutions for different initial and boundary conditions. Chem. Eng. Sci. 33:1471–1480.

Kreft, A., and A. Zuber. 1986. Comments on "Flux-averaged and volume-averaged concentrations in continuum approaches to solute transport" by J. C. Parker and M. Th. van Genuchten. Water Resour. Res. 22:1157–1158.

Krupp, H.K., J.W. Biggar, and D.R. Nielsen. 1972. Relative flow rates of salt and water in soil. Soil Sci. Soc. Am. Proc. 36:412–417.

Kueper, B.H., and E.O. Frind. 1988. Overview of immiscible fingering in porous media. J. Contam. Hydrol. 2:95–110.

Lassey, K.R. 1988. Unidimensional solute transport incorporating equilibrium and rate-limited isotherms with first-order loss. 1. Model conceptualizations and analytic solutions. Water Resour. Res. 3:343–350.

Lee, K.Y., and C.V. Chrysikopoulos. 1995. Numerical modeling of three-dimensional contaminant migration from dissolution of multicomponent NAPL pools in saturated porous media. Environ. Geol. 26:157–165.

Lee, K.Y., and C.V. Chrysikopoulos. 2002. Dissolution of a well-defined trichloroethylene pool in saturated porous media: Experimental results and model simulations. Water Res. 36:3911–3918.

Leij, F.J., and J.H. Dane. 1989. Determination of transport parameters from step and pulse displacement of cation and anions. Ala. Agric. Exp. Sta. Dept. of Agronomy and Soils Series 135.

Leij, F.J., and J.H. Dane. 1992. Moment method applied to solute transport with binary and ternary exchange. Soil Sci. Soc. Am. J. 56:667–674.

Leij, F.J., E. Priesack, and M.G. Schaap. 2000. Solute transport modeled with Green's functions with application to persistent solute sources. J. Contam. Hydrol. 41:155–173

Leij, F.J., and N. Toride. 1995. Discrete time- and length-averaged solutions of the advection–dispersion equation. Water Resour. Res. 31:1713–1724.

Leij, F.J., and M.Th. van Genuchten. 1995. Approximate analytical solutions for transport in two-layer porous media. Transp. Porous Media 18:65–85.

Leij, F.J., and M.Th. van Genuchten. 2000. Analytical modeling of nonaqueous phase liquid dissolution with Green's functions. Transp. Porous Media 38:141–166.

Leismann, H.M., and E.O. Frind. 1989. A symmetric-matrix time integration scheme for the efficient solution of advection–dispersion problems. Water Resour. Res. 25:1133–1139.

Li, S.-G., F. Ruan, and D. McLaughlin. 1992. A space-time accurate method for solving solute transport problems. Water Resour. Res. 28:2297–2306.

Lide, D.R. 1995. Handbook of chemistry and physics. CRC Press, Boca Raton, FL.

Lindstrom, F.T., and M.N.L. Narasimhan. 1973. Mathematical theory of a kinetic model for dispersion of previously distributed chemicals in a sorbing porous medium. SIAM J. Appl. Math. 24:469–510.

Lindstrom, F.T., and W.M. Stone. 1974. On the start up or initial phase of linear mass transport of chemicals in a water saturated sorbing porous medium. SIAM J. Appl. Math. 26:578–591.

Liu, C., and W.P. Ball. 1999. Application of inverse methods to contaminant source identification from aquitard diffusion profiles. Water Resour. Res. 35:1975–1985.

Mackay, D.M., P.V. Roberts, and J.A. Cherry. 1985. Transport of organic contaminants in groundwater. Environ. Sci. Technol. 19:364–392.

Mahar, S.P., and B. Datta. 1997. Optimal monitoring network and ground-water pollution source identification. J. Water Resour. Plan. Manag. 123:199–207.

McCarthy, J.M., and W.W.-G. Yeh. 1990. Optimal pumping test design for parameter estimation and prediction in groundwater hydrology. Water Resour. Res. 26:779–791.

Mercer, J.W., and R.M. Cohen. 1990. A review of immiscible fluids in the subsurface: Properties, models, characterization and remediation. J. Contam. Hydrol. 6:107–163.

Michalak, A.M., and P.K. Kitanidis. 2004. Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling. Water Resour. Res. 40, W08302, doi: 10.1029/2004WR003214.

Miller, C.T., M.M. Poirier-McNeill, and A. Mayer. 1990. Dissolution of trapped nonaqueous phase liquids: Mass transfer characteristics. Water Resour. Res. 26: 2783–2796.

Miller, C.T., C.S. Willson, S.N. Gleyzer, M.W. Farthing, J.F. McBride, and P.T. Imhoff. 1998. NAPL-aqueous phase mass transfer in heterogeneous porous media, p. 141–148. *In* Proceedings of the groundwater quality: Remediation and protection, Tubingen, Germany, September 1998. IAHS Publication Vol. 250.

Millington, R.J. 1959. Gas diffusion of porous solids. Science 130:100–102.

Millington, R.J., and J.P. Quirk. 1961. Permeability of porous solids. Trans. Faraday Soc. 57: 1200–1206.

Mordis, G.J., and D.L. Reddell. 1991. The Leplace transform finite difference method for simulation of flow through porous media. Water Resour. Res. 27:1873–1884.

Nauman, E.B., and B.A. Buffham. 1983. Mixing in continuous flow systems. John Wiley, New York.

Neuman, S.P. 1984. Adaptive Eulerian-Lagrangian finite element method for advection-dispersion. Int. J. Numer. Methods Eng. 20:321–337.

Neupauer, R.M., and J.L. Wilson. 1999. Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant. Water Resour. Res. 35:3389–3398.

Neupauer, R.M., and J.L. Wilson. 2001. Adjoint-derived location and travel time probabilities for a multidimensional groundwater. Water Resour. Res. 37:1657–1668.

Nielson, D.R., and J.W. Biggar. 1961. Mixible displacement in soils: I. Experimental Information. Soil Science Society of America Journal 25(1):1–5.

Nielsen, D.R., and J.W. Biggar. 1962. Miscible displacement: III Theoretical considerations. Soil Sci. Soc. Am. Proc. 26:216–221.

Nkedi-Kizza, P., J.W. Biggar, M.Th. van Genuchten, P.J. Wierenga, H.M. Selim, J.M. Davidson, and D.R. Nielsen. 1983. Modeling Tritium and Chloride 36 Transport through an Aggregated Oxisol. Water Resour. Res. 19:691–700.

Nkedi-Kizza, P., J.W. Biggar, H.M. Selim, M.Th. van Genuchten, P.J. Wierenga, J.M. Davidson, and D.R. Nielsen. 1984. On the equivalence of two conceptual models for describing ion exchange during transport through an aggregated Oxisol. Water Resour. Res. 20:1123–1130.

Nye, P.H. 1979. Diffusion of ions and uncharged solutes in soils and soil clays. Adv. Agron. 31:225–272.

O'Carroll, D.M., and B.E. Sleep. 2007. Hot water flushing for immiscible displacement of a viscous NAPL. J. Contam. Hydrol. 91:247–266.

Olsen, S.R., and W.D. Kemper. 1968. Movement of nutrients to plant roots. Adv. Agron. 20:91–151.

Oscarson, D.W., H.B. Hume, N.G. Sawatsky, and S.C.H. Cheung. 1992. Diffusion of iodide in compacted bentonite. Soil Sci. Soc. Am. J. 56:1400–1406.

Parker, J.C., and M.Th. van Genuchten. 1984. Determining transport parameters from laboratory and field tracer experiments. Virg. Exp. Stat. Bull. 84–3.

Pinder, G.F., and W.G. Gray. 1977. Finite elements in surface and subsurface hydrology. Academic Press, New York.

Pitzer, K.S. 1979. Activity coefficients in electrolyte solutions. CRC Press, Boca Raton, FL.

Powers, S.E., L.M. Abriola, and W.J. Weber. 1992. An experimental investigation of nonaqueous phase liquid dissolution in saturated subsurface systems: Steady state mass transfer rates. Water Resour. Res. 28:2691–2705.

Powers, S.E., C.E. Louriero, L.M. Abriola, and W.J. Weber. 1991. Theoretical study of the significance of nonequilibrium dissolution of nonaqueous phase liquids in subsurface systems. Water Resour. Res. 27:463–477.

Pruess, K., C. Oldenburg, and G. Moridis. 1999. TOUGH2 user's guide, version 2.0. Lawrence Berkeley National Laboratory Report LBNL-43134, Berkeley, CA.

Robin, M.J.L., R.W. Gillham, and D.W. Oscarson. 1987. Diffusion of strontium and chloride in compacted clay-based materials. Soil Sci. Soc. Am. J. 51(5):1102–1108.

Robin, M.J.L., E.A. Sudicky, R.W. Gillham, and R.G. Kachanoski. 1991. Spatial variability of Strontium distribution coefficients and their correlation with hydraulic conductivity in the Canadian forces base Borden aquifer. Water Resour. Res. 27:2619–2632.

Rubin, Y., and D. Or. 1993. Stochastic modeling of unsaturated flow in heterogeneous soils with water uptake by plant roots: The parallel columns model. Water Resour. Res. 29:619–631.

Sahloul, N.A., and I. Chatzis. 2002. Dissolution of residual non-aqueous phase liquids in porous media: Pore-scale mechanisms and mass transfer rates. Adv. Water Resour. 25:33–49.

Scheidegger, A.E. 1974. The Physics of Flow Through Porous Media. University of Toronto Press.

Schwarzenbach, R.P., P.M. Gschwend, and D.M. Imboden. 1993. Environmental organic chemistry. John Wiley, New York.

Schwille, F. 1988. Dense chlorinated solvents in porous and fractured media. Lewis, Chelsea, MI.

Sciortino, A., T.C. Harmon, and W.-G. Yeh. 2000. Inverse modeling for locating dense nonaqueous pools in groundwater under steady flow conditions. Water Resour. Res. 36:1723–1735.

Segol, G. 1994. Classic groundwater simulations: Proving and improving numerical models. PTR Prentice Hall, Englewood Cliffs, NJ.

Selim, H.M., J.M. Davidson, and R.S. Mansell. 1976. Evaluation of a Two-site Adsorption-Desorption Model for Describing Solute Transport in Soil. Proc. Summer Conf., Washington D.C., pp. 444–448.

Selim, H.M., and R.S. Mansell. 1976. Analytical solution of the equation for transport of reactive solutes through soils. Water Resour. Res. 12:528–532.

Selim, H.M., and R.S. Mansell. 1976. Analytical solution of the equation for transport of reactive solutes through soils. Water Resour. Res. 12:528–532.

Settari, A., and K. Aziz. 1975. Treatment of nonlinear terms in the numerical solution of partial differential equations for multiphase flow in porous media. Int. J. Multiphase Flow 1:817–844.

Sherwood, T.K., R.L. Pigford, and C.R. Wilke. 1975. Mass transfer. McGraw-Hill, Inc., New York.

Sidauruk, P., A.H.-D. Cheng, and D. Ouazar. 1998. Groundwater contaminant source and transport parameter identification by correlation coefficient optimization. Ground Water 36:208–214.

Simon, P., P. Reichert, and C. Hinz. 1997. Properties of exact and approximate traveling wave solutions for transport with nonlinear and nonequilibrium sorption. Water Resour. Res. 33(5):1139–1147.

Sitar, N., J.R. Hunt, and K.S. Udell. 1987. Movement of nonaqueous liquids in groundwater, p. 205–223. *In* Proc. Geotech. Pract. Waste Disposal '87. Geotechnical Division, ASCE Ann Arbor, MI.

Skaggs, T.H., and Z.J. Kabala. 1994. Recovering the history of a groundwater contaminant. Water Resour. Res. 30:71.

Skaggs, T.H., and Z.J. Kabala. 1995. Recovering the history of a groundwater contaminant plume: Method of quasi-reversibility. Water Resour. Res. 31:2669–2673.

Snodgrass, M.F., and P.K. Kitanidis. 1997. A geostatistical approach to contaminant source identification. Water Resour. Res. 33:537–546.

Sparks, D.L. 1989. Kinetics of soil chemical processes. Academic Press, San Diego.

Spiegel, M.R. 1965. Theory and problems of Laplace transforms. Schaum's Outline Series, McGraw-Hill, New York.

Sudicky, E.A. 1989. The Laplace transform Galerkin technique: A time continuous finite element theory and application to mass transport in groundwater. Water Resour. Res. 25:1833–1846.

Taylor, G.I. 1953. Dispersion of Soluble Matter in Solvent Flowing through a Tube. Proc. R. Soc. London A 219:186–203.

Toride, N., F.J. Leij, and M.Th. van Genuchten. 1993. A comprehensive set of analytical solutions for nonequilibrium solute transport with first-order decay and zero-order production. Water Resour. Res. 29:2167–2182.

Toride, N., and F.J. Leij. 1996a. Convective-dispersive stream tube model for field-scale solute transport: I. Moment analysis. Soil Science Society of America Journal 60(2):342–352.

Toride, N., and F.J. Leij. 1996b. Convective-dispersive stream tube model for field-scale solute transport: II. Examples and calibration. Soil Science Society of America Journal 60(2):352–361.

Toride, N., F.J. Leij, and M.Th. van Genuchten. 1995. The CXTFIT code for estimating transport parameters from laboratory or field tracer experiments. Version 2.0. U.S. Salinity Lab. Res. Rep. 137, Riverside, CA.

Uffink, G.J.M. 1989. Application of Kolmogorov's backward equation in random walk simulations of groundwater contaminant transport, p. 283–289. *In* H.E. Kobus and N. Kinzelbach (eds.) Contaminant transport in groundwater. Balkema, Rotterdam, the Netherlands.

van der Heijde, P., Y. Bachmat, J. Bredehoeft, B. Andrews, D. Holtz, and S. Sebastian. 1985. Groundwater management: The use of numerical models, 2nd Ed. American Geophysical Union, Washington, DC.

van der Laan, E.T. 1958. Notes on the diffusion-type model for the longitudinal mixing in flow. Chem. Eng. Sci. 7:187–191.

van der Waarden, M., W.M. Groenewound, and L.A.A. Bridié. 1971. Transport of mineral oil components to groundwater, I. Model experiments on the transfer of hydrocarbons from a residual oil zone to trickling water. Water Resour. Res. 5:213–226.

van der Zee, S.E.A.T.M. 1990. Analytical traveling wave solutions for transport with nonlinear and nonequilibrium adsorption. Water Resour. Res. 26:2563–2578.

van Genuchten, M.Th. 1978. Mass transport in saturated–unsaturated media. One-dimensional solutions. Princeton University Research Rep. 78-WR-11.

van Genuchten, M.Th. 1981. Non-equilibrium transport parameters from miscible displacement experiments. U.S. Salinity Lab. Res. Rep. 119, Riverside, CA.

van Genuchten, M.Th. 1988. Solute transport, p. 360–362. *In* S.P. Parker (ed.) McGraw-Hill yearbook of science and technology. McGraw-Hill Book Co., New York.

van Genuchten, M.Th., and W.J. Alves. 1982. Analytical solutions of the one-dimensional convective–dispersive solute transport equation. USDA Tech. Bull. 1661.

van Genuchten, M.Th., and R.W. Cleary. 1982. Movement of solutes in soils: Computer-simulated and laboratory results, p. 349–386. *In* G.H. Bolt (ed.) Soil chemistry. B. Physicochemical models. Elsevier, Amsterdam, the Netherlands.

van Genuchten, M.Th., and F.N. Dalton. 1986. Models for simulating salt movement in aggregated field soils. Geoderma 38:165–183.

van Genuchten, M.Th., and W.G. Gray. 1978. Analysis of some dispersion corrected numerical schemes for solution of the transport equation. Int. J. Numer. Methods Eng. 12:387–404.

van Genuchten, M.Th., and J.C. Parker. 1987. Parameter estimation for various contaminant transport models, p. 273–295. *In* C.A. Brebbia and G.A. Keramidas (eds.) Reliability and robustness of engineering software. Elsevier, New York.

van Genuchten, M.Th., and J.C. Parker. 1984. Boundary conditions for displacement experiments through short laboratory columns. Soil Sci. Soc. Am. J. 48:703–708.

van Genuchten, M.Th., and E.A. Sudicky. 1999. Recent advances in vadose zone flow and transport modeling, p. 155–193. *In* M.B. Parlange and J.W. Hopmans (eds.) Vadose zone hydrology: Cutting across disciplines. Oxford University Press, New York.

van Genuchten, M.Th., D.H. Tang, and R. Guennelon. 1984. Some exact solutions for solute transport through soils containing large cylindrical macropores. Water Resour. Res. 20:335–346.

van Genuchten, M.Th., and P.J. Wierenga. 1976. Mass transfer studies in sorbing porous media. I. Analytical solutions. Soil Sci. Soc. Am. J. 40:473–480.

van Genuchten, M.Th., and P.J. Wierenga. 1986. Solute dispersion coefficients and retardation factors. *In* A. Klute (ed.) Methods of soil analysis. I. Physical and mineralogical methods. SSSA, Madison, WI.

van Rees, K.C.J., E.A. Sudicky, P.S.C. Rao, and K.R. Reddy. 1991. Evaluation of laboratory techniques for measuring diffusion coefficients in sediments. Environ. Sci. Technol. 25:1605–1611.

Varoglu, E., and W.D.L. Finn. 1982. Utilization of the method of characteristics to solve accurately two-dimensional transport problems by finite elements. Int. J. Numer. Methods Fluids 2:173–184.

Voudrias, E.A., and M.F. Yeh. 1994. Dissolution of a toluene pool under constant and variable hydraulic gradients with implications for aquifer remediation. Ground Water 32:305–311.

Wagner, B.J. 1992. Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modeling. J. Hydrol. 135:275–300.

Warrick, A.W. 2003. Soil water dynamics. Oxford University Press, New York.

Whelan, M.P., E.A. Voudrias, and A. Pearce. 1994. DNAPL pool dissolution in saturated porous media: Procedure development and preliminary results. J. Contam. Hydrol. 15:223–237.

White, M.D., and M. Oostrom. 2000. STOMP subsurface transport over multiple phase: Theory guide PNNL-11216 (UC-2010). Pacific Northwest National Laboratory, Richland, WA.

Wierenga, P.J. 1977. Solute distribution profiles computed with steady-state and transient water movement. Soil Sci. Soc. Am. J. 41:1050–1055.

Willson, C.S., J.W. Weaver, and R.J. Charbeneau. 2006. A screening model for simulating DNAPL flow and transport in porous media: Theoretical development. Environ. Model. Softw. 21:16–32.

Woodbury, A., E. Sudicky, T.J. Ulrych, and R. Ludwig. 1998. Three-dimensional plume source reconstruction using minimum relative entropy inversion. J. Contam. Hydrol. 32:131–158.

Yeh, G.T. 1990. A Lagrangian–Eulerian method with zoomable hidden fine-mesh approach to solving advection–dispersion equations. Water Resour. Res. 26:1133–1144.

Yeh, H.-D., T.-H. Chang, and Y.-C. Lin. 2007. Groundwater contaminant source identification by a hybrid heuristic approach. Water Resour. Res. 43, W09420, doi:10.1029/2005WR004731.

Yeh, W.W.-G., and N.-Z. Sun. 1984. An extended identifiability in aquifer parameter identification and optimal pumping test design. Water Resour. Res. 20:1837–1847.

Zhang, C., C.J. Werth, and A.G. Webb. 2007. Characterization of NAPL source zone architecture and dissolution kinetics in heterogeneous porous media using magnetic resonance imaging. Environ. Sci. Technol. 41:3672–3678.

Zhao, W., and M.A. Ioannidis. 2007. Effect of NAPL film stability on the dissolution of residual wetting NAPL in porous media: A pore-scale modeling study. Adv. Water Resour. 30:171–181.

Zoller, U., and H. Rubin. 2001. Feasibility of in situ NAPL-contaminated aquifer bioremediation by biodegradable nutrient-surfactant. J. Environ. Sci. Health A 36:1451–1471.

Zwillinger, D. 1989. Handbook of differential equations. Academic Press, San Diego, CA.

<div style="text-align: right; font-size: 3em;">8</div>

Gas Transport in Soils

Dennis E. Rolston
University of California, Davis

Per Møldrup
Aalborg University

8.1 Introduction

Understanding gas transport in soils and the vadose zone is important for evaluation of soil aeration, which is the interchange of various gases between the atmosphere and the soil and the various processes that either produce or consume gases in the soil. Soil aeration is important for plant growth since O_2 is required for root respiration. The two major gases associated with aeration are oxygen (O_2) and carbon dioxide (CO_2) where O_2 moves from the atmosphere to soil and is consumed and CO_2 is produced in soil and moves from the soil to the atmosphere. Oxygen transport through the capillary fringe in regard to reaeration of groundwater is another important area of soil-gas transport. A thorough discussion of soil aeration is given in a book by Glinski and Stepniewski (1985), Scott (2000), Jury and Horton (2004), and Rolston (2005).

In solid waste landfills, methane and carbon dioxide transport through the cover soil will likely be by both diffusion and advection affected by variations in soil moisture conditions, soil type, and soil structure, while oxygen diffusion into the cover layer will be a main control of atmospheric methane oxidation and uptake. Movement of gaseous radionuclides from radioactive waste sites is important in many areas. Understanding gas transport is also important for estimating movement of volatile organic chemicals from contaminated waste sites within the vadose zone (Baehr et al., 1989). Understanding transport of radon and its

decay products in soils and into buildings (Nazaroff, 1992) is important for evaluating the possible adverse health effects of radon gas. In a longer-term perspective, design of optimal plant growth media for use on earth or in space requires detailed knowledge of the gas diffusivity behavior in aggregated porous media including soils at both normal gravity and microgravity (Jones et al., 2003).

Gas transport in soils occurs due to two processes: diffusion and advection. Diffusion is considered to be the principal mechanism in the interchange of gases between the soil and the atmosphere, although advection may be responsible for moving significant amounts of gas but generally over short time periods. The interchange due to diffusion results from concentration gradients established within soil by respiration of microorganisms, macrofauna, and plant roots; by production of gases associated with biological reactions, such as fermentation, nitrification, and denitrification; and by incorporation of chemicals, such as fumigants, anhydrous ammonia, pesticides, and volatile organic compounds (VOCs) from toxic waste sites. The diffusion of water vapor within the soil also occurs due to differences in vapor pressure gradients induced by temperature differences or by evaporative conditions at the soil surface.

Whereas diffusion is driven by concentration gradients, advection is driven by pressure gradients. Pressure gradients may develop due to barometric pressure changes in the atmosphere; wind blowing across the soil surface or against

a hill or other landscape feature; infiltration of water into the soil due to rainfall or irrigation; soil-water redistribution and evaporation; temperature gradients near the soil surface; and differences in gas density due to high concentrations of gases that have densities greater than air. Soil-gas pressure gradients are also created artificially using soil venting, or soil-vapor extraction, as techniques for removing harmful volatile gases from contaminated waste sites in the unsaturated zone (Rathfelder et al., 1995).

8.2 Soil-Air Composition

The composition of soil air depends on the relative magnitude of both the sources and sinks of the various gas components, the interchange between soil air and atmospheric air, and the partitioning of the gases between the gaseous, liquid, and solid (mineral and organic matter) phases of the soil. If a soil were completely "aerated," the concentrations of the gases in the soil air would be similar to that in the atmosphere. Oxygen concentrations in the soil air will be somewhat below that in the atmosphere (approximately 20% by volume), since O_2 is consumed in soil by plant root and microbial respiration (including soil mezofauna) and through chemical reactions. Under some conditions, O_2 concentrations can go to zero and the soil becomes anaerobic (anoxic). It is now widely accepted that under some conditions soil profiles do not have to be either fully aerated or fully anaerobic but may be partially aerobic and partially anaerobic. Anaerobic pockets or "hot spots" may exist within the soil due to pockets of very high O_2 consumption such as around incorporated carbon materials and/or due to very slow diffusion to regions of O_2 consumption. For example, the interior of large aggregates may be anaerobic due to slow diffusion to microbial "hot spots."

Carbon dioxide concentrations in the soil air can be as high as 10 times larger than in the atmosphere (0.036% by volume). Since nitrogen gas (N_2) is more abundant than other gases in the atmosphere (~76%) and there are generally no sources or sinks for N_2 in the soil (except N_2 produced during denitrification), the concentration of N_2 in the soil air will be similar to that in the atmosphere, varying only slightly depending on the production and consumption of other soil gases. The soil air will also contain varying amounts of nitric oxide and nitrous oxide (from nitrification and denitrification); methane, hydrogen sulfide, and ethylene (from anaerobic processes); water vapor; and trace amounts of inert gases such as argon. Human activities also result in the accidental or intentional introduction of gases in the soil profile such as fumigants, anhydrous ammonia, pesticides, and various organic chemicals that exist partially in the gas phase.

The concentrations of O_2 and CO_2 that occur in the soil-pore space vary widely, especially for CO_2, and depend on the rate of consumption and production and upon the rate that the soil is able to exchange these gases between the soil and the atmosphere through diffusion and advection. The diurnal and annual variability in the soil-gas concentrations is generally much greater

for clayey soils than for sandy soils due to the ability of sandy soils to transmit gases at a higher rate (larger soil-gas diffusion coefficients and air permeability) and maintain more constant concentrations.

8.3 Mechanisms of Gas Transport

8.3.1 Diffusion Processes and Definitions

The molecular transport of one gas relative to another is known as diffusion. Diffusion in porous media consists of several kinds of movement: *ordinary* or continuum diffusion (also known as molecular, mass, or concentration diffusion); *Knudsen* diffusion and *configurational* diffusion (free-molecule diffusion; Ho and Webb, 2006); *pressure* diffusion; *thermal* diffusion; *forced* diffusion; and *surface* diffusion. For the first of these, *ordinary diffusion*, molecules of a given species transfer momentum to the molecules of other species, and if a concentration or mole fraction gradient exists, there will be a net movement of a gas species from the more-concentrated region to the less-concentrated region. The probability of molecule–molecule collisions is much greater than molecule-wall collisions or, in other words, the distance that a molecule moves before it collides with another molecule (i.e., the mean free path), is much smaller than the soil-pore diameter.

For *Knudsen diffusion*, the probability of molecule–molecule collisions is negligible compared to molecule-wall collisions, or in other words, the mean free path is much larger than the soil-pore diameter. In this case, gas molecules transfer momentum predominately to the pore walls rather than to other gas molecules. The driving force for flow is due to a concentration gradient just as in ordinary diffusion. As the pore size gets even smaller, approaching that of single gas molecules, the flow mechanism changes and *configurational diffusion* (Ho and Webb, 2006) becomes important.

Pressure diffusion is the separation of gases of different molecular weight under a pressure gradient (Chapman and Cowling, 1970). Pressure diffusion causes the heavier molecules to diffuse toward the regions of greater pressures. Under the natural gravitational pressure gradient, the heavier molecules tend to diffuse downward along the direction of gravity. Pressure diffusion is generally negligible for vadose zone depths of less than 100 m (Amali and Rolston, 1993).

Thermal diffusion is the movement of gases due to a temperature gradient. This process appears to be minimal (Grew and Ibbs, 1952; Bird et al., 2007) because it requires large temperature gradients of 200°C or more, which are not possible under normal conditions in soils. Another diffusion process due to external forces such as a gravitational, electric, or magnetic field of some kind is called *forced diffusion*. For a case where gas molecules are adsorbed by a pore wall or surface, the gas molecule may migrate along the surface giving rise to *surface diffusion* where the binding energy is not too great. Since soil particles are generally covered by multiple layers of water, surface diffusion is either not likely or becomes incorporated into an

effective diffusion coefficient. For diffusion in soil and vadose zone materials, only ordinary and Knudsen diffusion are generally considered to be of most importance (Thorstenson and Pollock, 1989a, 1989b).

8.3.2 Gas Flux Definitions

As with diffusion processes, there are several different kinds of fluxes that arise from different driving forces in soils.

8.3.2.1 Ordinary Diffusive Flux

As mentioned earlier, ordinary or molecular diffusion occurs where there are no temperature or pressure gradients in systems where the pore size is much larger than the mean free path of the gas molecules or is dominated by molecule–molecule collisions. Although gas molecules are constantly in motion, there will only be a net flux (mass of gas moving across a cross-sectional area of soil per unit time) of a gas if a concentration or mole fraction gradient exists. Strictly speaking, ordinary diffusive flux only occurs for a binary mixture of gases that are of equal or nearly equal molecular weights, which results in equal countercurrent diffusion. Diffusion is considered to be "segregative" because the diffusing species has a different velocity than that of the gas as a whole (Cunningham and Williams, 1980).

For a two-component system consisting of gases i and j, the diffusive flux of gas i is given by Fick's first law,

$$N_i^F = -D_{ij}C\nabla y_i, \tag{8.1}$$

where

  N_i^F is the molar diffusive flux of gas i (mol m^{-2} s^{-1})
  C is the total molar density (mol m^{-3})
  D_{ij} is binary diffusion coefficient (m^2 s^{-1}) for the gas pair
  ∇y_i is the gradient in the mole fraction of gas i

Discussion of how Fick's law is adapted for use in porous media will be presented later.

8.3.2.2 Viscous Flux (Advection)

The flow of gas due to a pressure gradient is called viscous or advective flux. Viscous flux differs from that of diffusive flux in that the gas molecules lose momentum due to collisions with pore walls and there is no segregation of species since a specific gas species is moving at the same velocity as the gas as a whole (Cunningham and Williams, 1980). For soils, a viscous flux is due to the presence of a pressure gradient that arises from artificially applying a pressure difference in the soil or to such effects as barometric pressure changes, wind, and/or infiltration of water.

8.3.2.3 Nonequimolar Flux (Diffusion Slip Flux)

In general, ordinary diffusion of gases with equal molecular weights and equal countercurrent diffusion is rare in porous media.

The more common case is that diffusing gases differ in their molecular weights. If one considers two different gases of differing molecular weight initially separated that are allowed to diffuse into each other, two inseparable phenomena will occur. First, the lighter molecules will diffuse faster than the heavier molecules creating a higher concentration in the region toward which they flow. A pressure gradient thus develops that creates a nonsegregative viscous flux in a direction opposite to that of the direction of diffusion of the lighter molecules. Since this nonsegregative viscous flux arises as a result of a difference in concentration (nonequimolarity) it is called a "nonequimolar flux" (Cunningham and Williams, 1980). This flux is also called "diffusion slip flux." At some point, a quasi steady state will be reached, and the pressure gradient will become constant. Macroscopically, the only observable phenomena would be the equal and opposite net fluxes of the two different gases (Thorstenson and Pollock, 1989a, 1989b).

8.3.2.4 Knudsen Diffusive Flux

Knudsen diffusive or free-molecule flux occurs when the gas mean free path is much greater than the pore size. For this case, the gas molecules will collide only with the walls and rebounding molecules do not collide with other gas molecules. The flux is in response to a concentration gradient, there will be segregation of species, and specific species fluxes will be independent of one another and additive.

The Knudsen flux in a capillary tube can be described by

$$N_i^k = -D_i^k C\nabla y_i, \tag{8.2}$$

where

  N_i^k is the molar Knudsen flux of gas i (mol m^{-2} s^{-1})
  C is the total molar density (mol m^{-3})
  D_i^k is Knudsen diffusion coefficient (m^2 s^{-1}) for the gas i within the capillary tube
  ∇y_i is the gradient in the mole fraction of gas i (Reinecke and Sleep, 2002)

For unsaturated soils, a whole range of air-filled pore sizes will be available, and some method is needed to obtain an effective D_i^k for the porous medium. An approach used by Reinecke and Sleep (2002) is to make use of the Brooks–Corey equation relating soil-water potential to water-phase saturation (Corey, 1994). Using the Brooks–Corey relationship gives

$$D_i^k = D_i^{k0}(1 - S_e)^2(1 - S_e^{(1+\lambda)/\lambda}), \tag{8.3}$$

where

  D_i^{k0} is the Knudsen diffusion coefficient for an effective water saturation of zero (determined empirically)
  S_e is the effective water saturation
  λ is an empirical parameter called the pore-size distribution index that is determined from the soil-water characteristic for a particular soil or porous medium

8.3.2.5 Total Flux

The total molar diffusive flux, N_i^D, is the sum of the segregative ordinary diffusive flux and the nonsegregative nonequimolar diffusive flux given by

$$N_i^D = N_i^F + x_i \sum_{j=1}^{v} N_j^D, \qquad (8.4)$$

where $x_i \bullet \sum_{j=1}^{v} N_j^D$ is the nonequimolar flux.

The total molar gas flux is the sum of the total diffusive flux and the molar viscous flux given by

$$N_i^T = N_i^D + N_i^V. \qquad (8.5)$$

The equations for molar viscous flux will be given in a later section.

8.3.2.6 Diffusion versus Dispersion

According to the equation for total flux (Equation 8.5) above, the diffusive and the viscous flux are totally additive. However, it has been shown (Rolston et al., 1969) that when a pressure gradient is applied to soil columns inducing both viscous and diffusive flow that an additional gas mixing phenomena occurs called dispersion. Dispersion is a complicated phenomenon due to differences in gas velocities within a soil pore and between pores. Dispersion is usually handled mathematically by adding a dispersive flux term to the total advective or viscous flux (Jury and Horton, 2004). The dispersive flux is usually described by an equation identical to Fick's law, except the ordinary diffusion coefficient is replaced with a dispersion coefficient.

8.3.3 Gas/Liquid/Solid Partitioning

During gas transport through soils and vadose zone materials, some of the moving gas will dissolve in the soil water and some will adsorb onto soil surfaces. For oven-dry soil (no or little water adsorbed on soil surfaces), soils have a high capacity to adsorb gas. For example, Petersen et al. (1994, 1995) determined that the adsorption of toluene and trichloroethylene was 4 orders of magnitude higher on oven-dry soil than on moist soil. Water is preferentially adsorbed on soil surfaces over most soil gases, so for a moist soil, little gas will be adsorbed or partitioned on soil solids, but will be dissolved in soil water. The partitioning of soil gas between the solid and liquid phases is given by

$$C_s = K_d C_l, \qquad (8.6)$$

where

C_s is the mass concentration of adsorbed gas on the soil solids
K_d is the liquid/solid partitioning coefficient
C_l is the mass gas concentration in the liquid phase

The partitioning of soil gas between the gas and liquid phases is given by Henry's law,

$$C_g = K_H C_l, \qquad (8.7)$$

where

C_g is mass concentration of gas in the gas phase
K_H is the Henry's law constant

Soil gases vary significantly in their solubility in water. A table giving Henry's law constants for various soil gases and chemicals is given by Farrell et al. (2002).

8.3.4 Gas Reactions

8.3.4.1 Respiration

Oxygen is continuously consumed and CO_2 produced by plant roots and by soil microorganisms and other soil fauna. This process and the exchange of O_2 and CO_2 between soil and the atmosphere are called aeration. For a thorough description of soil aeration processes, see the texts of Glinski and Stepniewski (1985) and Scott (2000). The rates of O_2 consumption and CO_2 production are directly related to the rate of plant and microbial growth, which in turn is related to several environmental factors including air and soil temperature, substrate availability, and soil moisture. Many studies have shown that the bulk of the soil respiratory activity (>90%) occurs in regions of high organic matter near the soil surface (Glinski and Stepniewski, 1985). For aerobic conditions, the amount of CO_2 produced and O_2 consumed tends to be about equal. For anaerobic conditions, the CO_2 production will tend to be larger than the O_2 consumption because other reactions are occurring.

8.3.4.2 Oxidation/Reduction

These processes are connected with the transfer of electrons from soil organic matter (or organic contaminants) to oxidized inorganic compounds catalyzed by enzymes produced by soil microorganisms. For well-aerated conditions (aerobic), O_2 is the electron acceptor. When O_2 becomes limiting (anaerobic), other substances will accept electrons or be reduced. Examples of compounds that can be reduced under anaerobic conditions are nitrate, manganic manganese, ferric iron, sulfate, and perchlorate (a natural and anthropogenic contaminant). The reduction of nitrate and sulfate results in N_2 (and N_2O) and hydrogen sulfide gases, respectively. Another gas produced from reduction processes is methane (CH_4). Both CH_4 and N_2O are strong greenhouse gases and contribute to global warming. For further information on oxidation/reduction processes in soils, see Glinski and Stepniewski (1985).

8.3.4.3 Production and Consumption

There are a few gases produced in soils that are not necessarily associated with oxidation/reduction reactions. When fertilizer materials such as urea and ammonium salts are applied to the soil, reactions can occur, particularly in soils with a pH greater than

about 7.5 or 8 that produces ammonia gas (NH_3). Fairly large emission of NH_3 can occur, for instance, if urea is applied to the soil surface since urea hydrolysis results in a large increase in pH near the fertilizer granules, with NH_3 being emitted. The deeper the material is placed, the lesser the NH_3 will be emitted. Ammonia may also be produced in soil from the incorporation of animal waste.

Under aerobic conditions, ammonium-based materials either from fertilizer or mineralization of organic materials will be oxidized to nitrite and then to nitrate by microbial processes (nitrification). During nitrification, some N_2O and nitric oxide (NO) can be produced and emitted to the atmosphere (Venterea and Rolston, 2000). Nitric oxide enters into the tropospheric ozone cycle and can contribute to very small particle (<10 or <5 μm) generation in the atmosphere.

Besides O_2, gases including hydrocarbons, N_2O, NO, CH_4, and some sulfur gases may move from the atmosphere into the soil and be consumed by biological processes. Thus, the soil may act as a significant sink for atmospheric gases under some circumstances. At toxic waste sites, the vapors of VOCs may be prevalent in the gas phase of soil and may undergo various consumption and production processes. In other cases, fumigants, herbicides, and pesticides are added to soil for agricultural purposes, and the vapors of these chemicals may undergo consumption processes due to biodegradation.

8.4 Gas Transport Theory

8.4.1 Stefan–Maxwell Equations

In general, the diffusion velocities of gas mixtures in porous media are related to each other in a complex manner dependent upon the mole fraction of each gas, the molar fluxes of each gas, and the binary diffusion coefficient of each gas pair. Bird et al. (2007) indicate that the Stefan–Maxwell equations are simply an extension of Fick's law for multicomponent mixtures of gases. Curtiss and Hirschfelder (1949) gave the general equations for steady transport of a multicomponent mixture of gases, based on kinetic gas theory. If gravity effects are ignored or diffusion occurs only horizontally, the well-known Stefan–Maxwell equations provide the theoretical framework for 1D diffusion of gases in soils, given by

$$\sum_{j=1, i \neq j}^{n} \frac{y_j N_i - y_i N_j}{D_{ij}} = -\frac{P}{RT} \frac{dy_i}{dz}, \qquad (8.8)$$

where
- y_i and y_j are the mole fractions of gases i and j, respectively
- N_i and N_j are the molar fluxes (mol m^{-2} s^{-1}) of gas i and j, respectively
- D_{ij} is the binary diffusion coefficient (m^2 s^{-1}) for gas pairs i and j in the gas phase (no porous media)
- P is total pressure (kPa)
- R is the universal gas constant (m^3 kPa mol^{-1} K^{-1})
- T is absolute temperature (K)
- z is a spatial coordinate (m)

This equation results in n − 1 independent equations because

$$\sum_{i=1}^{n} y_i = 1. \qquad (8.9)$$

To use this equation for soils, the binary diffusion coefficient, D_{ij}, would need to be replaced with the soil-gas diffusion coefficient to be discussed later.

8.4.2 Fick's Equations

Fick's first law applies to steady-state gas transport, and Fick's second law is for unsteady (transient) transport derived by incorporating Fick's first law into the mass conservation equation. Fick's first law for diffusion is considered to be a restrictive case of the Stefan–Maxwell equations. Jaynes and Rogowski (1983) showed that Fick's law would result in incorrect estimations of mass distributions and fluxes for systems containing two or more vapor components when the concentrations of the diffusing species are not small compared to the bulk gas. They showed that in ternary systems, the Fickian diffusion coefficient depends not only on the binary diffusion coefficient and flux ratios but also on the mole fraction of the diffusing species. Leffelaar (1987) found that when the binary diffusion coefficients are different by a factor of two, the diffusion of gases in multicomponent gas mixtures could not accurately be described with Fick's law. Amali and Rolston (1993) showed that if the total mole fraction of VOC vapors in air exceeded about 0.05, Fick's law fluxes deviated from the true diffusive flux for each species by up to 5%. This indicated that the total mole fraction of the diffusing species also needs to be considered in addition to the requirement of Leffelaar (1987) that the binary diffusion coefficients of the diffusing species must be sufficiently different for multicomponent effects to be observed.

Thus, Fick's law is generally applicable for only a few special situations. One of these cases is for the diffusion of a trace gas in a binary mixture, meaning that the mole fraction of the tracer gas is small. The second special case is for diffusion of two gases in a closed system (total pressure remains constant or there is equimolar countercurrent diffusion). In this case, neither gas needs to be in trace amounts. A third case where Fick's law is applicable is for a three-gas system where there is equimolar, countercurrent diffusion of two of the gases and the third gas is stagnant (Jaynes and Rogowski, 1983). The consequences and potential errors of using Fick's law for situations other than the few special cases above are discussed in the application sections later on.

To use Fick's law for describing steady transport of a single gas species in soils, one must assume that the above conditions are met. Fick's first law given in mass fluxes and mass concentrations is

$$\frac{M_g}{At} = f_g^F = -D_p \frac{\partial C_g}{\partial z}, \qquad (8.10)$$

where

M$_g$ is the amount of gas diffusing (g gas)

A is the cross-sectional area of the soil (m^2 soil)

t is time (s)

f$_g^F$ is the gas flux density (g gas m^{-2} soil s^{-1})

C$_g$ is concentration in the gaseous phase (g gas m^{-3} soil air)

z is distance (m soil)

D$_p$ is the soil-gas diffusion coefficient (m^3 soil air m^{-1} soil s^{-1})

Soil air is considered to consist of a mixture of gases (i.e., N$_2$, O$_2$, CO$_2$, etc.). For unsteady transport of gases in soil, Fick's second law, including partitioning of the soil gas in soil water and onto soil surfaces and a source/sink term for production/consumption processes, is

$$R\varepsilon \frac{\partial C_g}{\partial t} = D_p \frac{\partial^2 C_g}{\partial z^2} - r_g, \qquad (8.11)$$

where R is the retardation factor given by

$$R = 1 + \frac{\theta}{\varepsilon K_H} + \frac{\rho_b K_d}{\varepsilon K_H} \qquad (8.12)$$

ε is the soil-air content (air-filled porosity; m^3 air m^{-3} soil)

r$_g$ is the sink or source (positive sign) that can be described by various kinetic equations

θ is soil-water content (m^3 water m^{-3} soil)

ρ$_b$ is soil dry bulk density (kg soil m^{-3} soil)

8.4.3 Dusty Gas Model

In order to use either the Stefan–Maxwell equations or Fick's law as given above to calculate soil-gas flux and/or concentration distributions, the equations are modified by inserting empirically obtained, soil-gas diffusion coefficients for the binary diffusion coefficients in free air (no porous media). Even without imposed pressure gradients in the soil, advective or viscous flow may occur in nonequimolar diffusion processes (discussed above). Only for equimolar gases will diffusion not result in advection. The phenomena of an advective flux due to diffusion of a non-equimolar system of gases have been implicitly included in the diffusion formulation (Bird et al., 2007), in that diffusion is in relation to the molar-average velocity of the gas as a whole (Ho and Webb, 2006). Coupling of diffusion and advection due to the nonequimolar phenomenon has been formulated with the development of the dusty gas model (DGM) by Mason et al. (1967). The DGM includes the effect of the porous media as a "dusty gas" component of the gas mixture, where the "dusty gas" is assumed to be large molecules fixed in space. The DGM combines ordinary and Knudsen diffusion through kinetic arguments and advective fluxes are added based on Chapman–Enskog kinetic

theory (Ho and Webb, 2006). Writing the DGM in terms of total molar flux (diffusive plus advective) relative to fixed coordinates (Thorstenson and Pollock, 1989a) gives

$$\sum_{\substack{j=1 \\ j \ne i}}^{n} \frac{y_i N_j^T - y_j N_i^T}{D_{ij}^e} - \frac{N_i^T}{D_i^k} = \frac{P\nabla y_i}{RT} + \left(1 + \frac{k_g P}{D_i^k \mu_g}\right)\frac{y_i \nabla P}{RT}, \qquad (8.13)$$

where

N$_j^T$ is the total molar flux of gas species j

D$_{ij}^e$ is the effective ordinary diffusion coefficient of gas species j in i

k$_g$ is the air permeability

μ$_g$ is the gas viscosity

The effective ordinary diffusion coefficient, D$_{ij}^e$, is equivalent to the soil-gas diffusion coefficient, D$_p$, in Fick's equations above. The first term on the left-hand side of the above equation considers molecule–molecule interactions and is recognized as the Stefan–Maxwell equations. The second term on the left-hand side takes into account molecule–particle interactions (Knudsen diffusion). The first and second terms on the right-hand side represent the driving forces for diffusion and advection due to gradients in gas partial pressure (concentration) and total pressure, respectively (Ho and Webb, 2006). If Knudsen diffusion is negligible and there are no total pressure gradients, the DGM simplifies to the Stefan–Maxwell equations. For a more thorough description of the DGM, see the books by Mason and Malinauskas (1983) and Ho and Webb (2006), and the papers by Thorstenson and Pollock (1989a, 1989b).

8.4.4 Darcy's Equation

Darcy's law can be used to describe advective gas transport in soils. For a nonreactive gas, the total molar advective (viscous) flux is given by

$$N^V = -\frac{P}{RT}\frac{k_g}{\mu_g}\nabla P. \qquad (8.14)$$

8.4.5 Convection–Dispersion Equation

If one assumes that for both diffusion and advection that the two processes are simply additive, then the total mass flux is the sum of the flux due to diffusion by Fick's law and that due to advection by Darcy's equation. Thus, the total mass flux is

$$f_g^T = f_g^F + f_g^D, \qquad (8.15)$$

where

f$_g^F$ is the mass flux due to Fick's first law

f$_g^D$ is the mass flux due to Darcy's equation

Combining these fluxes and the mass conservation equation (equation of continuity) and assuming that the gas is nonreactive gives

$$\frac{\partial C_g}{\partial t} = \frac{D_P}{\varepsilon} \frac{\partial^2 C_g}{\partial z^2} - \frac{f_g^D}{\varepsilon} \frac{\partial C_g}{\partial z}. \tag{8.16}$$

The term f_g^D/ε is the "average" velocity of gas $\overline{v}_g$ within pores. Due to additional mixing from dispersion (see definition above), the term $D_p/\varepsilon = D_H$ becomes somewhat of a fitting parameter and is called the dispersion coefficient (Rolston et al., 1969).

8.5 Measuring Gas Transport Parameters

8.5.1 Soil-Gas Diffusion Coefficient

In order to use the models above for calculating soil-gas fluxes and/or soil-gas concentration distributions, the fundamental parameter that must be determined either by measurement or from predictive equations is the soil-gas diffusion coefficient, D_P. Field or in situ methods are available, but the most common approach is a laboratory method using either undisturbed or repacked soil cores (Rolston and Moldrup, 2002). As with most other soil properties, the soil-gas diffusion coefficient is spatially variable, so multiple cores should be collected from each soil depth of interest. The diffusion coefficient is also strongly dependent upon the soil-air content (air-filled porosity), which is dependent upon the soil-water content. Thus, diffusion coefficients should be measured over a range of soil-air contents.

The standard laboratory method is based upon establishing a tracer gas concentration, C_o, within a chamber (Figure 8.1). One end of a soil core at concentration C_s is placed in contact with the gas of the chamber. The other end of the soil core is maintained at concentration C_s, usually by exposing the core to the atmosphere. The gas of interest diffuses either into or out of the chamber depending upon the gas concentration C_o compared to that outside the core. The other gases making up the atmosphere will diffuse in an opposite direction to that of the gas of interest. The rate of change of concentration versus time in the chamber is related to the soil-gas diffusion coefficient.

The gases to be used in measuring the soil-gas diffusion coefficient should be selected with some care to be sure that Fick's law is valid. As discussed above, this means that the gas of interest should be in trace amounts or that the self-diffusion coefficients of the gases in the diffusion apparatus are similar. Also, the gas should be basically nonreactive in order to simplify the mathematical derivation. For example, a trace gas that has been used for this purpose is Freon diffusing into air. Freon is nonreactive and can be measured in trace amounts by gas chromatography (Rolston and Moldrup, 2002). Another common approach is to use oxygen in air as the gas of interest diffusing into nitrogen gas. The two gases, O_2 and N_2, make up 99% of atmospheric air and their molecular weights are similar, basically making
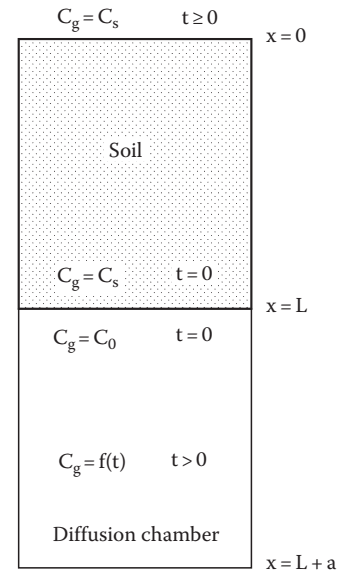
FIGURE 8.1 Schematic showing boundary and initial conditions used in the recommended laboratory method for determining the soil-gas diffusion coefficient. (Redrawn from Rolston, D.E., and P. Moldrup. 2002. Gas diffusivity, p. 1141–1158. *In* G.C. Topp and J.H. Dane (eds.) Methods of soil analysis, Part 4, Physical methods. SSSA, Madison, WI. With permission from Soil Science Society of America.)

up a binary mixture. Thus, the self-diffusion coefficients are very similar. The binary diffusion coefficient of O_2 diffusing into either air or N_2 is very similar to the self-diffusion coefficients of O_2 and N_2. Based upon these arguments, with O_2 as the tracer gas in measurements of the soil-gas diffusion coefficient and using Fick's law, using O_2 as the tracer gas appears to be a valid approach. Furthermore, Leffelaar (1987) indicated that if the binary diffusion coefficients do not differ by more than a factor of 2, then Fick's law should be valid. Since O_2 is potentially consumed in moist soil and CO_2 is produced by microbial respiration, an additional error may occur. However, since the diffusion measurements are taken over fairly short time periods of a few minutes to a few hours, it is assumed that the consumption of O_2 and production of CO_2 is small, and the error in using O_2 as the tracer gas is considered to be small (<1.5%) (Moldrup et al., 2000b).

The method described here is based on solutions to Fick's second law given by Currie (1960). Similar procedures have been used by several investigators (Rust et al., 1957; Gradwell, 1961; Shearer et al., 1966; Flühler, 1973; Schjønning, 1985; Resurreccion et al., 2008).

For unsteady diffusion of a gas that is nonreactive (physically, biologically, and chemically), Equation 8.11 can be simplified to give

$$\varepsilon \frac{\partial C_g}{\partial t} = D_P \frac{\partial^2 C_g}{\partial z^2}, \tag{8.17}$$

where ε is the soil-air content (m^3 air m^{-3} soil).

In developing Equation 8.11, it is assumed that the soil is uniform with respect to the diffusion coefficient and that ε is constant in space and time. Equation 8.17 may be solved subject to the boundary and initial conditions given in Figure 8.1. The solution (Carslaw and Jaeger, 1959, p. 128) for the relative concentration in the chamber C_r is

$$C_r = \frac{C_g - C_s}{C_o - C_s} = \sum_{n=1}^{\infty} \frac{2h \exp(-D_p \alpha_n^2 t/\varepsilon)}{L(\alpha_n^2 + h^2) + h}, \qquad (8.18)$$

where

C_g is the concentration in the chamber ($g\,m^{-3}$)

$h = \varepsilon/(a\varepsilon_c)$, ε_c is the air content of the chamber ($1.0\,m^3$ air m^{-3} chamber), a is the length of the chamber or the volume V, of the chamber per area A, of soil

α_n, with n = 1, 2,…, are the positive roots of

$$\alpha L \tan \alpha L = hL. \qquad (8.19)$$

At some time greater than zero, the terms for $n \geq 2$ are negligible with respect to the first term, and Equation 8.18 reduces to

$$C_r = \frac{2h \exp(-D_p \alpha_1^2 t/\varepsilon)}{L(\alpha_1^2 + h^2) + h}, \qquad (8.20)$$

where α_1 is the first root of Equation 8.19 above. Thus, for a plot of $\ln C_r$ versus time t, Equation 8.20 becomes linear with the slope $= -D_p \alpha_1^2/\varepsilon$ for sufficiently large t.

To make the diffusion coefficient measurement, the diffusion apparatus is set up according to the boundary and initial conditions of Figure 8.1 using a diffusion chamber setup similar to Figure 8.2. The tracer gas is measured within the diffusion chamber initially and then at several times after tracer gas is allowed to diffuse either into or out of the diffusion chamber. Graph ln
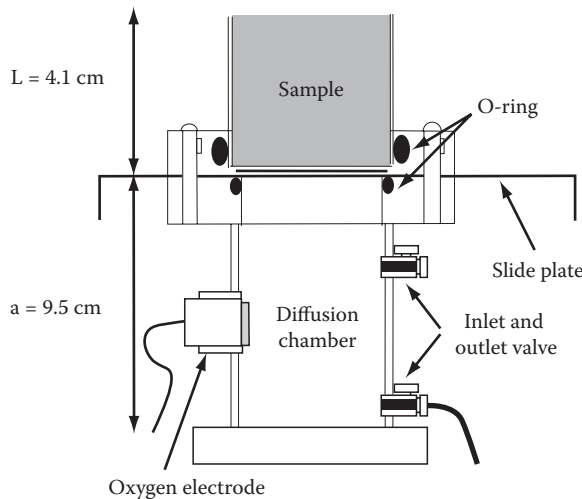
FIGURE 8.2 Schematic diagram of a diffusion apparatus that uses oxygen as the tracer gas.

TABLE 8.1 The First Root, $\alpha_1 L$, of $\alpha L \tan \alpha L = hL$ and the Roots of This Equation Are All Real If $hL > 0$

hL	$\alpha_1 L$	hL	$\alpha_1 L$	hL	$\alpha_1 L$
0	0	0.4	0.5932	8.0	1.3978
0.001	0.0316	0.5	0.6533	9.0	1.4149
0.002	0.0447	0.6	0.7051	10.0	1.4289
0.004	0.0632	0.7	0.7506	15.0	1.4729
0.006	0.0774	0.8	0.7910	20.0	1.4961
0.008	0.0893	0.9	0.8274	30.0	1.5202
0.01	0.0998	1.0	0.8603	40.0	1.5325
0.02	0.1410	1.5	0.9882	50.0	1.5400
0.04	0.1987	2.0	1.0769	60.0	1.5451
0.06	0.2425	3.0	1.1925	80.0	1.5514
0.08	0.2791	4.0	1.2646	100.0	1.5552
0.1	0.3111	5.0	1.3138	∞	1.5708
0.2	0.4328	6.0	1.3496		
0.3	0.5218	7.0	1.3766		

Source: Adapted from Rolston, D.E., and P. Moldrup. 2002. Gas diffusivity, p. 1141–1158. *In* G.C. Topp and J.H. Dane (eds.) Methods of soil analysis, Part 4, Physical methods. SSSA, Madison, WI. With permission from Soil Science Society of America.

C_r versus time, t, according to Equation 8.20, and determine the slope $= -D_p \alpha_1^2/\varepsilon$ from the linear part of the curve. To get D_p from this relationship, one must have a measured value of ε, and the value of α_1 must be determined from Table 8.1. For specifics on the diffusion apparatus, measurement methodologies, and alternative approaches for handling the data, see Rolston and Moldrup (2002). Also, see Rolston and Moldrup (2002) for a description of some field methods for measuring D_p.

8.5.2 Air Permeability

The fundamental soil property for advective gas transport in soils is the air permeability. It is similar to hydraulic conductivity for soil water but differs in that the air permeability of dry porous media is larger than the corresponding permeability determined from water flow due to the occurrence of slip flux along the pore walls (Klinkenberg, 1941). As with the soil-gas diffusion coefficient, the air permeability differs for various soil types and is highly dependent upon soil-water content.

In principle, the air permeability is determined by imposing a pressure gradient across some volume of soil (a soil column in the simplest case) to induce a steady-state advective flow of gas through the soil. The steady-state approach is used predominately for laboratory methods where either the pressure gradient is maintained constant and the flow rate is measured or the flow rate is maintained constant and the pressure is measured. The pressure gradient across the column and the flow rate through the soil volume are measured and the air permeability is calculated from Darcy's equation (Equation 8.14).

There have been many experimental variations and measurement approaches of steady-state laboratory and field approaches since the early designs of Grover (1955) as discussed by Ball and
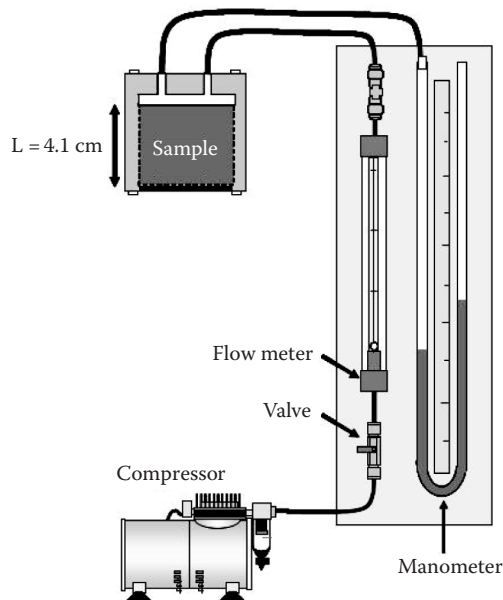
FIGURE 8.3 Schematic of a typical laboratory apparatus to measure the air permeability of a soil core.

Schjønning (2002). In a nonsteady-state field method first used by Kirkham (1946), air permeability was calculated from the rate of decrease in pressure in a tank supplying the air flowing through an in situ soil sample.

The laboratory technique recommended by Ball and Schjønning (2002) is the steady-state approach used by Ball et al. (1981) where the air permeability is measured on an intact or repacked soil core enclosed within a two-chamber apparatus. The two-chamber apparatus allows good control of the boundary conditions on each end of the soil core. For this technique, a constant flow of air is applied, and the pressure at each end of the soil core is accurately measured. For specifics on this technique, see Ball and Schjønning (2002). A typical apparatus is schematically shown by Figure 8.3.

The field method recommended by Ball and Schjønning (2002) is a steady-state, surface-chamber method where pressurized air is applied to a sharpened cylinder (recommended diameter and height of 200 mm) pressed into the soil. The top of the cylinder is sealed to create a closed chamber. Thus, air flows through the soil enclosed within the cylinder and flows out the bottom of the cylinder into the soil profile. Air flow rates and the pressure gradient across the cylinder of soil are measured precisely. Again, the air permeability is calculated from Darcy's law. This method is based on that of van Groenewoud (1968), Green and Fordham (1975), Liang et al. (1995), and Iversen et al. (2001). For specifics on this technique, see Ball and Schjønning (2002).

8.5.3 Dispersion Coefficient

The soil-gas dispersion coefficient is generally determined from laboratory soil column experiments (miscible displacement). An undisturbed or repacked soil column is set up with a system for pumping gas through columns, generally of differing soil-water content. The soil-air content for the column must be known. The flow rate of the gas is measured precisely, and a switching mechanism is established on the inflow end of the column so that the gas flow into the column can be switched from the background gas to the tracer gas. Initially, the column is flushed with a background gas such as N_2 until the flow becomes steady and the gas filled pores are completely filled with the background gas. Then, the inflow gas is switched to the tracer gas, and the concentration of the tracer gas is measured in the outlet gas stream. Some investigators have also measured the tracer gas concentration at various distances along the column in addition to the outlet gas stream. The concentration of the tracer gas in the outlet gas stream (and at sampling points along the column if applicable) is measured as a function of time, usually until the tracer gas concentration is equal to the concentration in the inflow stream. The inflow gas may be switched back to the background gas and tracer gas concentration in the outflow measured until the tracer gas approaches zero.

Solutions of the convective–dispersion equation (Equation 8.16) (Rolston et al., 1969; Costanza-Robinson and Brusseau, 2002, 2006) are then compared with the experimental data with the dispersion coefficient being a fitting parameter. Some investigators also curve fit the pore-gas velocity in the convective–dispersion equation (Hamamoto et al., 2009b) and then compare the fitted pore-gas velocity with the measured velocity determined from the gas flow rate and the soil-air content.

Various designs of the experimental apparatus have been used (Rolston et al., 1969; Costanza-Robinson and Brusseau, 2002; Poulsen et al., 2008), but Hamamoto et al. (2009b) constructed a unique apparatus that allows for the measurement of the soil-gas diffusion coefficient, the air permeability, and the dispersion coefficient on the same soil core shown schematically by Figure 8.4.

8.6 Predictive Equations

8.6.1 Soil-Gas Diffusion Coefficient

8.6.1.1 Repacked, Dry Soil

Prediction of the soil-gas diffusion coefficient (D_p) in dry porous media (void of water) normally can be described by the classical power law of Buckingham (1904),

$$\frac{D_P}{D_0} = A\varepsilon^B, \tag{8.21}$$

where

    D_P is the soil-gas diffusion coefficient (m^3 soil air m^{-1} soil s^{-1})
    D_0 is the gas diffusion coefficient in free air (m^2 air s^{-1})
    ε is the volumetric soil-air content (the air-filled porosity; m^3 soil air m^{-3} soil)
    A and B are empirical constants representing pore connectivity, continuity, and tortuosity of the dry porous medium
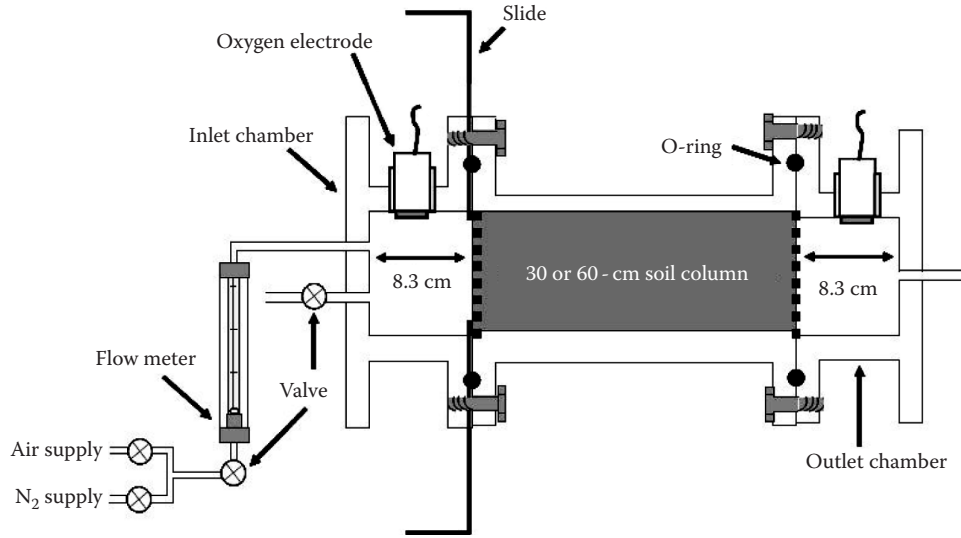
FIGURE 8.4 Schematic of a typical apparatus for measuring the dispersion coefficient of a soil core. Schematic simplified from unified measurement system (UMS) setup by Hamamoto et al. (2009b).

In the case of a dry porous medium, the soil-air content will equal the soil total porosity, Φ (m^3 pore space m^{-3} soil).

Moldrup et al. (2000b) tested three widely used models based on Equation 8.21: the Penman (1940) model (A = 0.66, B = 1), the Millington (1959) model (A = 1, B = 1.33), and the Marshall (1959) model (A = 1, B = 1.5). They found that all three models performed well against data for a wide range of sieved, repacked dry porous media (sand, soil, glass beads, steel and glass wool, and various mixtures) in the range of $0.2 < \Phi < 0.6$. The Marshall (1959) model,

$$\frac{D_p}{D_0} = \varepsilon^{1.5}, \tag{8.22}$$

performed best across the entire range of air-filled porosities and is generally recommended to predict the soil-gas diffusion coefficients in dry, nonaggregated soil.

In the case of aggregated soil with bimodal pore-size distribution, recent studies (Resurreccion et al., 2008, 2010) suggest that the change of D_p/D_0 with ε in the intraaggregate pore space is linear and often with a slope close to 0.5. Thus, if the soil is highly aggregated and with known values of interaggregate porosity (Φ_1; m^3 interaggregate pore space m^{-3} soil) and intraaggregate porosity ($\Phi_2 = \Phi - \Phi_1$), which could be determined from the measured soil-water retention curve (Resurreccion et al., 2009), the soil-gas diffusion coefficient for dry conditions may be estimated from

$$\frac{D_p(\varepsilon = \Phi)}{D_0} = (\Phi_1)^{1.5} + 0.5\Phi_2. \tag{8.23}$$

8.6.1.2 Repacked, Wet Soil

In the case of wet porous media, the soil water will typically be located in the narrow bottlenecks between pores, meaning a small amount of water can have a dramatic effect in reducing D_p/D_0 (Papendick and Runkles, 1965; Moldrup et al., 2000a). This is illustrated in Figure 8.5 for a sieved, repacked fine sand (Toyoura sand, Japan; data from Hamamoto et al., 2009a) at different moisture contents compared with dry
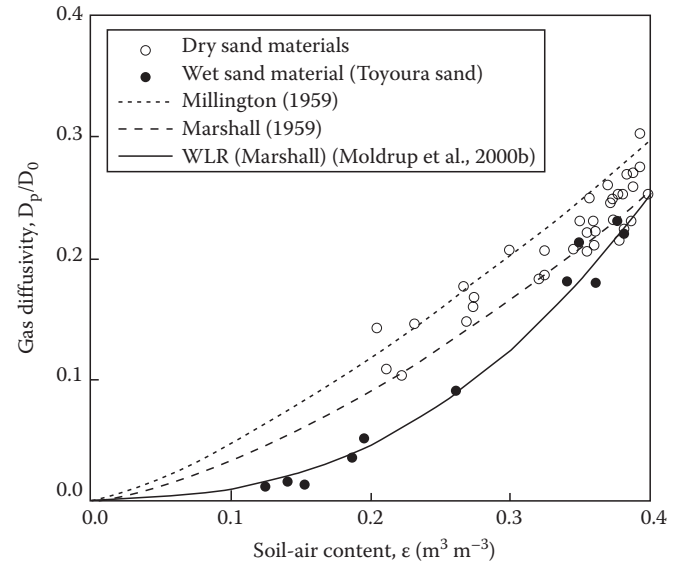
FIGURE 8.5 Prediction of the soil-gas diffusivity in dry and wet (variably saturated) sand. Models considered are the Millington (1959) and Marshall (1959) models for dry soil (void of water) and a WLR type of model for variably saturated soil (Moldrup et al., 2000b). (Data from Moldrup, P., T. Olesen, J. Gamst, P. Schjønning, T. Yamaguchi, and D.E. Rolston. 2000a. Predicting gas diffusivity in sieved, repacked soil: Water-induced linear reduction model. Soil Sci. Soc. Am. J. 64:1588–1594; Moldrup, P., T. Olesen, P. Schjønning, T. Yamaguchi, and D.E. Rolston. 2000b. Predicting the gas diffusion coefficient in undisturbed soil from soil water characteristics. Soil Sci. Soc. Am. J. 64:94–100; Hamamoto, S., P. Moldrup, K. Kawamoto, and T. Komatsu. 2009a. Effect of particle size and soil compaction on gas transport parameters in variably saturated, sandy soils. Vadose Zone J. 8:986–995.)

Toyoura sand and different other dry sands from the literature (data compiled by Moldrup et al., 2000a). In the dry state at different levels of compaction (open circles), the measured data are mostly placed in between the models by Millington (1959: A = 1 and B = 1.33 in Equation 8.21) and Marshall (1959, Equation 8.22). In the wet state at different moisture contents (filled-out circles), the D_p/D_0 is markedly reduced compared to the dry state at the same air-filled porosity. This leads to the development of the so-called water-induced linear reduction (WLR) class of models (Moldrup et al., 2000a; Hamamoto et al., 2009a), where the D_p/D_0 model for an equivalent dry media at a given ε (Equation 8.21) is multiplied with a water reduction factor,

$$F = \left(\frac{\varepsilon}{\Phi} \right)^N, \tag{8.24}$$

where N for repacked soils typically can be taken as 1 (Moldrup et al., 2000b; Hamamoto et al., 2009a; Resurreccion et al., 2009). If the Marshall (1959) model (Equation 8.22) is assumed at dry conditions and N = 1 is applied in Equation 8.24, the resulting WLR (Marshall) model becomes

$$\frac{D_p}{D_0} = \varepsilon^{1.5} \left(\frac{\varepsilon}{\Phi} \right). \tag{8.25}$$

This model with a high degree of accuracy predicted D_p/D_0 in sieved, repacked soil across the entire water content range from wet to dry, and with soil texture ranging from 0% to 54% clay (Moldrup et al., 2000a). Figure 8.6 shows the performance of the WLR (Marshall) model against data for 15 sands and differently textured soils from the literature (data compiled by Moldrup et al., 2000a and Hamamoto et al., 2009a), with total porosities (Φ) ranging from 0.4–0.5 m³ m⁻³ (average 0.45 m³ m⁻³). Also shown is the still most commonly used model in the literature for predicting gas diffusivity, the Millington and Quirk (1961) model,

$$\frac{D_p}{D_0} = \frac{\varepsilon^{3.33}}{F^2} = \varepsilon^{1.33} \left(\frac{\varepsilon}{\Phi} \right)^2. \tag{8.26}$$

The Millington and Quirk (1961) model when rewritten like the right-hand side of Equation 8.26 is recognized as an analogue to the WLR model, here based on Millington (1959; A = 1 and B = 1.33) at dry conditions and with N = 2 in Equation 8.24, representing a nonlinear water reduction factor. Both at high gas diffusivities (see Figure 8.6a with normal-scale D_p/D_0 axis) and at low gas diffusivities (see Figure 8.6b with log-scale D_p/D_0 axis), the WLR model (Equation 8.25) is more accurate than the Millington and Quirk model (Equation 8.26). The upper, middle, and lower prediction curves for each model (WLR solid line, Millington and Quirk dotted line) in Figure 8.6a and b represent Φ = 0.4, 0.45, and 0.5 m³ m⁻³, respectively.
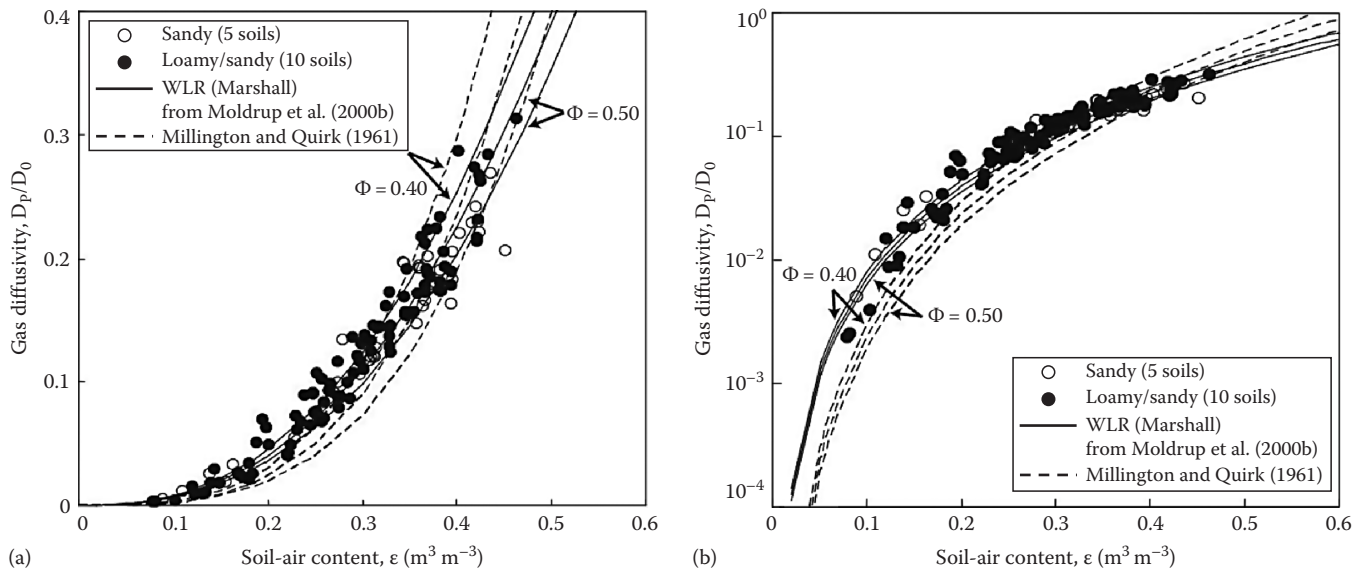
FIGURE 8.6 Prediction of the soil-gas diffusivity in variably saturated sands and soils with 0%–54% clay. Models considered are the widely used Millington and Quirk (1961) model (dotted line) and WLR model (Moldrup et al. 2000b; solid line). Model performance is shown in (a) normal-scale D_p/D_0 and (b) log-scale D_p/D_0 plots, and for values of soil total porosity (Φ) equal to 0.4 (upper line; representing minimum Φ for data), 0.45 (average for all data), and 0.5 (lower line; representing maximum Φ for data). (Data from Moldrup, P., T. Olesen, J. Gamst, P. Schjønning, T. Yamaguchi, and D.E. Rolston. 2000a. Predicting gas diffusivity in sieved, repacked soil: Water-induced linear reduction model. Soil Sci. Soc. Am. J. 64:1588–1594; Moldrup, P., T. Olesen, P. Schjønning, T. Yamaguchi, and D.E. Rolston. 2000b. Predicting the gas diffusion coefficient in undisturbed soil from soil water characteristics. Soil Sci. Soc. Am. J. 64:94–100; Hamamoto, S., P. Moldrup, K. Kawamoto, and T. Komatsu. 2009a. Effect of particle size and soil compaction on gas transport parameters in variably saturated, sandy soils. Vadose Zone J. 8:986–995.)

The reason why both measured data and predictive models suggest that the D_p/D_0 at a given air-filled porosity (ε) is larger at higher compaction (lower Φ) is due to the larger water to solid ratio, remembering that increasing water content typically has a dramatic effect on reducing air-filled pathways for gas diffusion, since water will form water blockages in bottlenecks between air-filled pores. For example, a soil with an air-filled porosity of $\varepsilon = 0.25\,m^3\,m^{-3}$ and a total porosity of $\Phi = 0.4\,m^3\,m^{-3}$ will contain a volumetric water content of $\theta = 0.15\,m^3\,H_2O\,m^{-3}$ soil and a volumetric solids content of $V_S = 0.6\,m^3$ solids m^{-3} soil. This yields a volumetric water/solids ratio, θ/V_S, of 0.25. If the same soil has a higher total porosity of $\Phi = 0.5$ (less compacted), the same ε as before will give a θ/V_S ratio of $0.25/0.5 = 0.5$ (double the ratio), which explains the markedly higher water blockage effects and smaller D_p/D_0 at larger Φ.

The Millington and Quirk (1961) model performs well in some cases. Equation 8.26 normally predicts D_p/D_0 well in uniform, coarse sandy soils (not included in Figure 8.6), since the water blockage effects will be more pronounced in media with larger particles and lower number of pores and bottlenecks between pores, explaining the sometimes higher N values for sandy media (Thorbjørn et al., 2008; Hamamoto et al., 2009a; Resurreccion et al., 2009).

It is further noted that in the case of aggregated soil (bimodal pore-size distribution), it will again be necessary to apply a two-region model, with the gas diffusivity (D_p/D_0) in the interaggregate pore space described by Equation 8.24 with Φ set equal to the interaggregate porosity Φ_1 (Jones et al., 2003) and $D_p(\varepsilon)/D_0$ in the intraaggregate pore space described by a linear relation (Resurreccion et al., 2009).

8.6.1.3 Undisturbed Soil

In the case of undisturbed soil, Werner et al. (2004) suggested that the WLR (Marshall) model, Equation 8.25, was the most accurate compared to a number of classical, soil-type independent models (including Equation 8.26) when tested against field-measured D_p/D_0 across soil types and moisture conditions. However, to some extent soil structure will influence gas diffusivity in undisturbed soil compared to sieved, repacked soil. Moldrup et al. (1999) therefore suggested using the Campbell pore-size distribution parameter b as a soil structure index for gas diffusivity. The Campbell b equals the negative slope of the soil-water retention curve in a pF (i.e., $\log -\psi$) versus $\log(\theta)$ plot, where ψ is soil-water matric potential in $cm\,H_2O$ and θ is volumetric soil-water content in $m^3\,H_2O\,m^{-3}$ soil (Campbell, 1974; Jury and Horton, 2004),

$$b = \frac{-d[\log(-\psi(\theta))]}{d[\log(\theta)]}, \tag{8.27}$$

where b will always be >0. The Campbell soil-water retention model normally performs best in the pF interval between 1.2 and 3.2. Moldrup et al. (1999) introduced and successfully tested a power law model for D_p/D_0 with the exponent equal to 2 + (3/b), following the capillary tube model for hydraulic conductivity by Burdine (1953),

$$\frac{D_p}{D_0} = \left(\frac{D_p^\star}{D_0} \right) \left(\frac{\varepsilon}{\varepsilon^\star} \right)^{2+(3/b)}, \tag{8.28}$$

where

$\varepsilon^\star$ is the air-filled porosity at $D_p = D_p^\star$

$D_p^\star$ is the soil-gas diffusion coefficient at a given reference-point soil-water content (e.g., zero water content equal to air saturation) or a given soil-water matric potential

At pF 2 ($-100\,cm\,H_2O$ of soil-water matric potential), where only the larger pores of $>30\,\mu m$ equivalent pore diameter are air filled, a good relation ($r^2 = 0.97$) between D_p/D_0 at pF 2 (D_{p100}/D_0) and air-filled porosity at pF 2 (ε_{100}) was found for 126 differently textured, undisturbed soils (Moldrup et al., 2000b),

$$\frac{D_{p100}}{D_0} = 2\varepsilon_{100}^3 + 0.04\varepsilon_{100}. \tag{8.29}$$

Using this as reference-point gas diffusivity in the Moldrup et al. (1999) b-dependent model yields for undisturbed, less-aggregated soil (Moldrup et al., 2000b):

$$\frac{D_p}{D_0} = (2\varepsilon_{100}^3 + 0.04\varepsilon_{100}) \left(\frac{\varepsilon}{\varepsilon^\star} \right)^{2+(3/b)}. \tag{8.30}$$

This macroporosity (ε_{100}) and b-dependent model well predicted gas diffusivity in differently textured, undisturbed soils across a wide range of soil and moisture conditions (Moldrup et al., 2000b, 2004; Rolston and Moldrup, 2002; Kawamoto et al., 2006b).

Recently, Moldrup et al. (2005) developed soil-gas diffusivity models for undisturbed soil linked to the van Genuchten (1980) soil-water retention model to establish a more direct and detailed link between pore-size distribution and soil-gas diffusivity. However, these models are presently more useful for description (curve-fitting to measured data) than prediction due to the combined high number of soil-water retention and pore tortuosity–connectivity parameters involved.

The above models represent only a few, although likely the most precise and useful currently, among the numerous predictive models available for soil-gas diffusivity. A more thorough listing and discussion of predictive and descriptive $D_p(\varepsilon)/D_0$ models can be found in Tick et al. (2007) and Allaire et al. (2008).

When using the predictive equations for relative gas diffusivity in simulation models for diffusive gas transport, the D_0 needs to be known for the actual gas under consideration at the given temperature. For example, the trace gas diffusion coefficient of oxygen in air at 20°C equals the trace gas diffusion coefficient of methane in air at 25°C, equal to $0.205\,cm^2\,s^{-1}$ (Andrussow and Schramm, 1969; Kruse et al., 1996). Further values for the binary diffusion coefficients of common gases at standard temperature and pressure are provided by Rolston and Moldrup (2002) that
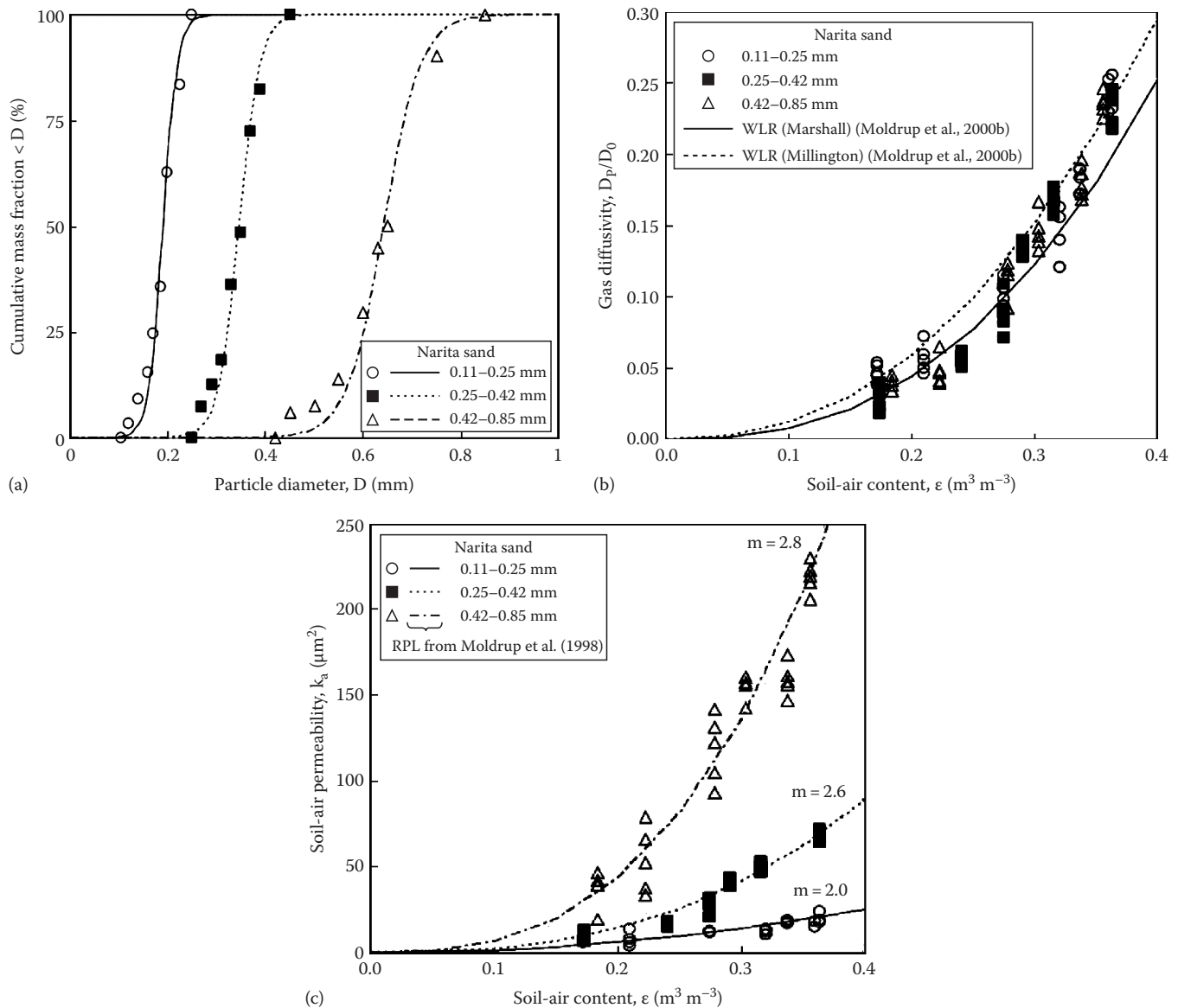
FIGURE 8.7 Relation between basic soil physical properties and gas transport parameters for three size fractions of Narita sand. (a) particle-size distribution, (b) gas diffusivity, and (c) air permeability. Also shown are predictions by (b) the WLR type of D_P/D_0 models (Moldrup et al., 2000b), and (c) a general power law model for k_a (Moldrup et al., 1998). (Data from Hamamoto, S., P. Moldrup, K. Kawamoto, and T. Komatsu. 2009a. Effect of particle size and soil compaction on gas transport parameters in variably saturated, sandy soils. Vadose Zone J. 8:986–995.)

also provides references for additional gas mixtures and an equation (Fuller et al., 1966) to calculate D_0 for nonpolar gases from molar mass and volume, pressure, and temperature.

8.6.2 Air Permeability and Soil-Gas Dispersion Coefficient

Robust predictive models for air permeability (k_a) and the soil-gas dispersion coefficient (D_H) are presently not available across soil type and scale. Both convective gas transport parameters are highly sensitive to small changes in particle and pore-size distribution and soil structure. This is contrary to the behavior for gas diffusivity (D_P/D_0, previous section) where the influence

of pore-size distribution and soil structure is minor, especially for repacked soil (Moldrup et al., 2001). This is illustrated in Figure 8.7 for three different particle-size fractions of a sandy soil (Narita sand, Japan, data from Hamamoto et al., 2009a). The particle-size distribution of the three size fractions are shown in Figure 8.7a. Despite the differences in particle (and thus pore)-size distribution, the difference in measured gas diffusivity (D_P/D_0) curves between the three size fractions are minor, and the WLR model based on either Millington (1959) or Marshall (1959) at dry conditions and with $N = 1$ (Equation 8.25) well predict the data (Figure 8.7b). In contrast to this result, the effect of particle (and thus pore)-size distribution on air permeability, k_a (μm^2), is dramatic (Figure 8.7c).

8.6.2.1 Air Permeability

Moldrup et al. (1998, 2003) suggested a reference-point power law model for $k_a(\varepsilon)$, analogous to Equation 8.28 for D_p/D_0,

$$k_a = k_a^* \left(\frac{\varepsilon}{\varepsilon^*} \right)^m,$$
(8.31)

where

k_a is the soil-air permeability (μm^2)

ε^* is the volumetric soil-air content

k_a^* is the air permeability at a given reference point of moisture content or soil-water matric potential

m is a tortuosity–connectivity parameter for convective gas transport

If the reference point is again (like for soil-gas diffusivity, Equation 8.30) chosen at pF 2 (–100 cm H_2O), the $k_a(\varepsilon)$ model becomes

$$k_a = k_{a100} \left(\frac{\varepsilon}{\varepsilon_{100}} \right)^m.$$
(8.32)

For the three particle-size fractions in Figure 8.7c, both k_{a100} and m markedly increase with particle size and thus median air-filled pore size. Kawamoto et al. (2006a) developed a predictive equation for k_{a100} from ε_{100}, and D_{p100} (applying Equation 8.29 for D_{p100}) by combining Fick's law for diffusive transport with Poiseuille's law for convective fluid transport and assuming soil pores to be uniform, tortuous and nonjointed tubes of similar diameter to give

$$k_{a100} = \left(\frac{d_{100}^2}{32} \right) \frac{D_{p100}}{D_0} = \left(\frac{d_{100}^2}{32} \right) (2\varepsilon_{100}^3 + 0.04\varepsilon_{100})$$
(8.33)

where the average equivalent pore diameter for convective gas transport at pF 2 was found to be $d_{100} = 150\,\mu m$ (i.e., around five times the minimum drained pore size at pF 2), Inserting this in Equation 8.33 gives

$$k_{a100} = 700(2\varepsilon_{100}^3 + 0.04\varepsilon_{100}).$$
(8.34)

Kawamoto et al. (2006a) suggested an expression for the tortuosity–connectivity parameter for convective gas transport, m in Equation 8.32, based on the Alexander and Skaggs (1986) capillary tube model for unsaturated water permeability, m = 1 + (3/b), where b is the Campbell pore-size distribution parameter (Equation 8.27). Combining Equations 8.32 and 8.34 and inserting m = 1 + (3/b) yield

$$k_a = 700(2\varepsilon_{100}^3 + 0.04\varepsilon_{100}) \left(\frac{\varepsilon}{\varepsilon_{100}} \right)^{1+(3/b)}.$$
(8.35)

This $k_a(\varepsilon)$ model described well the air permeability for differently textured undisturbed soils (100 cm³ scale) sampled from two deep vadose zone profiles. If Campbell b is not known, average b values of 3 for sandy soils (m = 2), 6 for silty soils (m = 1.5), and 10 for clayey soils (m = 1.3) are recommended. For coarse porous media like the two largest particle-size fractions of Figure 8.7, m values significantly higher than 2 can be observed (Figure 8.7c), and the d_{100} value (Equation 8.33) likely exceeds $150\,\mu m$. Compared to soil-gas diffusivity, the prediction inaccuracy for k_a is high (order of magnitude) and likely highly scale dependent. It is again stressed that reliable models for k_a across soil types and scales are not available, and this parameter if possible should be measured in situ.

8.6.2.2 Dispersion Coefficient

The same lack of models are evident for the soil-gas dispersion coefficient, D_H (m² s⁻¹) where very limited amount of studies and measurements have been carried out since the first major study by Rolston et al. (1969). Like for gas diffusivity and air permeability, Hamamoto et al. (2009b) found different behavior and $D_H(\varepsilon)$ curve shapes for structureless (unimodal) and aggregated (bimodal) soils packed in 30 or 60 cm columns. They also for the first time documented quantitative links between the soil-gas diffusion coefficient, air (gas) permeability, and soil-gas dispersion coefficient for structureless soils and in the interaggregate region of aggregated soils. They found that the soil-gas dispersivity (λ, in m),

$$\lambda = \frac{D_H}{\bar{v}_g},$$
(8.36)

where $\bar{v}_g$ is the pore-gas velocity, was closely related to the soil-gas diffusion coefficient and air-filled porosity. The best-fit relation was

$$\lambda = 0.0185 \left(\left(\frac{\varepsilon D_0}{D_p} \right)^{0.5} - 1 \right).$$
(8.37)

Combining Equations 8.36 and 8.37 with the WLR (Marshall) model for D_p, Equation 8.25, yields

$$D_H = \bar{v}_g \times 0.0185 \left(\left(\frac{\Phi}{\varepsilon^{1.5}} \right)^{0.5} - 1 \right),$$
(8.38)

where

D_H is in m² s⁻¹

$\bar{v}_g$ in m s⁻¹

This relation needs further testing and modification against a wider range of soil types, for undisturbed soils, and at different

measurement scales. Presently, Equation 8.38 is therefore suggested to be used only as an order of magnitude estimate for D_H.

8.7 Applications

There have been many studies to evaluate the applicability of the various models and equations for simulating the movement of gases within porous media for a whole range of chemicals, porous media type, and applications. For soil science and other disciplines, the bulk of the applications have used Fick's laws for modeling purposes. However, as noted above, Fick's laws are only applicable under a limited set of conditions. From the results of multiple studies, some general statements can be made about the applicability of the various models for various modeling situations. For cases of multinary system of gases and where the diffusion coefficients of the diffusing gases are very much different from one another, the Stephan–Maxwell equations would be more appropriate than Fick's law. Neither Fick's law nor the Stephan–Maxwell equations consider Knudsen diffusion. Thus, for porous media with very small pores and with small air permeabilities where Knudsen diffusion may be important, both of these models may be inadequate to model gas transport. The dividing line between molecular and Knudsen diffusion is still quite uncertain, but there is some evidence that Knudsen diffusion becomes important and should be included in models for air permeability less than 10^{-13} m^2 (Sleep, 1998; Webb and Pruess, 2003). The air permeability for many natural soils at field-water content range from about 10^{-10} to 10^{-12} m^2, except for heavy clay soils (>20% clay), where the air permeability may be as low as 10^{-14} m^2 (Moldrup et al., 1998; Kawamoto et al., 2006a). Fick's law and the Stefan–Maxwell equations also may prove inadequate when nonequimolar fluxes result in large pressure gradients and subsequent large viscous fluxes. For porous media with small air permeability, the DGM is generally considered to be superior over the other models because it not only considers Knudsen diffusion but also is applicable to multinary mixtures of counter diffusing gases (Abriola and Pinder, 1985; Thorstenson and Pollock, 1989a, 1989b; Massmann and Farrier, 1992; Sleep, 1998; Webb, 2006).

For cases where either naturally occurring or imposed pressure gradients occur within porous media, both advection and diffusion are controlling processes of gas transport. For high permeability and low concentration of the gas of interest, it is generally considered that the convection–dispersion equation is adequate for modeling. However, if concentration of the gas of interest is high and nonequimolar fluxes may occur, the DGM (with an advective component added) may be more appropriate (Abu-El-Sha'r and Abriola, 1997). For nonisobaric (pressure gradients) conditions and low permeability, the DGM is needed and is generally considered to be superior over the convection–dispersion equation (Webb, 2006). There is still a debate about the applicability of the various models to real porous media, and there is still large uncertainty in the appropriate parameter values to use in the models.

8.7.1 Soil Aeration

As discussed above, soil aeration is generally considered to be the exchange of gases between the soil and the atmosphere, but more specifically, it is the transport of O_2 from the atmosphere to the soil microbial and root respiration sites and the exchange of generated CO_2 from the soil to the atmosphere. The ratio of the production of CO_2 (moles) to the consumption of O_2 (moles) is called the respiration quotient (Freijer and Leffelaar, 1996). The respiration quotient is determined by the C–H–O ratio and the fraction of the carbon sources that is mineralized to CO_2. The respiration quotient varies from 0.8 to 1.1 for most constituents of soil organic matter (Freijer et al., 1995).

Soil respiration is basically a ternary system of diffusion of N_2, O_2, and CO_2 (Argon neglected). Fick's first and second laws have often been used in an effort to quantify soil respiration from measurements of the concentration gradients of O_2 and CO_2 within soil profiles. For example, de Jong and Schappert (1972) used field-measured CO_2 gradients to calculate the flux of CO_2 in order to estimate soil respiration rates. Freijer and Leffelaar (1996) used the Stefan–Maxwell equations to evaluate soil respiration. They concluded that only using Fick's law for a fairly wide range of the respiration quotient and the O_2 consumption rate that the errors in the CO_2 gradients were small (<5%), but that the errors in the O_2 gradients were substantial. Freijer and Leffelaar (1996) showed through comparison of Fick's law with the Stefan–Maxwell calculations that there will be no error in the O_2 gradient only when the respiration quotient is close to 0.789. Thus, using Fick's law and the gradient of CO_2, as did de Jong and Schappert (1972), soil respiration rate may be estimated reasonably well, but errors will occur if the gradient of O_2 is used to calculate the flux. Leffelaar (1987, 1988) developed a method to correct the Fickian fluxes by considering the diffusive flux to consist of the Fickian component and a correction term to account for the effect of the presence of other gases (nonequimolar diffusion).

8.7.2 Greenhouse Gas Emissions

The three most prevalent greenhouse gases that may be emitted from soils are CO_2, N_2O, and CH_4. The most common methodology to quantify the emissions of these gases is to measure the surface flux by placing chambers over the soil surface. Transport in soil is not considered for this methodology except to correct the diffusion flux due to nonlinear behavior of the chamber concentration versus time curves (Livingston et al., 2006; Venterea and Baker, 2008). As for soil aeration, one could use Fick's law or some other theory to calculate the diffusion fluxes from measured concentration gradients within the soil near the soil surface. Since the greenhouse gases are generally in trace amounts in soil compared to the other soil gases, one may consider a binary system comprised of a trace gas diffusing into air (especially near the soil surface). Thus, the effective diffusion coefficient from Fick's law would be equal to the binary diffusion coefficient (Jaynes and Rogowski, 1983), and Fick's law could be

used to calculate the fluxes. However, the large spatial variability of greenhouse gas concentration profiles in soils generally makes this estimate of emission very uncertain.

8.7.3 Soil Fumigation

Various fumigants, including methyl bromide and 1, 3-dichloropropene, have been extensively used in agriculture to control soilborne pests such as nematodes. Fumigants are generally applied by shank injection into the soil and the soil surface often covered with some kind of plastic tarping material in an attempt to slow the transport of the gas from the soil to the atmosphere in order to attain effective control of the pests. Although some of the early models only considered diffusive transport using Fick's law (Rolston et al., 1982), the process of gas movement within the soil can be by both diffusion and advection (from the injection and volatilization process, wind, barometric pressure changes). A later model (Chen et al., 1995) used a modified convection–dispersion equation to model the transport of a fumigant with barometric pressure fluctuations. Wang et al. (1997, 2000, 2004) used a convection–dispersion model coupled to a heat, water, and transport model to evaluate the movement of fumigants in soils. Although Chen et al. (1995) indicate good agreement between model simulations and measurement of surface volatilization of the fumigant, this case of high concentration of fumigant, cotransport of water vapor (Leffelaar, 1987; Petersen et al., 1996), and advective transport due to pressure gradients may be a situation where the DGM could have been applied.

Model applications include simulating scenarios of injection depth and either plastic film or high soil-water content near the soil surface in order to attempt to decrease the flux of fumigant out of soil and into the atmosphere (Allaire et al., 2004). Potentially as equally important as using the most appropriate gas transport model is to use the most appropriate predictive equation for D_p. For example, the Millington and Quirk (1961) relation would underestimate the soil-gas diffusion coefficient at the higher water contents.

8.7.4 Volatile Organic Compounds and Vapor Extraction

In recent years, there has been a lot of interest in the movement of VOCs from contaminated soil sites to sites where humans may be exposed to these chemicals. Also, soil-vapor extraction has been used as a means to remediate contaminated sites by removing and treating the vapor phase (Falta, 2006). The transport mechanism for soil-vapor extraction is primarily advection, and Darcy's law has been extensively used (Falta, 2006). Many VOCs have molecular weights much different than air, and consequently their diffusion coefficients are much different than air. Also, the vapor concentration of these chemicals in soil can also be quite high especially near the contaminated soil or aquifer. Since some of the volatile chemicals are also being biodegraded in the subsurface, O_2 will be consumed and CO_2 produced, setting up a complex transport system of nonequimolar gases. Investigators also have used the O_2 and CO_2 concentration gradients in soil in an attempt to estimate the rate that VOCs or hydrocarbons are being biodegraded. Thus, because of these properties, it is generally considered that Fick's law would not be a valid approach to modeling the movement of these gases, and the Stefan–Maxwell equations (Amali and Rolston, 1993; Amali et al., 1996; Evans et al., 2004), the DGM (Massmann and Farrier, 1992; Sleep, 1998; Webb and Pruess, 2003; Fen and Abriola, 2004), or "adjusted" Fick's law (Freijer and Leffelaar, 1996; van de Steene and Verplancke, 2006) are deemed more appropriate. This application will usually require knowledge of all three gas transport parameters (D_p, k_a, and D_H). Some studies (Sleep, 1998; Fen and Abriola, 2004) show that the DGM is superior to the others for modeling the movement of VOCs in soils.

8.7.5 Radon Transport into Structures

Radon gas is formed in soil from the radioactive decay of naturally occurring radium within the upper few meters of the vadose zone. Radon is a significant human health concern because of radon entering buildings and the inhalation of the gas that may potentially cause lung cancer. Radon moves through soil primarily by diffusion except in the near proximity to buildings. Near buildings, radon enters buildings primarily through cracks in basement floors and walls by advective gas transport due to pressure differences within and below the building (Nazaroff, 1992). This advective flow is driven by barometric pressure changes, wind, temperature differences, and by operation of heating and air-conditioning systems. It is estimated that diffusion accounts for about 10% of the transport into single family structures and that advection contributes the other 90% (Nazaroff, 1992). Since radon gas occurs in very small concentrations, it may be considered to be a trace gas and Fick's law for describing the diffusion part of the transport process should be applicable. For advection, Darcy's law has been used to describe the gas flow into and around buildings (Riley et al., 1996; Robinson et al., 1997). For a thorough discussion of radon transport through porous media and into buildings, see Nazaroff (1992). Again, this application will usually require knowledge of all three gas transport parameters (D_p, k_a, and D_H).

8.7.6 Volatile Organic Vapor Transport into Structures

A focus area for soil-gas transport in relation to human health stems from the numerous contaminated soil sites around the world, most of which are in urban areas, causing intrusion of volatile organic vapors (e.g., gasoline compounds such as benzene and chlorinated compounds such as trichloroethylene) from soil into buildings. Moisture conditions below buildings greatly impact vapor intrusion due to moisture controls on the soil-gas diffusivity, dispersion coefficient, and air permeability (Tillman and Weaver, 2007). The pathways and controls will be similar to those discussed for radon (previous section). Two technology

barriers preventing realistic assessment of indoor air problems from vapor intrusion are (a) lack of knowledge of actual moisture conditions and its temporal and spatial variability under and around buildings and (b) current risk assessment models for soil contribution to indoor air concentrations (e.g., USEPA, 1996; DKEPA, 2002) where general simulation tools for predicting diffusive and convective gas transport have only considered traditional models such as the Millington and Quirk model for gas diffusivity (Equation 8.26) that do not reflect intact soil conditions across soil types and moisture conditions. However, the new generation of simulation models for the soil vadose zone including the HYDRUS-1D and 2D packages (Simunek et al., 2008) allows for use of more realistic models based on transport parameters in both soil liquid and gaseous phases and will in time facilitate more realistic simulations of vapor transport and intrusion as well as the other key application areas of gas transport theory discussed in this chapter.

References

Abriola, L.M., and G.F. Pinder. 1985. A multiphase approach to the modeling of porous media contamination by organic compounds: Numerical simulation. Water Resour. Res. 21:19–26.

Abu-El-Sha'r, W., and L.M. Abriola. 1997. Experimental assessment of gas transport mechanisms in natural porous media: Parameter evaluation. Water Resour. Res. 33:505–516.

Alexander, L., and R.W. Skaggs. 1986. Predicting unsaturated hydraulic conductivity from water characteristic data. Trans. ASAE 29:176–184.

Allaire, S.E., J.A. Lafond, A.R. Cabral, and S.F. Lange. 2008. Measurement of gas diffusion through soil: Comparison of laboratory methods. J. Environ. Monit. 10:1326–1336.

Allaire, S.E., S.R. Yates, and F.F. Ernst. 2004. Effect of soil moisture and irrigation on propargyl bromide volatilization and movement in soil. Vadose Zone J. 3:656–667.

Amali, S., and D.E. Rolston. 1993. Theoretical investigation of multicomponent volatile organic vapor diffusion: Steady-state fluxes. J. Environ. Qual. 22:825–831.

Amali, S., D.E. Rolston, and T. Yamaguchi. 1996. Transient multicomponent gas-phase transport of volatile organic chemicals in porous media. J. Environ. Qual. 25:1041–1047.

Andrussow, L., and B. Schramm. 1969. Eigenschafen der materie in ihren aggregatzuständen: 5. Transportphänomenen I (viskosität und diffusion). *In* H. Borchers, H. Hausen, K.H. Hellwege, K.L. Schäfer, and E. Smidt (eds.) Zahlenwerte und funktionen aus physik-chemie-astronomie-geophysik-und technik. Springer-Verlag, Berlin, Germany.

Baehr, A.L., G.E. Hoag, and M.C. Marley. 1989. Removing volatile contaminants from the unsaturated zone by inducing advective air-phase transport. J. Contam. Hydrol. 4:1–26.

Ball, B.C., W. Harris, and J.R. Burford. 1981. A laboratory method to measure gas diffusion in soil and other porous materials. J. Soil Sci. 32:323–333.

Ball, B.C., and P. Schjønning. 2002. Air permeability, p. 1141–1158. *In* G.C. Topp and J.H. Dane (eds.) Methods of soil analysis, Part 4, Physical methods. SSSA, Madison, WI.

Bird, R.B., W.E. Stewart, and E.N. Lightfoot. 2007. Transport phenomena. Revised 2nd edn. John Wiley & Sons, New York.

Buckingham, E. 1904. Contributions to our knowledge of the aeration of soils. USDA, Bureau of soils Bulletin 25. U.S. Government Printing Office, Washington, DC.

Burdine, N.T. 1953. Relative permeability calculations from pore size distribution data. Trans. AIME 198:71–78.

Campbell, G.S. 1974. A simple method for determining unsaturated conductivity from moisture retention data. Soil Sci. 117:311–314.

Carslaw, H.S., and J.C. Jaeger. 1959. Conduction of heat in solids. 2nd edn. Clarendon Press, Oxford, U.K.

Chapman, S., and T.G. Cowling. 1970. The mathematical theory of non-uniform gases. 3rd edn. Cambridge University Press, Cambridge, MA.

Chen, C., R.E. Green, D.M. Thomas, and J.A. Knuteson. 1995. Modeling 1,3-dichloropropene fumigant volatilization with vapor-phase advection in the soil profile. Environ. Sci. Technol. 29:1816–1821.

Corey, A.T. 1994. Mechanics of immiscible fluids in porous media. Water Resources Publications, LLC, Highlands Ranch, CO.

Costanza-Robinson, M.S., and M.L. Brusseau. 2002. Gas phase advection and dispersion in unsaturated porous media. Water Resour. Res. 38:1036.

Costanza-Robinson, M.S., and M.L. Brusseau. 2006. Gas-phase dispersion in porous media, p. 121–132. *In* C.K. Ho and S.W. Webb (eds.) Gas transport in porous media. Vol. 20. Theory and applications of transport in porous media. Springer, Dordrecht, the Netherlands.

Cunningham, R.E., and R.J.J. Williams. 1980. Diffusion in gases and porous media. Plenum Press, New York.

Currie, J.A. 1960. Gaseous diffusion in porous media. Part 1. A non-steady state method. Br. J. Appl. Phys. 11:314–317.

Curtiss, C.F., and J.O. Hirschfelder. 1949. Transport properties of multicomponent gas mixtures. J. Chem. Phys. 17:550–555.

de Jong, E., and H.J.V. Schappert. 1972. Calculation of soil respiration and activity from CO_2 profiles in the soil. Soil Sci. 113:328–333.

DKEPA. 2002. Guidelines on the remediation of contaminated sites. Environmental guidelines no. 7. Danish Ministry of the Environment, Copenhagen, Denmark.

Evans, C.A., K.S. McLeary, G.P. Partridge, Jr., and R.S. Huebner. 2004. Modeling the impact of multicomponent VOCs on ground water using the Stefan–Maxwell equation. J. Am. Water Resour. Assoc. 40:409–417.

Falta, R. 2006. Environmental remediation of volatile organic compounds, p. 353–370. *In* C.K. Ho and S.W. Webb (eds.) Gas transport in porous media. Vol. 20. Theory and applications of transport in porous media. Springer, Dordrecht, the Netherlands.

Farrell, R.E., E. de Jong, and J.A. Elliott. 2002. Gas sampling and analysis, p. 1141–1158. *In* G.C. Topp and J.H. Dane (eds.) Methods of soil analysis, Part 4, Physical methods. SSSA, Madison, WI.

Fen, C.S., and L.M. Abriola. 2004. A comparison of mathematical model formulations for organic vapor transport in porous media. Adv. Water Resour. 27:1005–1016.

Flühler, H. 1973. Sauerstafdiffusion in boden. Mitt. Schweiz. Anst. Forstl. Versuchs wes. 49:125–250.

Freijer, J.I., S.C. Dekker, and W. Bouten. 1995. Gas activity in soils: A simulation model to study gas dynamics in soils. FGBL Report. 53. Landscape and Environmental Research Group, Amsterdam, the Netherlands.

Freijer, J.I., and P.A. Leffelaar. 1996. Adapted Fick's law applied to soil respiration. Water Resour. Res. 32:791–800.

Fuller, E.N., P.D. Schettler, and J.C. Giddings. 1966. A new method for prediction of binary gas-phase diffusion coefficients. Ind. Eng. Chem. 58:19–27.

Glinski, J., and W. Stepniewski. 1985. Soil aeration and its role for plants. CRC Press, Boca Raton, FL.

Gradwell, M.W. 1961. A laboratory study of the diffusion of oxygen through pasture topsoils. N. Z. J. Sci. 4:250–270.

Green, R.D., and S.J. Fordham. 1975. A field method for determining the air permeability in soil, p. 273–278. *In* Soil physical conditions and crop production. Technical Bulletin 29. HMSO, London, U.K.

Grew, K.E., and T.L. Ibbs. 1952. Thermal diffusion in gases. Cambridge University Press, Cambridge, MA.

Grover, B.L. 1955. Simplified air permeameters for soil in place. Soil Sci. Soc. Am. Proc. 19:414–418.

Hamamoto, S., P. Moldrup, K. Kawamoto, and T. Komatsu. 2009a. Effect of particle size and soil compaction on gas transport parameters in variably saturated, sandy soils. Vadose Zone J. 8:986–995.

Hamamoto, S., P. Moldrup, K. Kawamoto, T. Komatsu, and D.E. Rolston. 2009b. Unified measurement system for the gas dispersion coefficient, air permeability, and gas diffusion coefficient in variably-saturated soil. Soil Sci. Soc. Am. J. 73:1921–1930.

Ho, C.K., and S.W. Webb (eds.) 2006. Gas transport in porous media. Springer, Dordrecht, the Netherlands.

Iversen, B.V., P. Schjønning, T.G. Poulsen, and P. Moldrup. 2001. In situ, on-site and laboratory measurements of soil air permeability: Boundary conditions and measurement scale. Soil Sci. 166:97–106.

Jaynes, D.B., and A.S. Rogowski. 1983. Applicability of Fick's law to gas diffusion. Soil Sci. Soc. Am. J. 47:425–430.

Jones, S.B., D. Or, and G.E. Bingham. 2003. Gas diffusion measurement and modeling in coarse-textured porous media. Vadose Zone J. 2:602–610.

Jury, W.H., and R. Horton. 2004. Soil physics. 6th edn. John Wiley & Sons, Hoboken, NJ.

Kawamoto, K., P. Moldrup, P. Schjønning, B.V. Iversen, T. Komatsu, and D.E. Rolston. 2006a. Gas transport parameters in the vadose zone: Development and tests of power-law models for air permeability. Vadose Zone J. 5:1205–1215.

Kawamoto, K., P. Moldrup, P. Schjønning, B.V. Iversen, D.E. Rolston, and T. Komatsu. 2006b. Gas transport parameters in the vadose zone: Gas diffusivity in field and lysimeter soil profiles. Vadose Zone J. 5:1194–1204.

Kirkham, D. 1946. Field method for determination of air permeability of soil in its undisturbed state. Soil Sci. Soc. Am. Proc. 11:93–99.

Klinkenberg, L.J. 1941. The permeability of porous media to liquids and gases, p. 200–213. *In* Drilling and production practices. American Petroleum Institute, Washington, DC.

Kruse, C.W., P. Moldrup, and N. Iversen. 1996. Modeling diffusion and reaction in soils: II. Atmospheric methane diffusion and consumption in a forest soil. Soil Sci. 161:355–365.

Leffelaar, P.A. 1987. Dynamic simulation of multinary diffusion problems related to soil. Soil Sci. 143:79–91.

Leffelaar, P.A. 1988. Dynamics of partial anaerobiosis, denitrification, and water in a soil aggregate: Simulation. Soil Sci. 143:79–91.

Liang, P., C.C. Bowers, and H.D. Bowen. 1995. Finite element model to determine the shape factor for soil air permeability measurements. Trans. ASAE 38:997–1003.

Livingston, G.P., G.L. Hutchinson, and K. Spartalian. 2006. Trace gas emission in chambers: A non-steady-state diffusion model. Soil Sci. Soc. Am. J. 70:1459–1469.

Marshall, T.J. 1959. The diffusion of gases through porous media. J. Soil Sci. 10:79–82.

Mason, E.A., and A.P. Malinauskas. 1983. Gas transport in porous media: The dusty-gas model. Chemical engineering monographs 17. Elsevier, New York.

Mason, E.A., A.P. Malinauskas, and R.B. Evans. 1967. Flow and diffusion of gases in porous media. J. Chem. Phys. 46:3199–3216.

Massmann, J.W., and D.F. Farrier. 1992. Effects of atmospheric pressures on gas transport in the vadose zone. Water Resour. Res. 28:777–791.

Millington, R.J. 1959. Gas diffusion in porous media. Science 130:100–102.

Millington, R.J., and J.M. Quirk. 1961. Permeability of porous solids. Trans. Faraday Soc. 57:1200–1207.

Moldrup, P., T. Olesen, J. Gamst, P. Schjønning, T. Yamaguchi, and D.E. Rolston. 2000a. Predicting gas diffusivity in sieved, repacked soil: Water-induced linear reduction model. Soil Sci. Soc. Am. J. 64:1588–1594.

Moldrup, P., T. Olesen, T. Komatsu, P. Schjønning, and D.E. Rolston. 2001. Tortuosity, diffusivity, and permeability in the soil liquid and gaseous phases. Soil Sci. Soc. Am. J. 65:613–623.

Moldrup, P., T. Olesen, P. Schjønning, T. Yamaguchi, and D.E. Rolston. 2000b. Predicting the gas diffusion coefficient in undisturbed soil from soil water characteristics. Soil Sci. Soc. Am. J. 64:94–100.

Moldrup, P., T. Olesen, T. Yamaguchi, P. Schjønning, and D.E. Rolston. 1999. Modeling diffusion and reaction in soils: IX. The Buckingham-Burdine-Campbell equation for gas diffusivity in undisturbed soil. Soil Sci. 164:542–551.

Moldrup, P., T. Olesen, S. Yoshikawa, T. Komatsu, and D.E. Rolston. 2004. Three-porosity model for predicting the gas diffusion coefficient in undisturbed soil. Soil Sci. Soc. Am. J. 68:750–759.

Moldrup, P., T. Olesen, S. Yoshikawa, T. Komatsu, and D.E. Rolston. 2005. Predictive-descriptive models for gas and solute diffusion coefficients in variably saturated porous media coupled to pore-size distribution: II. Gas diffusivity in undisturbed soil. Soil Sci. 170:854–866.

Moldrup, P., T.G. Poulsen, P. Schjønning, T. Olesen, and T. Yamaguchi. 1998. Gas permeability in undisturbed soil. Measurements and predictive model. Soil Sci. 163: 180–189.

Moldrup, P., S. Yoshikawa, T. Olesen, T. Komatsu, and D.E. Rolston. 2003. Air permeability in undisturbed volcanic ash soils: Predictive model tests and soil structure fingerprints. Soil Sci. Soc. Am. J. 67:32–40.

Nazaroff, W.W. 1992. Radon transport from soil to air. Rev. Geophys. 30:137–160.

Papendick, R.I., and J.R. Runkles. 1965. Transient-state oxygen diffusion in soil: I. The case where rate of oxygen consumption is constant. Soil Sci. 100:251–261.

Penman, H.L. 1940. Gas and vapor movement in soil: The diffusion of vapors through porous solids. J. Agric. Sci. 30:437–462.

Petersen, L.W., Y.H. El-Farhan, P. Moldrup, D.E. Rolston, and T. Yamaguchi. 1996. Transient diffusion, adsorption and emission of volatile organic vapors in soils with fluctuating water contents. J. Environ. Qual. 25:1054–1063.

Petersen, L.W., P. Moldrup, Y.H. El-Farhan, O.H. Jacobsen, T. Yamaguchi, and D.E. Rolston. 1995. The effect of moisture and soil texture on the adsorption of organic vapors. J. Environ. Qual. 24:752–759.

Petersen, L.W., D.E. Rolston, P. Moldrup, and T. Yamaguchi. 1994. Volatile organic vapor diffusion and adsorption in soils. J. Environ. Qual. 23:799–805.

Poulsen, T.G., W. Suwarnarat, M.K. Hostrup, and N.V. Prasad. 2008. Simple and rapid method for measuring gas dispersion in porous media: Methodology and applications. Soil Sci. 173:169–174.

Rathfelder, K., J.R. Lang, and L.M. Abriola. 1995. Soil vapor extraction and bioventing: Applications, limitations, and future research directions. Rev. Geophys. 33:1067–1081.

Reinecke, S.A., and B.E. Sleep. 2002. Knudsen diffusion, gas permeability, and water content in an unconsolidated porous medium. Water Resour. Res. 38:1280.

Resurreccion, A.C., P. Moldrup, K. Kawamoto, S. Hamamoto, D.E. Rolston, and T. Komatsu. 2010. Hierarchical, bimodal model for gas diffusivity in aggregated, unsaturated soils. Soil Sci. Soc. Am. J. 74:481–491.

Resurreccion, A.C., P. Moldrup, K. Kawamoto, S. Yoshikawa, D.E. Rolston, and T. Komatsu. 2008. Variable pore connectivity factor model for gas diffusivity in unsaturated, aggregated soil. Vadose Zone J. 7:397–405.

Riley, W.J., A.J. Gadgil, Y.C. Bonnefous, and W.W. Nazaroff. 1996. The effects of steady winds on radon-222 entry from soil into houses. Atmos. Environ. 30:1167–1176.

Robinson, A.L., R.G. Sextro, and W.J. Riley. 1997. Soil-gas entry into houses driven by atmospheric pressure fluctuations—The influence of soil properties. Atmos. Environ. 31:1487–1495.

Rolston, D.E. 2005. Aeration, p. 17–21. *In* D. Hillel (ed.) Encyclopedia of soils in the environment. Elsevier, Amsterdam, the Netherlands.

Rolston, D.E., R.D. Glauz, and B.D. Brown. 1982. Comparisons of simulated with measured transport and transformation of methyl bromide gas in soils. Pestic. Sci. 13:653–664.

Rolston, D.E., D. Kirkham, and D.E. Nielsen. 1969. Miscible displacement of gases through soil columns. Soil Sci. Soc. Am. Proc. 33:488–492.

Rolston, D.E., and P. Moldrup. 2002. Gas diffusivity, p. 1141–1158. *In* G.C. Topp and J.H. Dane (eds.) Methods of soil analysis, Part 4, Physical methods. SSSA, Madison, WI.

Rust, R.H., A. Klute, and J. E. Gieseking. 1957. Diffusion-porosity measurements using a non-steady state system. Soil Sci. 84:453–463.

Schjønning, P. 1985. A laboratory method for determination of gas diffusion in soil (in Danish with English summary). Tidsskr. Planteavl. 89: Report no. S1773. The Danish Institute of Plant and Soil Science, Tjele, Denmark.

Scott, H.D. 2000. Soil physics: Agricultural and environmental applications. Iowa State University Press, Ames, IA.

Shearer, R.C., R.J. Millington, and J.P. Quirk. 1966. Oxygen diffusion through sands in relation to capillary hysteresis: 1. Calibration of oxygen cathode for use in diffusion studies. Soil Sci. 101:361–365.

Simunek, J., M.Th. van Genuchten, and M. Sejna. 2008. Development and applications of the HYDRUS and STANMOD software packages and related codes. Vadose Zone J. 7:587–600.

Sleep, B.E. 1998. Modeling transient organic vapor transport in porous media with the dusty gas model. Adv. Water Resour. 22:247–256.

Thorbjørn, A., P. Moldrup, H. Blendstrup, T. Komatsu, and D.E. Rolston. 2008. A gas diffusivity model based on air-, solid-, and water-phase resistance in variably saturated soil. Vadose Zone J. 7:1230–1240.

Thorstenson, D.C., and D.W. Pollack. 1989a. Gas transport in unsaturated zones: Multicomponent systems and the adequacy of Fick's laws. Water Resour. Res. 25:477–507.

Thorstenson, D.C., and D.W. Pollack. 1989b. Gas transport in unsaturated porous media: The adequacy of Fick's law. Rev. Geophys. 27:61–78.

Tick, G.R., C.M. McColl, I. Yolcubal, and M.L. Brusseau. 2007. Gas-phase diffusive tracer test for the in-situ measurement of tortuosity in the vadose zone. Water Air Soil Pollut. 184:355–362.

Tillman, F.D., and J.W. Weaver. 2007. Temporal moisture content variability beneath and external to a building and the potential effects on vapor intrusion risk assessment. Sci. Total Environ. 379:1–15.

USEPA. 1996. Soil screening guidance. User's guide. 2nd edn. Publ. 9355.4-23. EPA/540/R-96/018. Office of Solid Waste and Emergency Response. USEPA, Washington, DC.

van de Steene, J., and H. Verplancke. 2006. Adjusted Fick's law for gas diffusion in soils contaminated with petroleum hydrocarbons. Eur. J. Soil Sci. 57:106–121.

van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soil. Soil Sci. Soc. Am. J. 44:892–898.

van Groenewoud, H. 1968. Methods and apparatus for measuring air permeability of the soil. Soil Sci. 106:275–279.

Venterea, R.T., and J.M. Baker. 2008. Effects of soil physical non-uniformity on chamber-based gas flux estimates. Soil Sci. Soc. Am. J. 72:1410–1417.

Venterea, R.T., and D.E. Rolston. 2000. Nitric and nitrous oxide emissions following fertilizer application to agricultural soil: Biotic and abiotic mechanisms and kinetics. J. Geophys. Res. 105:15117–15129.

Wang, D., J.M. He, and J.A. Knuteson. 2004. Concentration-time exposure index for modeling soil fumigation under various management scenarios. J. Environ. Qual. 33:685–694.

Wang, D., J.A. Knuteson, and S.R. Yates. 2000. Two-dimensional model simulation of 1,3-dichloropropene volatilization and transport in a field soil. J. Environ. Qual. 29:639–644.

Wang, D., S.R. Yates, and J. Gan. 1997. Temperature effect on methyl bromide volatilization in soil fumigation. J. Environ. Qual. 26:1072–1079.

Webb, S. 2006. Gas transport mechanisms. *In* C.K. Ho and S.W. Webb (eds.) Gas transport in porous media. Vol. 20. Theory and applications of transport in porous media. Springer, Dordrecht, the Netherlands.

Webb, S.W., and K. Pruess. 2003. The use of Fick's law for modeling trace gas diffusion in porous media. Transp. Porous Media 51:327–341.

Werner, D., P. Grathwohl, and P. Höhener. 2004. Review of field methods for the determination of the tortuosity and effective gas-phase diffusivity in the vadose zone. Vadose Zone J. 3:1240–1248.

9

Soil Thermal Regime

Robert Horton
Iowa State University

Tyson Ochsner
Oklahoma State University

9.1 Introduction

What is temperature, and more specifically, what is soil temperature? Temperature is a measure of the average kinetic energy of the molecules of a substance. As kinetic energy increases, the temperature increases. At the macroscale, temperature is a physical property that determines the direction of heat flow between two substances in thermal contact. Heat flows from high temperature to low temperature. Temperature is not a measure of heat content. In order to specify the heat content of a particular substance (e.g., soil), both the heat capacity and the temperature must be known.

Why is soil temperature important? Temperature influences many biological, chemical, and physical processes that occur in the soil. Plant germination and growth are closely related to soil temperature. Below a soil temperature of about 5°C, growth of roots of most plants is negligible. When a soil freezes, most biological activities and chemical processes essentially cease. Pesticide degradation, residue decomposition, soil microbiological activity, and nutrient cycles are affected by soil temperature. Soil temperature gradients affect soil moisture and solute movement. Soil temperature affects soil hydraulic conductivity, transport properties, and important environmental processes such as soil–atmosphere gas exchanges, the global carbon cycle, and the transformations and transport of contaminants. Crop growth and evapotranspiration models require accurate submodels of the soil temperature regime. Climate modeling and remote sensing require additional ground-truth data for both soil heat flux and temperature of the surface soil.

Soil temperature is highly dynamic and is influenced by a multitude of factors. Those factors include soil thermal properties, surface mulch, tillage, precipitation, air temperature, incoming radiation, fire, vegetation, snow cover, soil color and organic-matter content, aspect and slope, elevation, and groundwater. All these factors have some influence on soil temperature, and they contribute to the great diversity of soil temperature regimes.

Soil temperature may range from highs of near 700°C at the surface during an intense forest fire (Neary et al., 1999) to lows of –20°C during the Arctic winter (Coulson et al., 1995). At the local scale, soil temperature varies horizontally, vertically, and temporally. Local horizontal variations in soil temperature arise from differences in plant cover and surface shading (Figure 9.1), variability in soil moisture and density, or distance from buried heat sources. Horizontal gradients in soil surface temperature under a partially developed row crop canopy (or partial mulch cover) can exceed 1°C cm^{-1} (Horton et al., 1984a).

Vertical soil temperature variations are primarily driven by the fluctuating temperature of the soil surface. Heat is both absorbed at and lost from the surface of the soil. The soil transmits heat
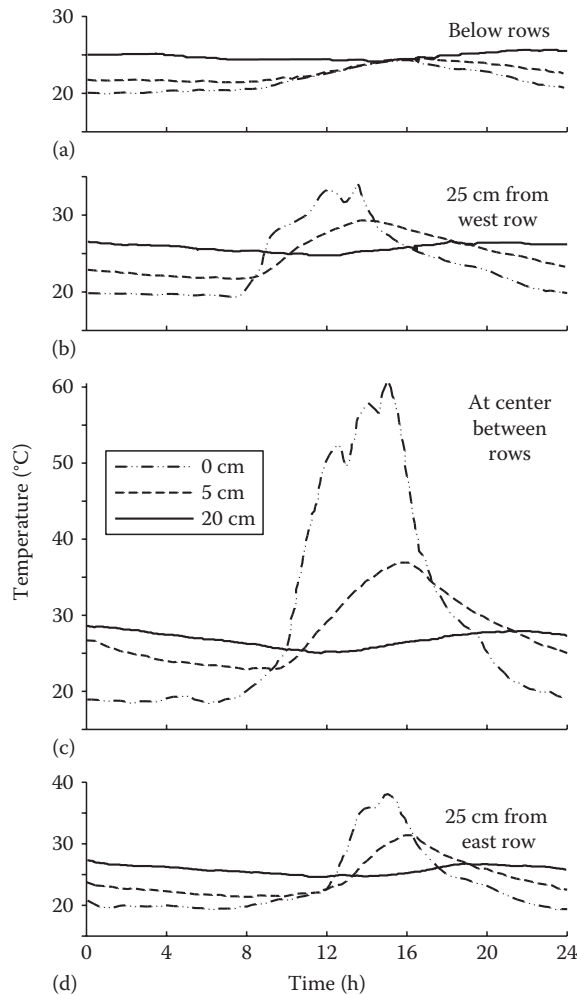
FIGURE 9.1 Soil temperature with time at 0, 5, and 20 cm below the soil surface as measured between two NE–SW-oriented rows of 60 cm tall chili (*Capsicum annuum* L.) plants. The rows were 100 cm apart, (a), (b), (c), and (d) show diurnal soil temperatures measured at locations beneath rows, 25 cm from the west row, at the center between rows, and 25 cm from the east row, respectively. (Reprinted from Horton, R., O. Aguirre-Luna, and P.J. Wierenga. 1984a. Observed and predicted two-dimensional soil temperature distributions under a row crop. Soil Sci. Soc. Am. J. 48:1147–1152.)

downward when the temperature near the surface is higher than the temperature below, and it transmits heat upward when the temperature is warmer within the soil than at the surface. During warming periods, the soil temperature is greatest near the surface and lower at depth. During cooling periods, the reverse is true. Thus, for most environments, both the highest and lowest observed soil temperatures occur at the surface. Under conditions of dry soil and high incoming radiation, vertical soil thermal gradients near the surface can exceed 10°C cm^{-1} (Heitman et al., 2008c).

Soil temperature also changes with time. In many, but not all, cases, soil temperatures exhibit distinct daily and annual cycles. The cycles deeper in the soil lag behind those near the surface. The daily cycles decrease in amplitude as depth increases, and below 35 cm, amplitudes become quite small in most soils (Figure 9.2). Annual cycles are observable to greater depths than

daily cycles, if seasonal air temperature differences occur. Thus, annual soil temperature cycles are much less pronounced near the equator than in the middle and high latitudes. Seasonal temperature differences decrease, and the annual cycles lag progressively as depth increases. The amplitude of the annual temperature cycle at depths >8 m is 5% or less of the amplitude at the soil surface, and the temperature at these depths approximates the mean annual temperature of the shallower soil (Baker and Baker, 2002; Smerdon et al., 2003).

For soils that freeze in winter, soil temperature is influenced by the release of heat when liquid water becomes ice. The amount of heat released is about 334 J g^{-1} of water that freezes. The heat of fusion energy must be dissipated while soil water freezes. The rate of thaw of frozen soils is relatively slow, because heat energy is required to both warm the soil and melt the ice. In areas of heavy snowfall, the snow provides an insulating blanket and soils do not freeze as deeply or may not freeze at all.

The soil thermal regime is one of the defining characteristics of a soil. The USDA soil classification system (Soil Survey Staff, 1999) includes soil temperature regimes based on temperature at a depth of 50 cm or at the upper boundary of a root-limiting layer, whichever is shallower. The soil temperature regimes, defined in terms of the mean annual soil temperature and the difference between mean summer and mean winter temperatures, are determined by the following groupings:

Gelisols and gelic suborders and great groups are assigned to soil temperature regimes based on their mean annual soil temperature as follows: *hypergelic* (–10°C or lower), *pergelic* (–4°C to –10°C), and *subgelic* (1°C to –4°C).

Other soils that have a difference in soil temperature of 6°C or more between mean summer and mean winter temperatures are assigned to soil temperature regimes based on their mean annual soil temperature as follows: *frigid* (lower than 8°C), *mesic* (8°C–15°C), *thermic* (15°C–22°C), and *hyperthermic* (22°C or higher).

All other soils are assigned to soil temperature regimes based on their mean annual soil temperature as follows: *isofrigid* (lower than 8°C), *isomesic* (8°C–15°C), *isothermic* (15°C–22°C), and *isohyperthermic* (22°C or higher).

9.2 Mechanisms of Heat Transfer in Soil

Three principal mechanisms, radiation, convection, and conduction, are responsible for the transfer of heat in soil simultaneously. Radiative energy transfer includes incoming direct and diffuse solar (shortwave) radiation, as well as longwave sky radiation to the soil surface and longwave radiation emitted outward from the soil surface. Radiative transfer is a significant component of heat transfer at the soil surface, but its significance decreases below the soil surface. Convective heat transfer in soil is associated with a net flux of fluids. Convection may be responsible for a major portion of the soil heat transfer during periods of large water flux (e.g., during rainfall or irrigation). Convection is also important via vapor fluxes in shallow unsaturated soil layers when large thermal gradients occur. Although radiation and convection are each important mechanisms of soil heat transfer
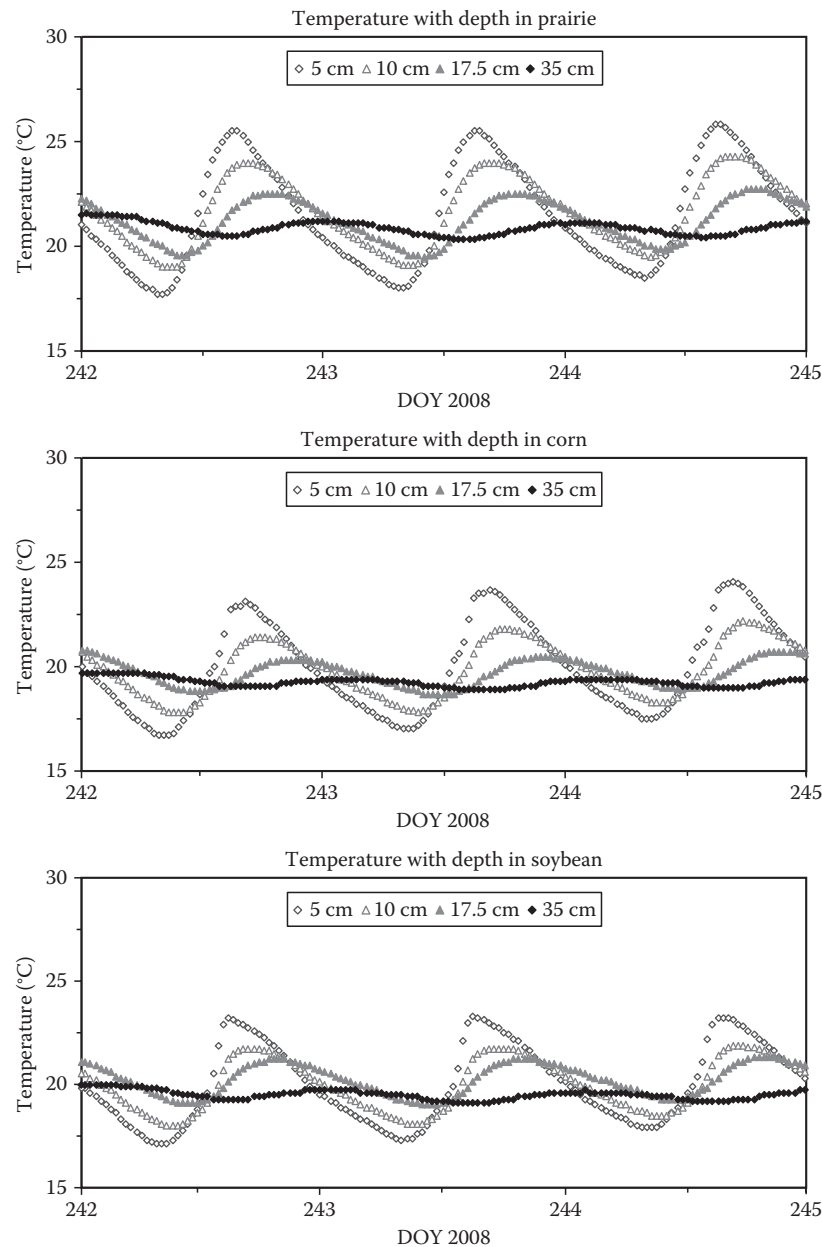
FIGURE 9.2   Soil temperature variation with depth and time for several Iowa fields.

under certain conditions, conduction is typically the dominant mechanism of heat transfer within the soil. Conduction heat transfer involves the transfer of heat at a molecular scale from positions of large kinetic energy (high temperature) to positions of small kinetic energy (low temperature).

We will consider that radiation heat transfer is restricted to the surface. That leaves convection and conduction as heat transfer mechanisms to be considered within soil. Most authors focus on conduction heat transfer as the primary mechanism of heat transfer in soil. They assume that conduction predominates at times other than rainfall or irrigation. Buchan (2001) is somewhat unique in his presentation of soil heat transfer processes in that he includes conduction and convection heat transfer

simultaneously. This is good because in soil, conduction and convection often occur together.

Transfer of heat and water in soil is generally coupled, which is to say that the transfer processes are intertwined. Water moving into and through the soil carries heat with it, and temperature gradients in the soil cause movement of both liquid water and water vapor. In a typical diurnal cycle, the soil surface loses heat during the night. The surface temperature drops, and the resulting upward thermal gradient results in transfer of heat to the surface from the relatively warm soil below. If the surface temperature drops to the dew point temperature, water vapor condenses at the cooled soil surface, liberating more heat. Throughout the night, liquid water flows upward, rewetting

surface layers that had dried during the previous day and redistributing heat upward. In the morning, solar radiation warms the soil surface. As the surface temperature increases, the thermal gradient shifts direction. Liquid water near the soil surface vaporizes and is transported into the atmosphere, taking large quantities of heat with it. Liquid water and water vapor are also driven by the thermal gradient from the hot surface to the cooler soil below. As the sun goes down, the dynamic and intricate cycle begins again.

9.3 Surface Energy Balance

The complexities and significance of the soil thermal regime find full expression in the deceptively simple and widely used equation for the energy balance of the land surface:

$$R_n + H + LE + G = 0, \qquad (9.1)$$

where

R_n is net radiation at the surface
H is sensible heat flux between the surface and the atmosphere
LE is latent heat flux between the surface and the atmosphere
G is the total heat flux (sensible and latent) between the surface and the subsurface

Here, all the fluxes are positive toward the surface and expressed in units of watt per square meter.

Comprehensive surface energy balance studies have been conducted since the 1950s. Works by Lettau and Davidson (1957) and Lemon (1963) are early examples of these types of studies. In recent years, with technical advancement of ground-based and remote sensing instrumentation, surface energy balance measurements have become much more common. This trend is evidenced by the development of the global energy balance archive (GEBA) for documentation of current climatic conditions and facilitation of the study of past and future climate (Gilgen and Ohmura, 1999). The spatial scale of energy balance studies has also expanded with advancing sensor technology. For example, interest in global climate change has prompted several efforts to estimate the earth's annual mean energy budget (e.g., Ohmura and Gilgen, 1993; Kiehl and Trenberth, 1997).

Some authors have defined H, LE, and G as positive away from the surface, in which case the "plus" signs are replaced with "minus" signs in Equation 9.1. With that convention, Equation 9.1 is often rearranged to set R_n equal to the sum of the other three terms. Such presentation promotes the idea that net radiation is partitioned into heating the air, heating the soil, and evaporating water. This thinking is only partially true. By isolating the net radiation term, proper importance is assigned to the radiative energy that derives from the sun. However, the net radiation includes more than solar energy. The implication that R_n is the driver and that the other processes are the followers is not completely true. A more accurate representation is given in Equation 9.1. The four components of the energy balance are interdependent. The surface temperature affects all four terms, including R_n. The surface temperature, in turn, is a result of the energy partitioning that includes all four terms of Equation 9.1. In the following sections, the four terms of Equation 9.1 are briefly considered for the case of a bare soil surface. We refer readers to Chapter 6 for a more comprehensive presentation of surface energy balance.

9.3.1 Net Radiation

Net radiation (R_n) is defined as incoming minus outgoing radiation at the land surface. Net radiation can be described by the following:

$$R_n = (1 - a)R_g + R_l - \varepsilon\sigma T_s^4, \qquad (9.2)$$

where

a is surface albedo
R_g is global (i.e., direct plus diffuse solar) radiation (W m^{-2})
R_l is downward longwave sky irradiance (W m^{-2})
ε is surface emissivity
σ is the Stefan–Boltzmann constant (5.67 × 10^{-8} W m^{-2} K^{-4})
T_s is surface temperature (K)

The soil surface temperature can be measured with thermocouples or by infrared thermometers (Ham and Senock, 1992).

9.3.2 Sensible and Latent Atmospheric Heat Fluxes

Horton and Chung (1991) used the following equations to calculate surface latent and sensible heat fluxes:

$$LE = \frac{L(\rho_{va} - \rho_{vs})}{r_{va}}, \qquad (9.3)$$

$$L = 2.49463 \times 10^9 - 2.247 \times 10^6 \, (T_s - 273.15), \qquad (9.4)$$

$$H = \frac{C_a(T_a - T_s)}{r_A}, \qquad (9.5)$$

where

E is the evaporative flux (m s^{-1})
L is the latent heat of vaporization (J kg^{-1})
ρ_{va} is vapor density of the air (kg m^{-3})
ρ_{vs} is vapor density of the surface (kg m^{-3})
r_{va} and r_A are the aerodynamic boundary layer resistances (s m^{-1}) to vapor and heat transfer
C_a is air volumetric heat capacity (J m^{-3} °C^{-1})
T_a is air temperature

Some studies suggest that an additional soil surface resistance parameter be added to the numerator in Equation 9.3 to more accurately represent evaporation when the soil surface is dry (Saito et al., 2006).

The vapor density at the surface, ρ_{vs}, can be calculated from the Kelvin equation as

$$\rho_{vs} = \rho_{vs}^{\star} \exp\left[\frac{(M_w \psi)}{(RT_s)}\right], \qquad (9.6)$$

where

$\rho_{vs}^{\star}$ is the saturated vapor density (kg m^{-3})

M_w is the molecular weight of water (kg mol^{-1})

ψ is the water potential (sum of osmotic and matric) at the surface (J kg^{-1})

R is the universal gas constant (8.314 J mol^{-1} K^{-1})

Note that Equations 9.2, 9.4 through 9.6 show that R_n, LE, and H are each influenced by surface temperature. Therefore, the terms cannot be considered to be independent of each other. Several approaches for solving the surface energy balance equations for T_s have been developed (e.g., Bristow, 1987; Horton and Chung, 1991).

9.3.3 Sensible and Latent Soil Heat Flux

The final term in Equation 9.1 is G, the soil heat flux. The amount of sensible and latent energy that moves through an area of soil in a unit of time is the soil heat flux or heat flux density, G. The ability of a soil to conduct heat impacts how fast its temperature changes during a day or between seasons. Soil heat flux is important because it effectively couples energy transfer processes at the surface (surface energy balance) with energy transfer processes in the soil (soil thermal regime). This interaction between surface and subsurface energy transfer processes has led to detailed investigations of soil heat flux for a wide variety of agricultural systems (Sauer, 2002; Sauer and Horton, 2005).

The magnitude of G as a component of the surface energy balance varies with surface cover, soil moisture content, and solar irradiance (Sauer and Horton, 2005). Daytime peak hourly values of G for a bare, dry soil in midsummer, could be in excess of 300 W m^{-2} (Fuchs and Hadas, 1972). By contrast, hourly G for a moist soil beneath a plant canopy, residue layer, or snow cover will often be less than ±20 W m^{-2}. Surface soil heat flux typically represents 1%–10% of daily R_n for growing crops with full canopy coverage (Denmead, 1969; Szeicz et al., 1973; Brown, 1976; Uchijima, 1976; Baldocchi et al., 1985; Clothier et al., 1986). This percentage can exceed 50% in the fall and spring when R_n is low and the soil is cooling/warming or in arid climates when there is no vegetation (Monteith, 1958; Idso et al., 1975; Choudhury et al., 1987).

Figure 9.3 provides example curves of the energy balance terms for a corn (*Zea mays* L.) residue-covered soil in central Iowa (Sauer et al., 1998). The data in Figure 9.3 were obtained in November when the soil and fresh residue layers were dry and the mean surface temperature was 6.8°C. Daytime G averaged 14.4% and 16.8% of R_n for days 310 and 311, respectively. Much of the energy that enters the soil during the day returns to the atmosphere at night through terrestrial longwave radiation. For this reason, G is often the smallest component of the daily
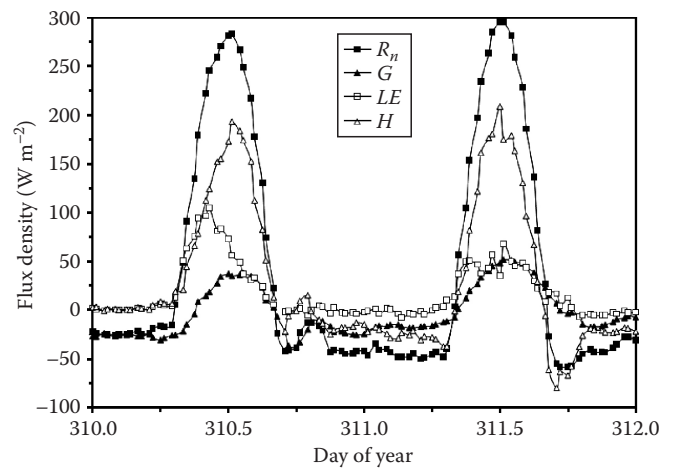
FIGURE 9.3 Energy balance components measured above a corn residue-covered soil surface in 1994 at a site near Ames, Iowa. Net radiation (R_n) is positive toward the surface. The other terms are positive away from the soil surface. (Reprinted from Sauer, T.J., J.L. Hatfield, J.H. Prueger, and J.M. Norman. 1998. Surface energy balance of a corn residue-covered field. Agric. For. Meteorol. 89:155–168, Copyright (1998), with permission from Elsevier.)

surface energy balance and has, in some cases, been ignored. However, there are often significant transfers of energy into and out of a soil during both day- and nighttime hours, and failure to include G in short-term (i.e., hourly) energy balance determinations can lead to significant errors. Later in this chapter, we consider methods to estimate and measure G.

9.4 Soil Thermal Properties

The three main soil thermal properties are volumetric heat capacity (C), thermal conductivity (λ), and thermal diffusivity (α). Because $\alpha \equiv \lambda C^{-1}$, knowledge of any two of the three properties is sufficient information to calculate the third property. Soil thermal properties are generally expressed as a function of soil water content. For example, Ren et al. (1999) reported soil thermal property values for two soils as a function of water content (Figure 9.4). Ochsner et al. (2001) studied soil thermal properties as a function of soil solid, liquid, and gas fractions. For a range of soils, they found that soil air content exerted the most dominant influence on soil thermal properties (Figure 9.5). The effective or apparent thermal properties of soil at temperatures below freezing can also be strongly temperature dependent due to latent heat effects associated with the liquid–solid phase change of soil water (Figure 9.6; Ochsner and Baker, 2008).

9.4.1 Soil Thermal Conductivity

Heat flow in soil has often been considered analogous to heat flow in a solid to which Fourier's law is applied:

$$G = -\lambda \frac{dT}{dz} \qquad (9.7)$$
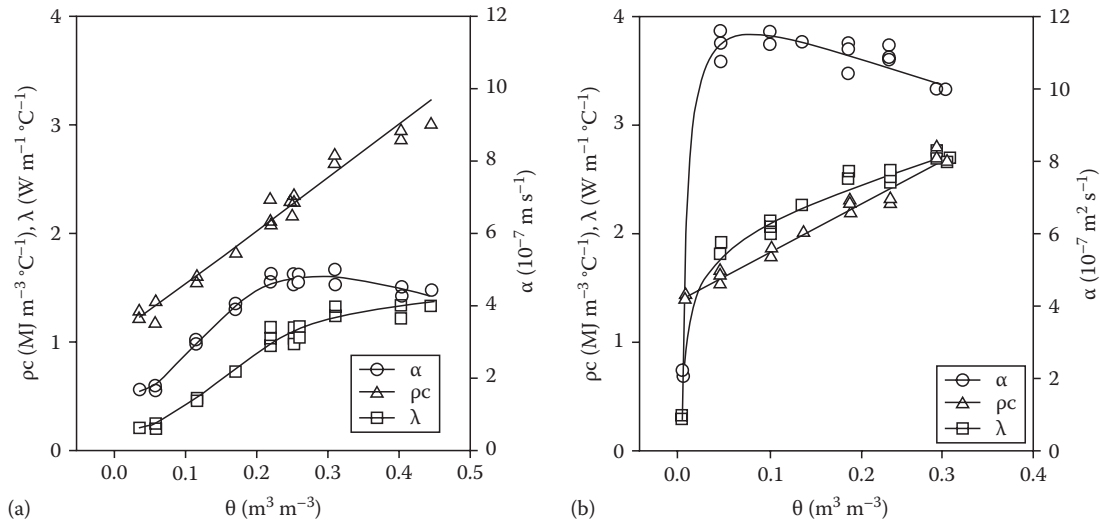
FIGURE 9.4 Thermal properties of clay loam soil (a) and silica sand (b) as functions of volumetric water content. (Reprinted from Ren, T., K. Noborio, and R. Horton. 1999. Measuring soil water content, electrical conductivity, and thermal properties with a thermo-TDR probe. Soil Sci. Soc. Am. J. 63:450–457. With permission. Soil Science Society of America.)
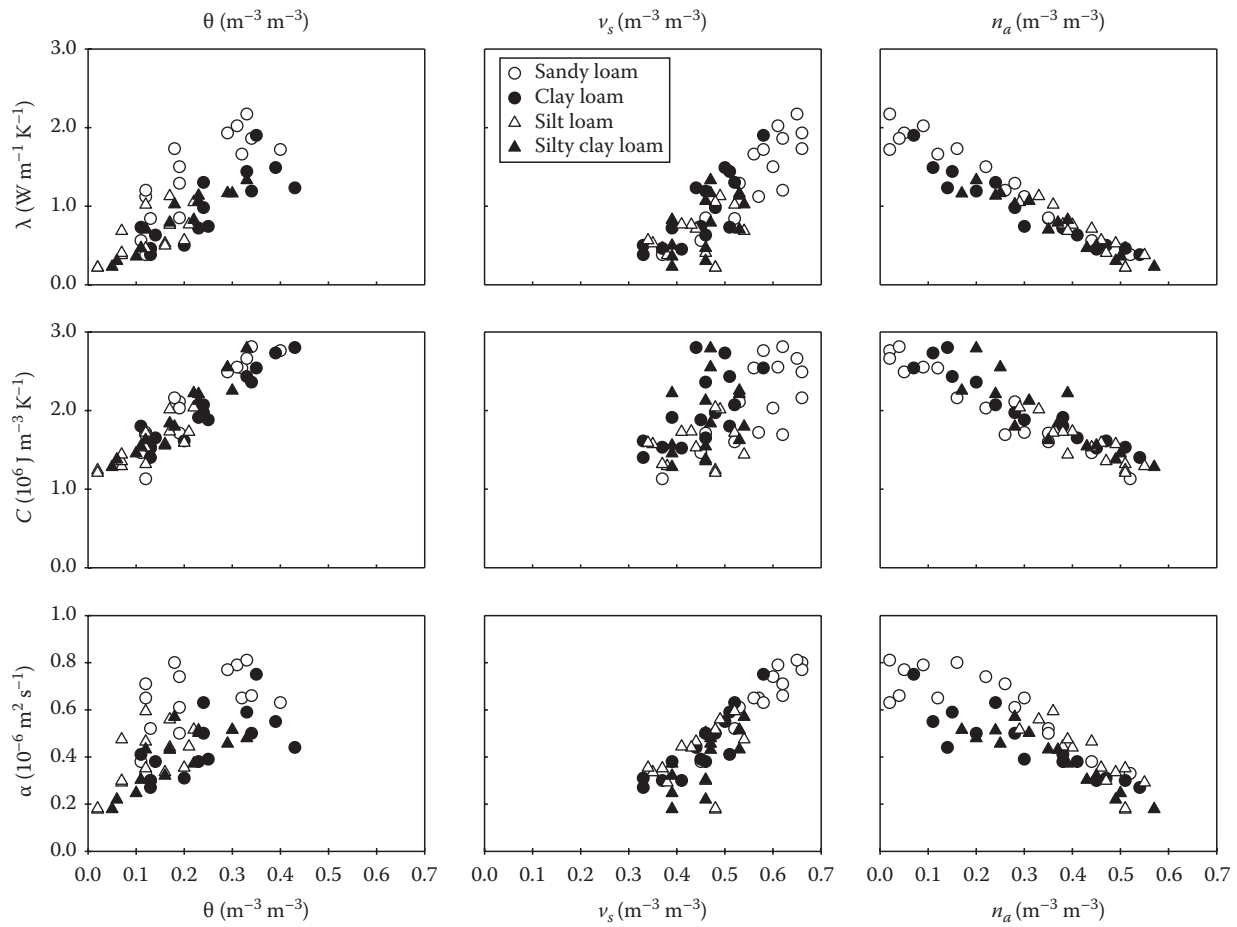
FIGURE 9.5 Thermal conductivity, heat capacity, and thermal diffusivity versus volume fractions of water (θ), solids (v_s), and air (n_a) for four medium-textured soils. (Reprinted from Ochsner, T.E., R. Horton, and T. Ren. 2001. A new perspective on soil thermal properties. Soil Sci. Soc. Am. J. 65:1641–1647. With permission. Soil Science Society of America.)
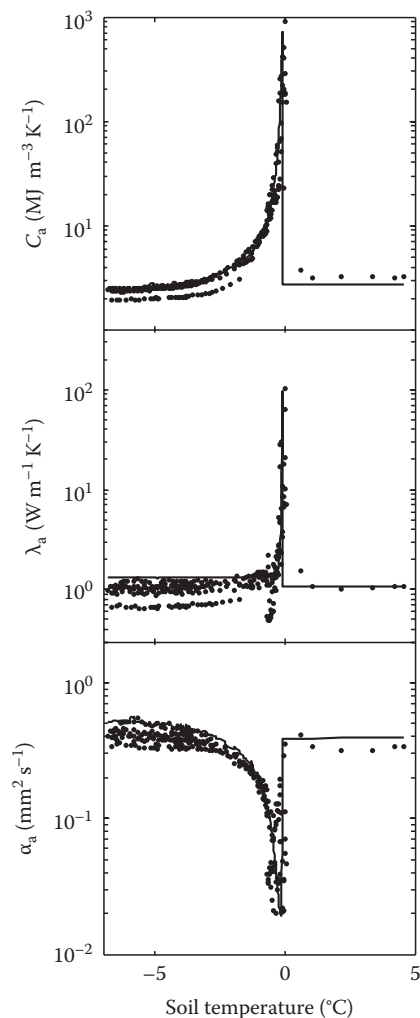
FIGURE 9.6 Apparent thermal properties of silt loam soil for temperatures just above and below 0°C. Thermal properties were measured in situ (symbols) using an HP method. Thermal properties were modeled (lines) using the approach of Fuchs et al. (1978). (Reprinted from Ochsner, T.E., and J.M. Baker. 2008. In situ monitoring of soil thermal properties and heat flux during freezing and thawing. Soil Sci. Soc. Am. J. 72:1025–1032. With permission. Soil Science Society of America.)

where

 G is the heat flux density (W m^{-2})
 λ is thermal conductivity (W m^{-1} K^{-1})
 d*T*/d*z* is the temperature gradient (K m^{-1})

Thermal conductivity is a measure of the soil's ability to transmit heat. It is the ratio of the magnitude of the heat flux through the soil to the magnitude of the temperature gradient. Fourier's law is easily and directly utilized in many engineering applications; however, in a porous, three-phase media like soil, specifying λ is considerably more difficult.

 Unsaturated soils consist of three phases: solid, liquid, and gas. The different materials within soils have different thermal conductivities. Both fractional makeup and physical arrangement of

TABLE 9.1 Density, Specific Heat and Thermal Conductivity of Common Soil Constituents at 10°C

Soil Constituent	Density (ρ) (Mg m^{-3})	Specific Heat (*c*) (kJ kg^{-1} K^{-1})	Thermal Conductivity (λ) (W m^{-1} K^{-1})
Quartz	2.66	0.75	8.8
Clay minerals	2.65	0.76	3
Soil organic matter	1.3	1.9	0.3
Water	1.00	4.18	0.57
Ice (0°C)	0.92	2.0	2.2
Air	0.00125	1.0	0.025

 Source: After de Vries, D.A. 1963. Thermal properties of soils, p. 210–235. *In* W.R. van Wijk (ed.) Physics of plant environment. North-Holland Publishing Co., Amsterdam, the Netherlands, Table 7.1.

materials affect the overall thermal conductivity of soils. Two of the major factors affecting thermal conductivity of porous materials are water content (θ) and bulk density (ρ_b). In general, λ of solid > λ of liquid > λ of gas. Therefore, increasing the bulk density or water content of a soil should lead to increases in λ. Al Nakshabandi and Kohnke (1965), for instance, measured an eightfold increase in the thermal conductivity of a silt loam soil as its water content increased. The thermal conductivity of soil also varies by composition of the solid fraction (e.g., mineral type, particle size, and amount of organic matter) (Al Nakshabandi and Kohnke, 1965; Abu-Hamdeh and Reeder, 2000). In particular, the thermal conductivity increases with quartz content because quartz has a higher thermal conductivity than other common soil minerals (Bristow, 1998). de Vries (1963) presented average values of thermal conductivity for various soil constituents and those values are reproduced here in Table 9.1.

 Temperature is another parameter that influences thermal conductivity. de Vries (1963) found that thermal conductivity of a sandy soil increased with increasing water content and with temperature until 60°C. At a temperature of 75°C, the thermal conductivity increased with increasing water content until 0.1 m^3 m^{-3}, but then decreased as water content increased. In unsaturated soil, conduction is not the only heat transfer mechanism influencing λ. Evaporation, movement, and condensation of water vapor in the soil pores also influence λ. These latent heat transfer processes become more important as soil temperature increases, but they can be inhibited when liquid water content fills the majority of the pore space.

 The properties that influence λ vary between soils, spatially at the soil surface for the same soil, between layers within a soil, and over time. Because of this variability, much effort has been made to develop λ models based on easily measurable soil parameters (Kersten, 1949; de Vries, 1963; Johansen, 1975; Campbell; 1985; Côté and Konrad, 2005). The available models can be divided into two groups: semi-theoretical models and empirical models.

 The de Vries (1963) model is a physically based, semi-theoretical model, which has been applied in numerous studies, including predicting the response of λ to temperature change (Hopmans

and Dane, 1986; Campbell et al., 1994). In the de Vries model, the thermal conductivity of soil is calculated as the weighted average of the conductivities of the various soil constituents according to the formula:

$$\lambda = \frac{\sum_{i=0}^{n} k_i \lambda_i x_i}{\sum_{i=0}^{n} k_i x_i},$$ (9.8)

where

x_i is the volume fraction of each constituent
λ_i is the thermal conductivity of each constituent
n is the number of soil constituents

The weighting factors, k_i, depend on the shape and orientation of the granules of the soil constituents and on the ratio of the conductivities of the constituents. The subscript 0 refers to the continuous fluid surrounding the solid particles (i.e., air for dry soil and water for moist soil) with $k_0 = 1$. Other values of k_i are calculated from

$$k_i = \frac{1}{3} \sum_{j=1}^{3} \left[1 + \left(\frac{\lambda_i}{\lambda_0} - 1 \right) g_j \right]^{-1},$$ (9.9)

where g_j represents the shape factors for the ith constituent, and $g_1 + g_2 + g_3 = 1$. Assuming g_1 and g_2 are equal, only one shape factor needs to be estimated for each constituent.

The thermal conductivity of the air-filled pores is considered to be the sum of λ_a and λ_v, where λ_a is the thermal conductivity of dry air (0.025 W m^{-1} K^{-1} at 20°C), and λ_v accounts for heat transfer across the air-filled pores by water vapor. Above some critical water content, the air-filled pores are assumed to be saturated with water vapor, and λ_v is assumed to be 0.074 W m^{-1} K^{-1} at 20°C. Below the critical water content, λ_v is assumed to decrease linearly with water content to a value of zero for oven-dry soil. Assuming a soil solid particle shape factor of 0.144, then for water contents above the critical water content, the value of g_1 for the air-filled pores is given by

$$g_1 = 0.333 - \frac{n_a}{(1 - v_s)}(0.333 - 0.035),$$ (9.10)

where n_a and v_s are the volume fractions of air and solids, respectively.

For water contents below the critical water content, the value of g_1 for the air-filled pores is given by

$$g_1 = 0.013 + \frac{\theta}{\theta_c}(g_{1c} - 0.013),$$ (9.11)

where

θ_c is the critical water content
g_{1c} is the value of Equation 9.10 at the critical water content

The de Vries model has proven useful for many years, but proper (and somewhat subjective) selection of model parameters, such as critical water content and shape factors, is necessary to predict λ accurately (Horton and Wierenga, 1984; Ochsner et al., 2001).

There is a great diversity of empirical models for λ. One of the earliest is that of Kersten (1949). This model only requires bulk density as an input parameter, but the model is not suitable for predicting λ at lower water contents (Côté and Konrad, 2005). Campbell (1985) introduced an empirical function to calculate λ. The empirical function has five parameters, some of which are difficult to estimate. Johansen (1975) proposed the concept of normalized thermal conductivity and established a simple empirical model based on the degree of saturation and soil mineral composition. For many soils, the Johansen (1975) model provides accurate predictions of λ (Farouki, 1981, 1982; Tarnawski and Wagner, 1992). Nevertheless, the Johansen (1975) model does not cover the entire water content range. Lu et al. (2007) presented an improved model describing the relationship between thermal conductivity and volumetric water content. The new model was calibrated using measured thermal conductivity from eight soils, and it was validated by comparing predicted λ with measured λ on 19 additional soils. The model performed well for a wide range of soils and was applicable over the entire range of soil water contents. Frozen soil represents a special condition. Farouki (1986) provides insight on observations and predictions of thermal properties for frozen soil conditions.

9.4.2 Soil Heat Capacity

Soil heat capacity characterizes the quantity of heat added to or removed from soil per unit change in temperature. Heat capacity has an important role in determining the magnitude of diurnal and annual soil temperature variations. In soil with low heat capacity, the addition or removal of heat causes a relatively large soil temperature change, and in soil with large heat capacity, the addition or removal of heat causes a relatively small temperature change.

The volumetric heat capacity of soil is calculated as the weighted sum of the heat capacities of the various soil constituents (Kluitenberg, 2002). Thus,

$$C = \sum_{i=1}^{n} x_i C_i = \sum_{i=1}^{n} x_i \rho_i c_i,$$ (9.12)

where x_i and C_i (J m^{-3} K^{-1}) represent the volume fractions and volumetric heat capacities, respectively, of n soil constituents. This relationship shows that the volumetric heat capacity of each constituent is the product of its density ρ_i (Mg m^{-3}) and its specific heat c_i (J kg^{-1} K^{-1}). Equation 9.12 can be expressed in terms of mass fractions ϕ_i by using the relationship:

$$x_i = \frac{\phi_i \rho_b}{\rho_i},$$ (9.13)

where ρ_b is the soil bulk density. Substituting Equation 9.12 into Equation 9.13 yields

$$C = \rho_b \sum_{i=1}^{n} \phi_i c_i. \tag{9.14}$$

In order to use Equation 9.14, knowledge is needed of the specific heats of the various constituents present in soil. Approximate specific heat values for soil constituents are provided in Table 9.1. Because soil air makes a negligible contribution to C, it is common to write Equation 9.14 as

$$C = \rho_b(c_m \phi_m + c_o \phi_o + c_w \theta_g), \tag{9.15}$$

where
the subscripts m, o, and w indicate mineral material, organic material, and water, respectively
θ_g is the gravimetric water content

The mineral and organic fractions of soil can be lumped to give an average specific heat of the solid constituents, c_s,

$$C = \rho_b(c_s + c_w \theta_g). \tag{9.16}$$

9.4.3 Soil Thermal Diffusivity

Soil thermal diffusivity, α, is defined as the ratio of thermal conductivity to volumetric heat capacity. Thermal diffusivity often appears in partial differential equations describing soil heat transfer and in the solutions to those equations. Soil thermal diffusivity is the parameter that describes the rate of transmission of a temperature change through the soil. The higher the thermal diffusivity, the more rapid the propagation of temperature changes through the soil. Thermal diffusivity can be estimated by using appropriate models to calculate thermal conductivity and heat capacity and then taking the ratio. Procedures for measuring thermal properties, including diffusivity, are discussed below.

9.4.4 Measuring Soil Thermal Properties

Considerable effort has gone into developing methods for measuring soil thermal properties. The line-source thermal sensor (with single needle or multiple needles) has been the primary method. When using a single-needle sensor, temperature with time data are obtained during and just after the needle is heated. The thermal conductivity can be calculated from the temperature data (Jackson and Taylor, 1965; Horton and Wierenga, 1984; Shiozawa and Campbell, 1990).

de Vries and Peck (1958) reported that it may be possible to determine λ and α simultaneously with a single-needle sensor. The unknown contact resistance between the line-source sensor

and soil is a practical inhibitor to accurate results. Bruijn et al. (1983) and van Haneghem et al. (1983) presented and tested the theory for simultaneously determining λ and C by accounting for contact resistance between sensor and soil. Other investigators have recognized that measurement of temperature (multineedle sensors) at a known distance away from a heat source provides opportunity for simultaneous determination of thermal properties (Lubimova et al., 1961; Chudnovski, 1962; Jaeger, 1965; Nix et al., 1967, 1969; Larson, 1988; Kluitenberg et al., 1993, 1995; Bristow et al., 1994; Bilskie et al., 1998).

Bristow et al. (1994) showed that using short-duration heat-pulse (HP) theory, λ, C, and α could be determined from a single HP measurement using a dual-probe sensor as described by Campbell et al. (1991). This method is commonly referred to as the dual-probe HP method or simply the HP method. Figure 9.7 shows diagrams of a typical dual-probe HP sensor. The HP method is based on the theory of radial heat conduction of a short-duration heat pulse from an infinite line source. In an infinite medium, the temperature change as a function of time at a radial distance from the HP source is given by (de Vries, 1952; Kluitenberg et al., 1993):

$$\Delta T(r, t) = \frac{Q}{4\pi\alpha}\left[Ei\left(\frac{-r^2}{4\alpha(t - t_0)}\right) - Ei\left(\frac{-r^2}{4\alpha t}\right)\right], \quad t > t_0 \tag{9.17}$$

where
ΔT is the temperature change (°C)
t is time (s)
t_0 is the HP duration (s)
r is the radial distance (m)
$-Ei(-x)$ is the exponential integral

The source strength is defined as $Q = q'/C$, where q' is the rate of heat liberation per unit length of probe (W m^{-1}). Figure 9.8 shows a typical temperature rise curve obtained from an HP measurement.

Based upon Equation 9.17, soil thermal properties can be expressed analytically as (Kluitenberg et al., 1993; Bristow et al., 1994):

$$\alpha = \frac{(r^2/4)[(1/(t_m - t_0)) - (1/t_m)]}{\ln[t_m/(t_m - t_0)]}, \tag{9.18}$$

$$C = \frac{q'}{4\pi\alpha\Delta T_m}\left[Ei\left(\frac{-r^2}{4\alpha(t_m - t_0)}\right) - Ei\left(\frac{-r^2}{4\alpha t_m}\right)\right], \tag{9.19}$$

$$\lambda = \alpha C, \tag{9.20}$$

where
t_m is the time (s) at which the temperature maximum occurred
ΔT_m is the maximum temperature change (°C)
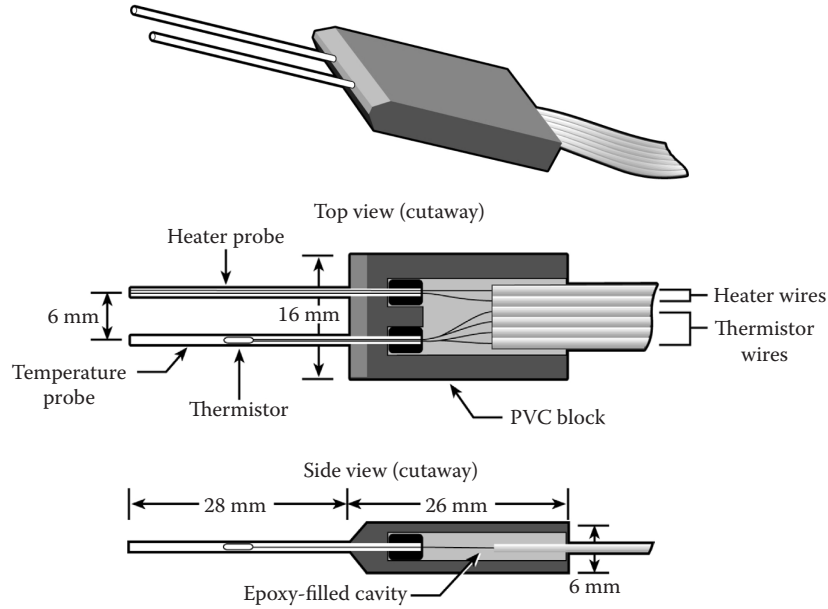
FIGURE 9.7 Diagram of a dual-probe HP sensor. (Reprinted from Kluitenberg, G.J. 2002. Heat capacity and specific heat, p. 1201–1208. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI. With permission. Soil Science Society of America.)
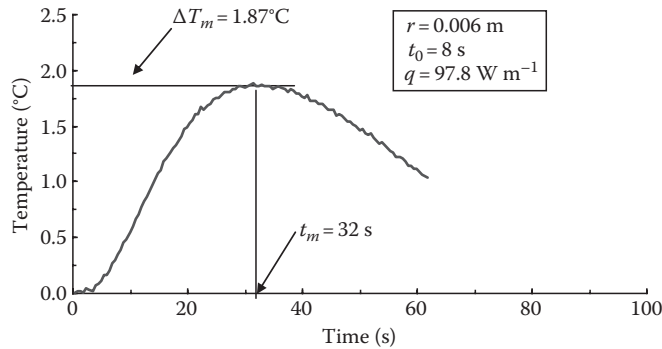
FIGURE 9.8 Typical temperature versus time curve for a dual-probe HP measurement. The maximum temperature rise (ΔT_m) and time of the maximum temperature rise (t_m) are indicated.

A more recent simplification of Equation 9.19 was developed by Knight and Kluitenberg (2004):

$$C = \frac{q' t_0}{e \pi r^2 \Delta T_m} \left(1 - \frac{\varepsilon^2}{8} \left\{ \frac{1}{3} + \varepsilon \left[\frac{1}{3} + \frac{\varepsilon}{8} \left(\frac{5}{2} + \frac{7\varepsilon}{3} \right) \right] \right\} \right), \quad (9.21)$$

where

$\varepsilon = t_0 / t_m$

e is the base of the natural logarithm

Equation 9.21 employs a telescoped polynomial rather than the exponential integral and is thus easier to use in spreadsheets and on-board dataloggers. Further details regarding the application of the HP method are given by Kluitenberg (2002) and Bristow (2002). One important application of the HP method is in monitoring soil water content (Tarara and Ham, 1997; Campbell et al., 2002;

Heitman et al., 2003; Ochsner et al., 2003). Soil water content can be estimated with the HP method by utilizing the linear dependence of C on the soil water content (Equation 9.16).

9.5 Modeling the Soil Thermal Regime

Heat transfer modeling offers the following contributions: (1) it describes energy partitioning at the soil surface; (2) it describes soil heat flux within the soil profile; and (3) it describes soil temperature distribution. The surface-energy partitioning and resulting soil temperature distribution directly and indirectly affect plant growth and development, and many biological, chemical, and physical processes in the soil.

9.5.1 Conduction Models

Analytical and numerical solutions to the following 1D, transient conduction equation for various initial and boundary conditions have been presented in the literature:

$$C \frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \left(\lambda \frac{\partial T}{\partial z} \right). \quad (9.22)$$

In many cases, heat transfer by radiation, convection, and conduction is modeled by the conduction equation alone (Wierenga and de Wit, 1970; Hanks et al., 1971; Gupta et al., 1981, 1983; Horton et al., 1984a, 1984b; Parton, 1984; Persaud and Chang, 1984). Apparent thermal properties rather than real thermal properties (Jackson and Kirkham, 1958) are assumed to account for both conductive and nonconductive heat flow. Several investigators have also modeled soil heat transfer in two dimensions (Takakura et al., 1971; Jury and Bellantuoni, 1976a, 1976b; Mahrer and Katan, 1981;

Mahrer, 1982; Horton et al., 1984a, 1984b; Chung and Horton, 1987; Horton, 1989; Benjamin et al., 1990; Kluitenberg and Horton, 1990; Novak, 1993). Two-dimensional models have been used to analyze soil heat transfer beneath row crops, in ridged surfaces, and beneath partial mulch covers. Some of the models include conduction heat transfer alone, and some consider soil water impacts on thermal properties and/or soil water movement (Wierenga and de Wit, 1970; van Bavel and Hillel, 1975, 1976; Sophocleous, 1979; Milly, 1982; Campbell, 1985; Ross et al., 1985; Bristow et al., 1986; Lascano et al., 1987; Pruess and Wang, 1987; Bear and Bensabat, 1989; Nassar and Horton, 1989b, 1989c, 1992; Horton and Chung, 1991; Hares and Novak, 1992a, 1992b; Nassar et al., 1992a, 1992b; Olivella et al., 1994; Scanlan and Milly, 1994; Bear and Gilman, 1995; Gawin et al., 1995; Bristow and Horton, 1996; Noborio et al., 1996a, 1996b; Saito et al., 2006).

For homogeneous, isotropic conditions, the transient conduction equation is as follows:

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial z^2}. \tag{9.23}$$

Analytical solutions to Equation 9.23 exist for certain initial and boundary conditions. Carslaw and Jaeger (1959) have compiled a number of solutions to Equation 9.23 for heat transfer in solids. Although soil does not consist of solids alone, with the assumptions of uniform, constant thermal properties and heat conduction processes dominating, conduction heat transfer theory for solids can be used to approximate heat transfer in soil.

van Wijk and de Vries (1963) present an analytical solution to Equation 9.23 for periodic surface temperature variation for a homogeneous soil. For initial conditions that result from infinite periodic surface temperature variation, semi-infinite space, and surface boundary as

$$T(0, t) = \bar{T} + A\sin(\omega t + \phi), \tag{9.24}$$

where

$\bar{T}$ is the average temperature
A is the amplitude of the surface temperature wave
ϕ is the phase constant
ω is the angular frequency (2π/period, e.g., 2π day^{-1} for diurnal variation)

and the analytical solution is

$$T(z, t) = \bar{T} + A\exp\left(-z\sqrt{\frac{\omega}{2\alpha}}\right)\sin\left(\omega t + \phi - z\sqrt{\frac{\omega}{2\alpha}}\right). \tag{9.25}$$

9.5.1.1 Apparent Thermal Diffusivity

Equation 9.25 provides a basis for estimating the apparent soil thermal diffusivity using measured soil temperature time series. Soil temperature measurements at two or more depths are required. Temperature observations from the shallowest depth are used to establish the upper boundary condition, and the lower depth(s) observations are used to determine α.

Wierenga et al. (1969) presented two methods of determining α based upon Equation 9.25. The first is called the amplitude method. Note that Equation 9.25 indicates that temperature amplitude should decrease exponentially with depth. Wierenga et al. (1969) used the following equation to calculate the apparent thermal diffusivity from the amplitude of the soil temperature wave at two depths:

$$\alpha = \frac{\omega}{2}\left[\frac{z_2 - z_1}{\ln(A_1/A_2)}\right]^2. \tag{9.26}$$

The second approach is the phase method. Note that Equation 9.25 indicates a phase shift as a function of depth. Wierenga et al. (1969) used the following equation to analyze the phase shift of their observations:

$$\alpha = \frac{1}{2\omega}\left[\frac{z_2 - z_1}{\delta t}\right]^2, \tag{9.27}$$

where δt is the time interval between occurrence of maximum soil temperatures at depths z_1 and z_2. Equations 9.26 and 9.27 have been used to give first-order approximations of α. These methods work well for data obtained below a depth of 10 cm (Wierenga et al., 1969). Near the soil surface, Equations 9.26 and 9.27 provide less accurate estimates of α because the soil water content is not uniform and the soil temperature fluctuations are not sinusoidal. Effective use of these equations requires soil temperature measurements at two or more depths at frequent times. The observation period should include two or more sunny days before the days for which soil temperature will be analyzed. The sunny days will allow the soil temperature initial condition to more closely match the theoretical requirements.

As an alternative to Equation 9.24, Horton et al. (1983) used the following general surface boundary condition:

$$T(0, t) = \bar{T} + \sum_{n=1}^{M} A_n \sin(n\omega t + \phi_n), \tag{9.28}$$

and the method of superposition to obtain the following general analytical solution to Equation 9.23:

$$T(z, t) = \bar{T} + \sum_{n=1}^{M} A_n \exp\left(-z\sqrt{\frac{n\omega}{2\alpha}}\right)\sin\left(n\omega t + \phi_n - z\sqrt{\frac{n\omega}{2\alpha}}\right),$$

$$\tag{9.29}$$

where

M is the number of harmonics
A_n and ϕ_n are the amplitude and phase constant, respectively, of the nth harmonic for the upper boundary temperature

The coefficients A_n, ϕ_n, and $\overline{T}$ can be determined by least squares fitting of Equation 9.28 to the observed upper boundary soil temperature data. Using these coefficients in Equation 9.29 leaves only α as an unknown value. The value of α can be determined by fitting Equation 9.29 to a set of soil temperature observations measured at a depth below the upper boundary depth. The value of α is selected to minimize the sum of squares of the differences between the measured and calculated temperatures.

It is important to note that when α is determined from inverse fitting of conductive heat transfer theory to observed temperature, it is not strictly correct to assume that the fitted $\alpha = \lambda/C$. In moist soil, conduction is not the only mechanism of heat transfer. All of the nonconductive heat transfer (sensible and latent convective heat transfer) influences the parameter α. It is more correct to call this α the apparent thermal diffusivity instead of the true or real thermal diffusivity (Jackson and Kirkham, 1958). The real thermal diffusivity involves conduction heat transfer alone, while the apparent thermal diffusivity includes convection heat transfer in addition to conduction heat transfer. The apparent thermal diffusivity is a lumped parameter developed for a practical reason, and while useful for predicting soil temperature, it oversimplifies the many mechanisms of heat transfer within porous media. We should not permit its use to become a hindrance to our basic understanding of the numerous mechanisms of heat and water transfer in soil.

Not all approaches assume that the apparent thermal diffusivity is constant with depth. Nassar and Horton (1989a, 1990) presented methods for determining apparent thermal diffusivity as a function of depth. Again, in these cases, the apparent thermal diffusivity lumps together conductive and convective heat transfer processes. Wiltshire (1982, 1983) and Massman (1993a) presented analytical solutions to describe temperature distributions in nonuniform soil.

9.5.1.2 Soil Heat Flux Models

Transient soil heat flux (G) has also been estimated by analytical solutions to Equation 9.23. The theory assumes that only conduction heat transfer is occurring, so all of the actual convection heat transfer must be accounted for by apparent thermal properties. Conduction heat transfer theory is utilized in a manner to approximate both conduction and convection heat transfer conditions. Horton and Wierenga (1983) presented the following equation for soil heat flux:

$$G(z,t) = \sum_{n=1}^{M} A_n C \sqrt{n\omega\alpha} \, \exp\!\left(-z\sqrt{\frac{n\omega}{2\alpha}} \right)$$
$$\times \, \sin\!\left(n\omega t + \phi_n + \pi/4 - z\sqrt{\frac{n\omega}{2\alpha}} \right). \tag{9.30}$$

Equation 9.30 reduces to the following for surface heat flux:

$$G(0, t) = \sum_{n=1}^{M} A_n C \sqrt{n\omega\alpha} \, \sin\!\left[n\omega t + \phi_n + \left(\frac{\pi}{4}\right) \right] \tag{9.31}$$

Horton and Wierenga (1983) used this analytical solution to estimate heat flux for a bare soil and grass-covered field. The analytical solution worked well for both sunny and cloudy days.

9.5.2 Conduction–Convection Models

A few efforts have been made to measure or describe convective heat transfer under field conditions. Convective energy transfer in soil is associated with a net flux of fluids. Some observations of thermal effects on soil water infiltration have been made (Jaynes, 1990). Increases in seepage or infiltration rate were observed in response to temperature increases (Constantz et al., 2003). Gao et al. (2003) presented an analytical model of transient conduction and steady convection heat transfer. Jaynes (1990) presented a numerical model of heat and water transfer in soil during ponded irrigation. Shao et al. (1998) developed an analytical solution describing coupled conduction and convection heat transfer during ponded infiltration. Shao et al. (1998) reported periodic diurnal fluctuations in ponded infiltration rate. This infiltration rate was correlated with the water temperature. As water temperature oscillated, the infiltration rate oscillated. During infiltration, convection heat transfer had significant effects on soil temperature. Convection is especially important for heat transfer deeper in the soil profile. At a depth of 0.6 m, diurnal temperature variation due to conduction was negligible; however, convection heat transfer caused noticeable diurnal variation in soil temperature. Wierenga et al. (1970) observed irrigation water temperature impacting soil temperature, and Ren et al. (2000), Wang et al. (2002), Hopmans et al. (2002), and Ochsner et al. (2005) have presented conduction–convection models to determine water velocity based upon upstream and downstream temperature measurements in response to an application of a heat pulse.

9.5.3 Coupled Heat and Water Transfer Models

Coupled transfer of heat and water occurs in soil. For example, near the surface of cropland soil, weather fluctuations often cause simultaneous heat and water flow to occur. Buried radioactive material, underground energy storage, and buried electric cables produce heat that causes temperature gradients, which cause water to move. Temperature gradients can cause water to move in the vapor phase from a hot site and to condense at a cold site. After vapor condensation, a total pressure head gradient can drive water from the cold site to the warm site. Example profiles showing the redistribution of soil moisture under imposed temperature gradients are plotted in Figure 9.9.

Coupling of soil heat and water transfer has long been recognized (e.g., Bouyoucos, 1915). Thus, many studies have attempted to address coupled soil heat and water transfer and provide theoretical description of the dominant transfer mechanisms (e.g., Philip and de Vries, 1957; Jury and Letey, 1979; Sophocleous, 1979; Milly and Eagleson, 1980; Milly, 1982; Nassar and Horton, 1997). Several studies have been performed to investigate interactions between heat and water (Philip and de Vries, 1957; Cary, 1966; Globus, 1983). Milly and Eagleson (1980) and Milly (1982) described simultaneous heat and water
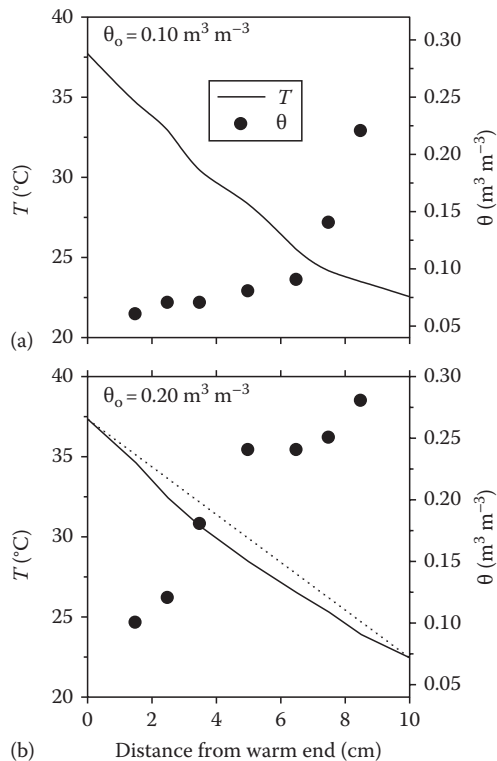
FIGURE 9.9 Comparison of paired temperature (T) and water content (θ) distributions for silt loam. Plots (a) and (b) show conditions after 96 h for 150°C m^{-1} temperature gradient at 30°C mean temperature. The dotted line in (b) represents a linear temperature distribution. (Reprinted from Heitman, J.L., R. Horton, T. Ren, and T.E. Ochsner. 2007. An improved approach for measurement of coupled heat and water transfer in soil cells. Soil Sci. Soc. Am. J. 71:872–880. With permission. Soil Science Society of America.)

transfer in terms of matric pressure head and temperature gradients. Their formulation accounted for inhomogeneous soil and hysteresis of the moisture retention process.

Nassar and Horton (1989c) introduced a theory describing water redistribution for conditions involving temperature, water, and solute concentration gradients. This theory was expanded by Nassar and Horton (1992) to include heat, water, and solute movement under soil water content, temperature, and solute concentration gradients. The theory includes several heat and mass transfer coefficients. The water, heat, and solute transport coefficients are usually affected by the water content, temperature, and solute concentration.

There are currently only a limited number of published reports that include values of these coefficients, especially under the combined effect of water content, temperature, and solute concentration (e.g., Nassar and Horton, 1997). Effects of a wide range of soil temperatures, 5°C–55°C, on water redistribution in solute-free soil columns were reported by Nassar et al. (1992a). A net water transfer was found toward the cold ends of the soil columns with increasing mean soil temperature. It was concluded that the temperature dependence of several soil water transfer coefficients affected the water transfer rates.

Values for hydraulic conductivity and isothermal and thermal vapor diffusivities as functions of soil water content and temperature were reported by Scanlon and Milly (1994). Hydraulic conductivity increased with increasing soil water content and temperature. Also, the isothermal and thermal vapor diffusivities increased with increasing temperature whereas they increased with increasing soil water content to certain values then decreased with further increases in the soil water content. Nassar and Horton (1997) describe how to determine experimentally and/or theoretically derived values of the transport coefficients of heat, water, and solute in soils. They present coefficient values as functions of temperature, water content, and solute concentration.

There are limitations in the models used to predict coupled heat and water transfer in soil (e.g., Philip and de Vries, 1957; Cary and Taylor, 1962; Milly, 1982; Nassar and Horton, 1997; Cahill and Parlange, 1998). Cahill and Parlange (1998) and Parlange et al. (1998) report that existing diffusion-based theory could not explain the large vapor fluxes and evaporation rates measured in their field study. They proposed that diurnal heating of the soil causes convective transport of soil air due to temperature-driven expansion of the air, and this convective enhancement (soil breathing) drives the large unexplained water vapor fluxes observed in field studies. This concept has not yet been confirmed. Philip and de Vries (1957) introduced a thermal vapor enhancement factor to account for pore-scale features (e.g., unequal thermal gradients in various phases and nongas phase diffusive pathways in the pores). Using Philip and de Vries theory to describe nonisothermal evaporation, Bachmann et al. (2001) found a thermal vapor enhancement factor of 3.5 (meaning that thermal vapor transport was about 3.5 times larger than expected) in a wettable soil, but only 0.35 in a hydrophobic soil. In other words, soil columns that differed only in wettability showed very different water evaporation rates and water distribution. No existing model can describe the pore-scale behaviors that led to these greatly different thermal vapor enhancement factors. There is growing appreciation for the utility of pore-scale models (e.g., Ewing and Horton, 2007) for gaining insight into the relative importance of coupled transport processes. Hopefully, in the near future, pore-scale models will become further developed, tested, and used to improve understanding and description of coupled heat and water processes in soil.

Diffusion-based theory first introduced by Philip and de Vries (1957) has long been used to describe coupled soil heat and water transfer. Discrepancies between measurements and theory point to the need for more careful evaluation of the theory. Heitman et al. (2008a) used data obtained from laboratory experiments with transient boundary conditions in an attempt to both calibrate and test calculations of coupled heat and water transfer. Simple calibration for a single boundary and initial condition through adjustment of saturated hydraulic conductivity and thermal vapor enhancement factors provided a reasonable match for transient boundary conditions with gradual changes and at additional initial conditions. The findings of Heitman et al. (2008a) indicate that after soil-specific calibration to one coupled heat and water transfer condition,

the Philip and de Vries theory can be used to predict coupled heat and water transfer under other soil conditions.

9.6 Measuring Soil Temperature

Soil temperature is commonly measured using nonelectric thermometers, thermocouples, or thermistors. Nonelectric thermometers operate based on thermal expansion of liquids or solids. These simple devices are appropriate when high accuracy or precision is not required. A thermocouple produces a small voltage (via the Seebeck effect) proportional to the temperature difference between its measuring junction and its reference junction. A sensitive measuring device can be used to measure and record the voltage and corresponding temperature at the measuring junction if the temperature of the reference junction is known by some other means. Thermistors are ceramic semiconductors whose resistance is strongly temperature dependent. By measuring the resistance, one can infer the temperature. Thermistors are generally more sensitive than either nonelectric thermometers or thermocouples. The World Meteorological Organization recommends routine soil temperature measurement depths of 5, 10, 20, 50, and 100 cm (WMO, 2006). A measurement frequency of once every 2 h is often sufficient to describe the diurnal soil temperature cycle. Measurement objectives, available equipment, and site characteristics may dictate different depth and measurement frequency choices. For additional information on soil temperature measurement, consult the chapters by McInnes (2002) and Ochsner (2008).

A significant recent development is the application of distributed temperature sensing in soils. A distributed temperature sensing instrument can be used to determine the temperature at each location along a fiber-optic cable. Temperatures can be measured with a precision of 0.1°C and spatial resolution of 1 m along a cable, which is several kilometers long (Selker et al., 2006). With the aid of a special plow, the cables can be installed at shallow depths below the soil surface and used to measure the spatial distribution of soil temperature over large areas (Steele-Dunne et al., 2010).

9.7 Measuring and Interpreting Soil Heat Fluxes

Researchers have noted the strong connection between soil-water evaporation and soil heat flux (e.g., Gardner and Hanks, 1966; Mayocchi and Bristow, 1995). Unfortunately, an inability to measure heat transfer at the fine scale necessary to observe evaporative processes has limited investigations. Soil heat flux is typically measured below the soil surface, and a correction is made for the change in sensible heat storage above the flux measurement depth (e.g., Fuchs and Tanner, 1968; Massman, 1993b). This approach ignores the transient, depth-dependence of the evaporation front below the soil surface (de Vries and Philip, 1986).

Evaporation plays a central role in the movement of heat and water within the soil (Yamanaka et al., 1998; Wythers et al., 1999) and the atmosphere. For example, evaporation of soil water

accounts for 20% of the heat budget in the Negev during the dry season (Agam et al., 2004). Despite this importance, evaporation is often erroneously described as occurring at the soil surface rather than within the soil. The surface soil heat flux is generally assumed only to account for conductive heat transfer into the soil. This is a restrictive assumption, because as a soil dries from the surface downward, an increasing fraction of the soil-water evaporation occurs below the surface (Mayocchi and Bristow, 1995; Yamanaka et al., 1998). The assumption that all evaporation occurs at the soil surface does not account for the evaporation zone that propagates downward into the soil.

The Community Land Model (CLM3.5; Oleson et al., 2007), perhaps the most detailed of the third-generation land surface models, assumes that evaporation occurs at the soil surface. As a result, soil evaporation estimates are likely to be systematically incorrect (Mahfouf and Noilhan, 1991; Milly, 1992; Guo and Dirmeyer, 2006). The current deficient treatment of the near-surface land/atmosphere coupling is largely due to a lack of appropriate measurement techniques. Aboveground techniques such as eddy covariance (McInnes and Heilman, 2005) and Bowen ratio (Fritschen and Fritschen, 2005) provide useful approaches to estimate evapotranspiration. However, the inability to measure near-surface movement of heat and water has prohibited an accurate assessment of the near-surface water evaporation and heat balance (de Silans et al., 1997). Recent advances in using the HP method for measuring soil thermal properties, heat flux, and heat storage (e.g., Ochsner et al., 2006, 2007) provide a new opportunity to investigate time and depth patterns of heat transfer and evaporation within soil.

9.7.1 Measuring Soil Heat Fluxes

Soil heat flux measurements are most frequently obtained with the use of soil heat flux plates (Sauer, 2002). These small rigid plates contain a thermopile whose voltage output is proportional to the temperature difference between the upper and lower faces of the plate. The relationship between that temperature difference and the conductive heat flux through the plate is calibrated in a laboratory setting. In the field, the assumption is made that the heat flux through the plate is the same as that through the surrounding soil at the same depth; thus, the voltage output from the plate (mV) is divided by an empirical constant (mV m^2 W^{-1}) to estimate the soil heat flux (W m^{-2}). Several criteria must be met for heat flux plates to produce accurate data: (1) there must be negligible thermal contact resistance between the plate and the soil, (2) the thermal properties of the plate must be the same as those of the surrounding soil, and (3) heat transfer through the soil must be by conduction alone. These criteria are rarely, if ever, satisfied in the field; thus, significant errors may be unavoidable. The typical result of violations of the three criteria above is an underestimation of *G*. Recent studies have found that commercially available heat flux plates underestimate *G* from 2% to 73% in the laboratory (Sauer et al., 2003, 2007) and from 18% to 66% in the field (Ochsner et al., 2006). The data in Figure 9.10a illustrate the underestimation, which is typical of soil heat flux plates. These data show that the magnitude of *G* at a 5 cm depth
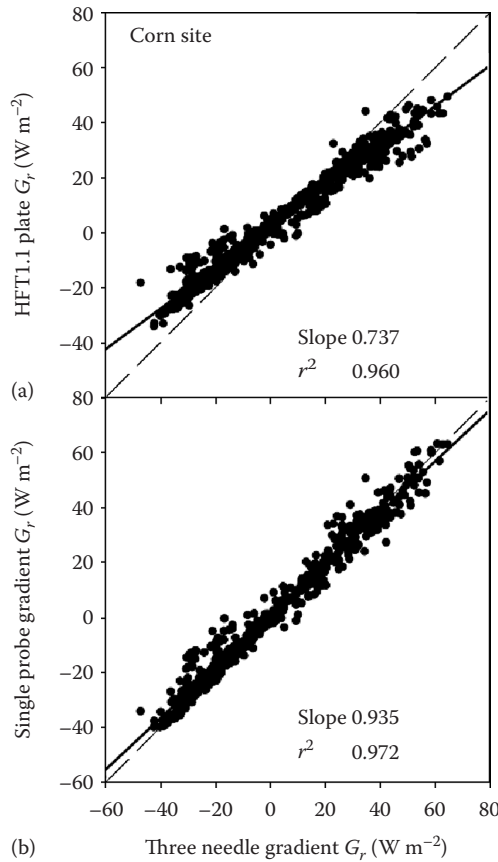
FIGURE 9.10 Soil heat flux at 5 cm (G_r) under corn as measured by heat flux plates (REBS HFT1.1) shown in (a) and the single-probe gradient method (b) versus the three-needle gradient method. The solid lines are the regression lines, and the dashed lines are the one-to-one lines. (Reprinted from Ochsner, T.E., T.J. Sauer, and R. Horton. 2006. Field tests of the soil heat flux plate method and some alternatives. Agron. J. 98:1005–1014. With permission. American Society of Agronomy.)

was underestimated by ~26% using some of the best available soil heat flux plates. Some other commercially available plates exhibit significantly worse performance (Ochsner et al., 2006).

Two primary alternatives to the traditional soil heat flux plate method are the three-needle gradient method and the self-calibrating plate method. The gradient method typically employs a three-needle HP sensor capable of measuring the ambient soil temperature gradient and soil thermal conductivity (Cobos and Baker, 2003; Ochsner et al., 2006). Soil heat flux is then calculated directly from Equation 9.7. This three-needle gradient method has produced excellent results in field tests (Cobos and Baker, 2003; Ochsner et al., 2006). Figure 9.10b illustrates the performance of the three-needle gradient method by comparing it to an independent single-probe gradient method. The magnitude of G at 5 cm measured by these two methods agrees to within ~6%. The three-needle gradient method offers additional benefits in that soil temperature, thermal properties, and water content can be obtained simultaneously with the same sensor. The method has been recently extended to measuring G under freezing and thawing conditions (Ochsner and Baker, 2008).

The self-calibrating plate method uses a soil heat flux plate to which a film heater is added on the upper face. The plate violates the three criteria identified in the first paragraph of this section; however, by introducing a known heat flux via the heater at predetermined intervals (e.g., 3 h), an in situ calibration constant is obtained that varies with the soil conditions. The procedure results in more accurate soil heat flux data than the traditional plate method (Ochsner et al., 2006). Increased adoption of the three-needle gradient method and the self-calibrating plate method should lead to more accurate soil heat flux and surface energy balance data.

9.7.2 Interpreting Soil Heat Fluxes

It has been suggested that a sensible heat balance offers a means to track the soil evaporation front within the soil (Gardner and Hanks, 1966; Mayocchi and Bristow, 1995). Evaporation of soil water represents a large heat sink. A sensible heat balance is used to determine the amount of latent heat involved with vaporization of soil water following Gardner and Hanks (1966):

$$(H_0 - H_1) - \Delta S = LE \qquad (9.32)$$

where
  H_0 and H_1 are soil heat fluxes (W m^{-2}) at depths 0 and 1, respectively
  ΔS (W m^{-2}) is the change in soil sensible heat storage between depths 0 and 1
  L (J m^{-3}) is the latent heat of vaporization
  E is evaporation (m s^{-1})

Using measurements of soil thermal conductivity (λ, W m^{-1} °C^{-1}) and the temperature gradient (dT/dz, °C m^{-1}), values for H can be determined from Fourier's law (Ochsner et al., 2006). Ochsner et al. (2007) provide several approximations for determining ΔS from measured temperature (T, °C) and soil volumetric heat capacity (C, J m^{-3} °C^{-1}). One approximation is

$$\Delta S = \sum_{i=1}^{N} C_{i,j-1} \frac{T_{i,j} - T_{i,j-1}}{t_j - t_{j-1}} (z_i - z_{i-1}), \qquad (9.33)$$

where
  z (m) is depth
  the subscripts i and j are index variables for depth layers and time steps, respectively

The residual from the energy balance provided by the left side of Equation 9.32 is attributed to LE, which cannot be measured directly. Values for L can be calculated from Equation 9.4 using the soil temperature at the depth of interest. Having values for H, ΔS, and L, it is possible to calculate E from Equation 9.32.

Heitman et al. (2008b, 2008c) placed several three-needle HP sensors at shallow depths in a soil. The central needles of the sensors were positioned at 6, 12, 18, 24, 45, and 60 mm below the soil surface. The plane formed by the three needles

of each sensor was oriented perpendicular to the soil surface. (a) The sensors were connected to a data acquisition system, and HP measurements and soil temperature measurements were obtained each 4 and 0.5 h, respectively. They were able to use their measurements to estimate sensible and latent heat fluxes with depth and time. The latent heat flux estimates allowed them to estimate soil water evaporation as a function of depth and time, thus revealing the temporal patterns of in situ evaporation and condensation. This approach is measurement driven, so it does not require determination of coupled heat and water transfer coefficients (e.g., Nassar and Horton, 1997) or soil-specific hydraulic properties.

Results for a 6 day period are presented in Figures 9.11 and 9.12. Data for the uppermost HP sensor (centered at 6 mm) are

FIGURE 9.12 Evaporation determined by heat balance using HP sensors. Data were collected in 2007 from a bare field in Iowa following rainfall on DOY 172. (Reprinted from Heitman, J.L., X. Xiao, R. Horton, and T.J. Sauer. 2008c. Sensible heat measurements indicating depth and magnitude of subsurface soil water evaporation. Water Resour. Res. 44:W00D05. With permission.)

shown in Figure 9.10. During this time period, daily maximum ambient T generally increases after rainfall through day of year (DOY) 178 at the 0, 6, and 12 mm depths (Figure 9.11). Drying in the upper portion of the soil profile also produces decreases in both C and λ (Figure 9.11). Accompanying these changes are shifts in the magnitude of dT/dz at the 3 and 9 mm depths. The magnitude of dT/dz is similar with depth through DOY 175 (Figure 9.11). On DOY 176, the peak magnitude of dT/dz at 3 mm begins to increase and thereafter remains relatively large. A shift also occurs at 9 mm on DOY 176, but peak magnitudes remain well below those at 3 mm until DOY 178 when the gradients begin to converge. This indicates that the zone of evaporation has moved below 9 mm. Driven by dT/dz, heat flux density demonstrates a similar pattern (Figure 9.11). The 3 and 9 mm deep heat flux densities are nearly identical through DOY 175. However, beginning on DOY 176, divergence in the heat flux density with depth indicates significant heat loss (>150 W m^{-2}) as heat is transferred through the soil. The amount of heat partitioned to ΔS can be quantified and remains consistently small throughout this period (<25 W m^{-2}). The difference between heat flux density at 3 and 9 mm (i.e., G_1 and G_2) appreciably exceeds ΔS through DOY 178 and thereby allows determination of LE with Equation 9.33.

Results for sensors measuring the 3–9, 9–15, 15–21, and 21–27 mm depth increments are shown in Figure 9.12. Note that data are presented as evaporation rate, E (mm h^{-1}), rather than LE. The evaporation rate remains near 0 for all measured depth increments through DOY 175, which indicates that the evaporation zone has not passed below the 3 mm soil depth. All sensible heat transferred through the 3 mm depth is accounted for through ΔS or heat flux density at lower depths. On DOY 176, the peak magnitude of E increases to 0.3 and 0.15 mm h^{-1} during the afternoon at the 3–9 and 9–15 mm depth increments, respectively. The peak magnitude of E continues to increase for both depth increments on DOY 177

FIGURE 9.11 Measurements obtained with HP sensors: T is soil temperature, C is volumetric heat capacity and λ is thermal conductivity, dT/dz is temperature gradient, and G and ΔS are heat flux densities and change in sensible heat storage. Data were collected in 2007 from a bare field in Iowa following rainfall on DOY 172. (Reprinted from Heitman, J.L., X. Xiao, R. Horton, and T.J. Sauer. 2008c. Sensible heat measurements indicating depth and magnitude of subsurface soil water evaporation. Water Resour. Res. 44:W00D05. With permission.)

before declining on subsequent days. Net evaporation rates are generally highest immediately following rainfall, assuming that atmospheric demand is not limiting (Lemon, 1956). Thus, the relatively lower E on DOY 176 than 177 does not necessarily suggest less total evaporation. Rather, it indicates that evaporation is still occurring in the soil layer above the 3 mm depth on DOY 176. Transition to evaporation at deeper soil depths is beginning to occur on DOY 176 as soil water stored above the 3 mm depth is depleted and cannot meet atmospheric demand. After DOY 177, measurements of the 3 mm heat flux density (Figure 9.11) suggest that nearly all evaporation occurs below the 3 mm soil depth. The shifting pattern of evaporation continues on subsequent days as soil water storage is further depleted near the surface and peak E at 9–15 mm begins to exceed E at 3–9 mm (Figure 9.12). The declining peak magnitudes of E for all depth increments after DOY 177 are indicative of decreased total evaporation as soil water becomes limiting.

9.8 Managing the Soil Thermal Regime

Agricultural management practices, including irrigation, drainage, tillage, and crop residue management, have the potential to affect the thermal properties of soils and therefore soil thermal regime. In particular, the effect of tillage and crop residue management on soil heat flux has been the subject of several studies (Allmaras et al., 1977; Pikul et al., 1985; Enz et al., 1988; Azooz et al., 1997; Richard and Cellier, 1998; Sauer et al., 1998). Tillage loosens the surface soil, although some local compaction may also occur. Lower soil bulk density generally translates to lower λ; thus, lower G has been observed in tilled soil as compared to untilled or compacted soil (Azooz et al., 1997; Richard and Cellier, 1998). Tillage can also influence surface energy partitioning through its effects on albedo, emissivity, and near-surface water content.

Crop residues can strongly influence the soil thermal regime. They can shade the soil, increase the shortwave albedo, and change the longwave emissivity of the land surface. Crop residues slow the rate of soil water evaporation and can result in higher soil water content. Water content, in turn, influences soil physical properties and surface energy partitioning. Compared to bare soil, crop residues can reduce extremes of heat and mass fluxes at the soil surface. Managing crop residues can result in more favorable agronomic soil conditions. The quantity, quality, architecture, and surface distribution of crop residues affect soil surface radiation and energy balances, soil water content, and soil temperature (e.g., Bristow, 1988). Crop residue has a low thermal conductivity and, whether lying on the soil surface or incorporated into the soil by tillage, inhibits heat transfer into the soil. Residue layers also have a shortwave reflectivity that is higher than most soils and provide a barrier to vapor flow (Gausman et al., 1975; Horton et al., 1996; Sauer et al., 1997). Thus, soils with a large proportion of the surface covered by crop residue tend to have higher water contents, lower temperatures, and lower G. Such changes in soil thermal regime, of course,

have implications for the surface energy balance and evaporation. Horton et al. (1996) provide a review of crop residue effects on surface radiation and energy balances.

Plastic "mulches" (i.e., plastic film) placed on the soil surface are another method often used to manage the soil thermal regime. Plastic mulches modify the soil thermal regime by changing the radiation budget of the surface and by suppressing evaporation (Ham et al., 1993). Plastic mulches usually result in higher soil temperatures, but the effects of plastic mulch depend on the optical properties of the mulch and on thermal contact resistance between the plastic and the soil surface (Ham and Kluitenberg, 1994).

9.9 Soil Temperature and Global Climate Change

Soil temperature and temperature of underlying geologic formations are strongly linked with global climate. The heat capacity of soil and rock is much larger than that of air; thus, if heat is accumulating on earth, we should expect more heat storage in the soil and rock than in the atmosphere. Baker and Baker (2002) examined a long-term data set of subsurface temperature measurements to quantify subsurface heat accumulation. They found an average annual downward flux of 1.1 MJ m^{-2} at the 1.6 m depth over a 37-year period. At the 12.8 m depth, the average annual downward heat flux was 0.22 MJ m^{-2}. Similarly, Beltrami et al. (2002) found a net downward long-term heat flux at the land surface with a global average of 1.2 MJ m^{-2} annually in the period from 1950 to 2000. Total heat accumulation in the continental lithosphere from 1950 to 2000 (9.1×10^{21} J) exceeded heat accumulation in the atmosphere (6.6×10^{21} J) and was second only to heat accumulation in the world's oceans (182×10^{21} J).

There is evidence that subsurface heat storage is not adequately simulated in general circulation models (GCMs), which are used for climate change projections. Current GCMs have 0 heat flux bottom boundary conditions at depths of between 1.5 and 10 m below the land surface (Smerdon and Stieglitz, 2006). The results of Baker and Baker (2002) indicate that >20% of subsurface heat storage occurs below the 10 m depth; thus, even the GCM with the deepest bottom boundary (ECHO-G, 10 m) may contain significant errors. In a simulation study, a land model with a 600 m deep bottom boundary stored 6.2 times more heat than ECHO-G under Intergovernmental Panel on Climate Change (IPCC) emission scenario A2 (MacDougall et al., 2008). To put this result in perspective, the difference in predicted continental heat storage between the models with shallow and deep bottom boundaries was 8.2×10^{22} J over the period from 1991 to 2100, while the estimated heat accumulation in the atmosphere from 1955 to 2003 was only 0.5×10^{22} J. Neglecting deep-subsurface heat storage in GCMs may lead to overpredicting soil temperature changes, overpredicting ocean- and atmosphere-temperature changes, and misrepresenting a host of atmosphere–land surface feedbacks and linkages. Clearly, a better understanding of the soil thermal regime and its connection with global climate is needed.

9.10 Concluding Remarks

Research to date has produced a rich store of knowledge on the soil thermal regime at scales ranging from the core sample to the field soil profile. This chapter has hopefully served to briefly introduce the key topics and direct the reader to the wealth of available literature. Still, there remain wide frontiers of unexplored phenomena at outlying scales both small (e.g., aggregate, pore, and particle scales) and large (e.g., field, watershed, and continental scales). With the continued advancement of measurement capabilities, there should be no shortage of research possibilities in the ongoing study of the soil thermal regime.

References

Abu-Hamdeh, N.H., and R.C. Reeder. 2000. Soil thermal conductivity: Effects of density, moisture, salt concentration, and organic matter. Soil Sci. Soc. Am. J. 64:1285–1290.

Agam, N., P.R. Berliner, A. Zangvil, and E. Ben-Dor. 2004. Soil water evaporation during the dry season in an arid zone. J. Geophys. Res. 109. doi: 10.1029/2004JD004802.

Allmaras, R.R., E.A. Hallauer, W.W. Nelson, and S.D. Evans. 1977. Surface energy balance and soil thermal property modifications by tillage-induced soil structure. Minn. Agric. Exp. Sta. Tech. Bull. 306:41.

Al Nakshabandi, G., and H. Kohnke. 1965. Thermal conductivity and diffusivity of soils as related to moisture tension and other physical properties. Agric. Meteorol. 2:271–279.

Azooz, R.H., B. Lowery, T.C. Daniel, and M.A. Arshad. 1997. Impact of tillage and residue management on soil heat flux. Agric. For. Meteorol. 84:207–222.

Bachmann, J., R. Horton, and R.R. van der Ploeg. 2001. Isothermal and nonisothermal evaporation from four sandy soils of different water repellency. Soil Sci. Soc. Am. J. 65:1599–1607.

Baker, J.M., and D.G. Baker. 2002. Long-term ground heat flux and heat storage at a mid-latitude site. Clim. Change 54:295–303.

Baldocchi, D.B., S.B. Verma, and N.J. Rosenberg. 1985. Water use efficiency in a soybean field: Influence of plant water stress. Agric. For. Meteorol. 34:53–65.

Bear, J., and Bensabat, J. 1989. Advective fluxes in multiphase porous media under nonisothermal conditions. Transp. Porous Media 4:423–448.

Bear, J., and Gilman, A. 1995. Migration of salts in the unsaturated zone caused by heating. Transp. Porous Media 19:139–156.

Beltrami, H., J.E. Smerdon, H.N. Pollack, and S. Huang. 2002. Continental heat gain in the global climate system. Geophys. Res. Lett. 29:1–3.

Benjamin, J.G., M.R. Ghaffarzadeh, and R.M. Cruse. 1990. Coupled water and heat transport in ridged soils. Soil Sci. Soc. Am. J. 54:963–969.

Bilskie, J.R., R. Horton, and K.L. Bristow. 1998. Test of a dual-probe heat-pulse method for determining thermal properties of porous materials. Soil Sci. 163:346–355.

Bouyoucos, G.T. 1915. Effect of temperature on the movement of water vapor and capillary moisture in soils. J. Agric. Res. 5:141–172.

Bristow, K.L. 1987. On solving the surface energy balance equation for surface temperature. Agric. For. Meteorol. 39:49–54.

Bristow, K.L., 1988. The role of mulch and its architecture in modifying soil temperatures. Aust. J. Soil Res. 26:269–280.

Bristow, K.L. 1998. Measurement of thermal properties and water content of unsaturated sandy soil using dual-probe heat-pulse probes. Agric. For. Meteorol. 89:75–84.

Bristow, K.L. 2002. Thermal conductivity, p. 1209–1226. In J.H. Dane and G.C. Topp (eds.) Methods of soil analysis: Part 4. Physical methods. SSSA, Madison, WI.

Bristow, K.L., G.S. Campbell, R.I. Papendick, and L.F. Elliott. 1986. Simulation of heat and moisture transfer through a surface residue-soil system. Agric. For. Meterol. 36:193–214.

Bristow, K.L., and R. Horton. 1996. Modeling the impact of partial surface mulch on soil heat and water flow. Theor. Appl. Climatol. 54:85–98.

Bristow, K.L., G.J. Kluitenberg, and R. Horton. 1994. Measurement of soil thermal properties with a dual-probe heat-pulse technique. Soil Sci. Soc. Am. J. 58:1288–1294.

Brown, K.W. 1976. Sugar beet and potatoes, p. 65–86. In J.L. Monteith (ed.) Vegetation and the atmosphere. Vol. 2. Academic Press, New York.

Bruijn, P.J., I.A. van Haneghem, and J. Schenk. 1983. An improved nonsteady-state probe method for measurements in granular materials. Part I: Theory. High Temp. High Press. 15:359–366.

Buchan, G.D. 2001. Soil temperature regime, p. 539–594. In K.A. Smith and C.E. Mullins (eds.) Soil and environmental analysis: Physical methods. Marcel Dekker, New York.

Cahill, A.T., and M.B. Parlange. 1998. On water vapor transport in field soils. Water Resour. Res. 34:731–739.

Campbell, G.S. 1985. Soil physics with BASIC: Transport models for soil-plant systems. Elsevier, New York.

Campbell, G.S., C. Calissendorff, and J.H. Williams. 1991. Probe for measuring soil specific heat using a heat-pulse method. Soil Sci. Soc. Am. J. 55:291–293.

Campbell, G.S., J.D. Jungbauer, W.R. Bidlake, and R.D. Hungerford. 1994. Predicting the effect of temperature on soil thermal conductivity. Soil Sci. 158:307–313.

Campbell, D.I., C.E. Laybourne, and I.J. Blair. 2002. Measuring peat moisture content using the dual-probe heat pulse technique. Aust. J. Soil Res. 40:177–190.

Carslaw, H.S., and J.C. Jaeger. 1959. Conduction of heat in solids. 2nd edn. Oxford University Press, Oxford, U.K.

Cary, J.W. 1966. Soil moisture transport due to thermal gradients: Practical aspects. Soil Sci. Soc. Am. Proc. 30:428–433.

Cary, J.W., and S.A. Taylor. 1962. The interaction of the simultaneous diffusions of heat and water vapor. Soil Sci. Soc. Am. J. 26:413–416.

Choudhury, B.J., S.B. Idso, and J.R. Reginato. 1987. Analysis of an empirical model for soil heat flux under a growing wheat crop for estimating evaporation by an infrared-temperature based energy balance equation. Agric. For. Meteorol. 39:283–297.

Chudnovski, A.F. 1962. Heat transfer in soil. Translated from Russian. Published for the National Science Foundation, Washington DC, by the Israel Program for Scientific Translations, Jerusalem, Israel.

Chung, S.O., and R. Horton. 1987. Soil heat and water flow with a partial surface mulch. Water Resour. Res. 23:2175–2186.

Clothier, B.E., K.L. Clawson, P.J. Pinter, Jr., M.S. Moran, R.J. Reginato, and R.D. Jackson. 1986. Estimation of soil heat flux from net radiation during the growth of alfalfa. Agric. For. Meteorol. 37:319–329.

Cobos, D.R., and J.M. Baker. 2003. In situ measurement of soil heat flux with the gradient method. Vadose Zone J. 2:589–594.

Constantz, J., S.W. Tyler, and E. Kwicklis. 2003. Temperature-profile methods for estimating percolation rates in arid environments. Vadose Zone J. 2:12–24.

Côté, J., and J.-M. Konrad. 2005. A generalized thermal conductivity model for soils and construction materials. Can. Geotech. J. 42:443–458.

Coulson, S.J., I.D. Hodkinson, A.T. Strathdee, W. Block, N.R. Webb, J.S. Bale, and M.R. Worland. 1995. Thermal environments of arctic soil organisms during winter. Arct. Alp. Res. 27:364–370.

Denmead, O.T. 1969. Comparative micrometeorology of a wheat field and a forest of *Pinus radiata*. Agric. Meteorol. 6:357–371.

de Silans, A.P., B.A. Monteny, and J.P. Lhomme. 1997. The correction of soil heat flux measurements to derive an accurate surface energy balance by the Bowen ratio method. J. Hydrol. 188–189:453–465.

de Vries, D.A. 1952. A nonstationary method for determining thermal conductivity of soil in situ. Soil Sci. 73:83–89.

de Vries, D.A. 1963. Thermal properties of soils, p. 210–235. *In* W.R. van Wijk (ed.) Physics of plant environment. North-Holland Publishing Co., Amsterdam, the Netherlands.

de Vries, D.A., and A.J. Peck. 1958. On the cylindrical probe method of measuring thermal conductivity with special reference to soils. Part 1: Extension of theory and discussion of probe characteristics. Aust. J. Phys. 11:255–271.

de Vries, D.A., and J.R. Philip. 1986. Soil heat flux, thermal conductivity, and the null-alignment method. Soil Sci. Soc. Am. J. 50:12–17.

Enz, J.W., L.J. Brun, and J.K. Larsen. 1988. Evaporation and energy balance for bare and stubble covered soil. Agric. For. Meteorol. 43:59–70.

Ewing, R.P., and R. Horton. 2007. Thermal conductivity of a cubic lattice of spheres with capillary bridges. J. Phys. D: Appl. Phys. 40:4959–4965.

Farouki, O.T. 1981. Thermal properties of soils. USA Cold region research engineering laboratory. Monograph 81-1. Cold Regions Research and Engineering Laboratory, Hanover, NH.

Farouki, O.T. 1982. Evaluation of methods for calculating soil thermal conductivity. USA cold region research engineering laboratory. CRREL report 82-8. Cold Regions Research and Engineering Laboratory, Hanover, NH.

Farouki, O.T. 1986. Thermal properties of soils. Series on rock and soil mechanics. Vol. 11. Trans Tech Publications, Clausthal-Zellerfeld, Germany.

Fritschen, L.J., and C.L. Fritschen. 2005. Bowen ratio energy balance method, p. 397–405. *In* J.L. Hatfield and J.M. Baker (eds.) Micrometeorology in agricultural systems. American Society of Agronomy, Madison, WI.

Fuchs, M., G.S. Campbell, and R.I. Papendick. 1978. An analysis of sensible and latent heat flow in partially frozen unsaturated soil. Soil Sci. Soc. Am. J. 42:379–385.

Fuchs, M., and A. Hadas. 1972. The heat flux density in a non-homogeneous bare loessial soil. Boundary-Layer Meteorol. 3:191–200.

Fuchs, M., and C.B. Tanner. 1968. Calibration and field test of soil heat flux plates. Soil Sci. Soc. Am. Proc. 32:326–328.

Gao, Z., X. Fan, and L. Bian. 2003. An analytical solution to one-dimensional thermal conduction–convection in soil. Soil Sci. 168:99–107.

Gardner, H.R., and R.J. Hanks. 1966. Evaluation of the evaporation zone in soil by measurement of heat flux. Soil Sci. Soc. Am. Proc. 30:425–428.

Gausman, H.W., A.H. Gerbermann, C.L. Wiegand, R.W. Leamer, R.R. Rodriguez, and J.R. Noriega. 1975. Reflectance differences between crop residues and bare soils. Soil Sci. Soc. Am. Proc. 39:752–755.

Gawin, D., P. Baggio, and B.A. Schrefler.1995. Coupled heat, water and gas flow in deformable porous media. Int. J. Numer. Methods Fluids 20:969–987.

Gilgen, H., and A. Ohmura. 1999. The global energy balance archive. Bull. Am. Meteorol. Soc. 80:831–850.

Globus, A.M. 1983. Physics of non-isothermal soil moisture transfer. 1st edn. Hydrometeorological Publishing, Leningrad, Russia.

Guo, Z., and P.A. Dirmeyer. 2006. Evaluation of the second global soil wetness project soil moisture simulation: 1. Intermodel comparison. J. Geophys. Res. 111:D22S02.

Gupta, S.C., W.E. Larson, and D.R. Linden, 1983. Tillage and surface residue effects on upper boundary temperatures. Soil Sci. Soc. Am. J. 47:1212–1218.

Gupta, S.C., J.K. Radke, and W.E. Larson. 1981. Predicting temperature of bare and residue covered soils with and without a corn crop. Soil Sci. Soc. Am. J. 45:405–412.

Ham, J.M., and G.J. Kluitenberg. 1994. Modeling the effect of mulch optical properties and mulch-soil contact resistance on soil heating under plastic mulch culture. Agric. For. Meteorol. 71:403–424.

Ham, J.M., G.J. Kluitenberg, and W.J. Lamont. 1993. Optical properties of plastic mulches affect the field temperature regime. J. Am. Soc. Hortic. Sci. 118:188–193.

Ham, J.M., and R.S. Senock. 1992. On the measurement of soil-surface temperature. Soil Sci. Soc. Am. J. 56:370–377.

Hanks, R.J., D.D. Austin, and W.T. Ondrechen. 1971. Soil temperature estimation by a numerical method. Soil Sci. Soc. Am. J. 35:665–677.

Hares, M.A., and M.D. Novak. 1992a. Simulation of surface energy balance and soil temperature under strip tillage. I. Model description. Soil Sci. Soc. Am. J. 56:22.

Hares, M.A., and M.D. Novak. 1992b. Simulation of surface energy balance and soil temperature under strip tillage. II. Field test. Soil Sci. Soc. Am. J. 56: 29.

Heitman, J.L., J.M. Basinger, G.J. Kluitenberg, J.M. Ham, J.M. Frank, and P.L. Barnes. 2003. Field evaluation of the dual-probe heat-pulse method for measuring soil water content. Vadose Zone J. 2:552–560.

Heitman, J.L., R. Horton, T. Ren, I.N. Nassar, and D. Davis. 2008a. A test of coupled soil heat and water transfer prediction under transient boundary conditions. Soil Sci. Soc. Am. J. 72:1197–1207.

Heitman, J.L., R. Horton, T. Ren, and T.E. Ochsner. 2007. An improved approach for measurement of coupled heat and water transfer in soil cells. Soil Sci. Soc. Am. J. 71:872–880.

Heitman, J.L., R. Horton, T.J. Sauer, and T.M. DeSutter. 2008b. Sensible heat observations reveal soil-water evaporation dynamics. J. Hydrometeorol. 9:165–171.

Heitman, J.L., X. Xiao, R. Horton, and T.J. Sauer. 2008c. Sensible heat measurements indicating depth and magnitude of subsurface soil water evaporation. Water Resour. Res. 44:W00D05.

Hopmans, J.W., and J.H. Dane. 1986. Thermal conductivity of two porous media as a function of water content, temperature, and density. Soil Sci. 142:187–195.

Hopmans, J.W., J. Simunek, and K.L. Bristow. 2002. Indirect estimation of soil thermal properties and water flux using heat pulse probes measurements: Geometry and dispersion effects. Water Resour. Res. 38. doi:10.1029/2000WR000071.

Horton, R. 1989. Canopy shading effects on soil heat and water flow. Soil Sci. Soc. Am. J. 53:669–679.

Horton, R., O. Aguirre-Luna, and P.J. Wierenga. 1984a. Observed and predicted two-dimensional soil temperature distributions under a row crop. Soil Sci. Soc. Am. J. 48:1147–1152.

Horton, R., O. Aguirre-Luna, and P.J. Wierenga. 1984b. Soil temperature in a row crop with incomplete surface cover. Soil Sci. Soc. Am. J. 48:1225–1232.

Horton, R., K.L. Bristow, G.J. Kluitenberg, and T.J. Sauer. 1996. Crop residue effects on surface radiation and energy balance—Review. Theor. Appl. Climatol. 54:27–37.

Horton, R., and S. Chung. 1991. Soil heat flow, p. 397–438. *In* J. Hanks, and J.T. Ritchie (eds.) Modeling plant and soil systems. ASA, CSSA, and SSSA, Madison, WI.

Horton, R., and P.J. Wierenga. 1983. Estimating the soil heat flux from observations of soil temperature near the surface. Soil Sci. Soc. Am. J. 47:14–20.

Horton, R., and P.J. Wierenga. 1984. The effect of column wetting on soil thermal conductivity. Soil Sci. 138:102–108.

Horton, R., P.J. Wierenga, and D.R. Nielsen. 1983. Evaluation of methods for determining the apparent thermal diffusivity of soil near the surface. Soil Sci. Soc. Am. J. 47:25–32.

Idso, S.B., R.D. Jackson, R.J. Reginato, B.A. Kimball, and F.S. Nakayama. 1975. The dependence of bare soil albedo on soil water content. J. Appl. Meteorol. 14:109–113.

Jackson, R.D., and D. Kirkham. 1958. Method of measurement of the real thermal diffusivity of moist soil. Soil Sci. Soc. Am. Proc. 22:479–482.

Jackson, R.D., and S.A. Taylor. 1965. Heat transfer, p. 349–356. *In* C.A. Black et al. (ed.) Methods of soil analysis: Part 1. Agronomy Monograph No. 9. ASA, Madison, WI.

Jaeger, J.C. 1965. Application of the theory of heat conduction to geothermal measurements. *In* W.H.K. Lee (ed.) Terrestrial heat flow. Am. Geophys. Union Monogr. 8:7–23.

Jaynes, D.B. 1990. Temperature variation effect on field-measured infiltration. Soil Sci. Soc. Am. J. 54:305–312.

Johansen, O. 1975. Thermal conductivity of soils. Ph.D. Thesis. Institute for Kjoletechnik, Trondheim, Norway. (CRREL draft translation 637, 1977.)

Jury, W.A., and B. Bellantuoni. 1976a. Heat and water movement under surface rocks in a field soil: I. Thermal effects. Soil Sci. Soc. Am. J. 40:505–509.

Jury, W.A., and B. Bellantuoni. 1976b. Heat and water movement under surface rocks in a field soil: II. Moisture effects. Soil Sci. Soc. Am. J. 40:509–513.

Jury, W.A., and J. Letey. 1979. Water vapor movement in soil: Reconciliation of theory and experiment. Soil Sci. Soc. Am. J. 43:823–827.

Kersten, M.S. 1949. Laboratory research for the determination of the thermal properties of soils. ACFEL Technical Report 23. AD 712516.

Kiehl, J.T., and K.E. Trenberth. 1997. Earth's annual global mean energy budget. Bull. Am. Meteorol. Soc. 78:197–208.

Kluitenberg, G.J. 2002. Heat capacity and specific heat, p. 1201–1208. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

Kluitenberg, G.J., K.L. Bristow, and B.S. Das. 1995. Error analysis of the heat pulse method for measuring soil heat capacity, diffusivity and conductivity. Soil Sci. Soc. Am. J. 59:719–726.

Kluitenberg, G.J., J.M. Ham, and K.L. Bristow. 1993. Error analysis of the heat-pulse method for measuring the volumetric heat capacity of soil. Soil Sci. Soc. Am. J. 57:1208–1215.

Kluitenberg, G.J., and R. Horton. 1990. Analytical solution for two-dimensional heat conduction beneath a partial surface mulch. Soil Sci. Soc. Am. J. 54:1197–1206.

Knight, J.H., and G.J. Kluitenberg. 2004. Simplified computational approach for dual-probe heat-pulse method. Soil Sci. Soc. Am. J. 68:447–449.

Larson, T.H. 1988. Thermal measurement of soils using a multineedle probe with a pulsed point-source. Geophysics 53:266–270.

Lascano, R.J., C.H.M. van Bavel, J.L. Hatfield, and D.R. Upchurch. 1987. Energy and water balance of a sparse crop: Simulated and measured soil and crop evaporation. Soil Sci. Soc. Am. J. 51:1113–1121.

Lemon, E.R. 1956. The potentialities for decreasing soil moisture evaporation loss. Soil Sci. Soc. Am. Proc. 20:120–125.

Lemon, E.R. 1963. The energy budget at the earth's surface—Part 1. U.S. army electronic proving ground production research report no. 71. USDA, Washington, DC.

Lettau, H., and B. Davidson (eds.). 1957. Exploring the atmosphere's first mile—Vol. I. Instrumentation and data evaluation. Pergamon Press, New York.

Lu, S., T. Ren, Y. Gong, and R. Horton. 2007. An improved model for predicting room temperature soil thermal conductivity versus water content. Soil Sci. Soc. Am. J. 71:8–14.

Lubimova, H.A., L.M. Lusova, F.V. Firsov, G.N. Starikova, and A.P. Shushpanov. 1961. Determination of surface heat flow in Mazesta (USSR). Annali di Geophys. XIV:157–167.

MacDougall, A.H., J.F. González-Rouco, B.M. Stevens, and H. Beltrami. 2008. Quantification of subsurface heat storage in a GCM simulation. Geophys. Res. Lett. 35:L13702. doi:10.1029/2008GL034639.

Mahfouf, J.F., and J. Noilhan. 1991. Comparative study of various formulations of evaporation from bare soil using in situ data. J. Appl. Meteorol. 30:1354–1365.

Mahrer, Y. 1982. A theoretical study of the effect of soil surface shape upon the soil temperature profile. Soil Sci. 134:381–387.

Mahrer, Y., and J. Katan. 1981. Spatial soil temperature regime under transparent polyethylene mulch: Numerical and experimental studies. Soil Sci. 131:82–87.

Massman, W.J. 1993a. Periodic temperature variations in an inhomogeneous soil: A comparison of approximate and exact analytical expressions. Soil Sci. 155:331–338.

Massman, W.J. 1993b. Errors associated with the combination method for estimating soil heat flux. Soil Sci. Soc. Am. J. 57:1198–1202.

Mayocchi, C.L., and K.L. Bristow. 1995. Soil surface heat flux: Some general questions and comments on measurements. Agric. For. Meteorol. 75:43–50.

McInnes, K.J. 2002. Soil heat: 5.1 Temperature, p. 1183–1199. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. Physical methods. SSSA, Madison, WI.

McInnes, K.J., and J.L. Heilman. 2005. Relaxed eddy accumulation, p. 437–453. *In* J.L. Hatfield and J.M. Baker (eds.) Micrometeorology in agricultural systems. ASA, Madison, WI.

Milly, P.C.D. 1982. Moisture and heat transport in hysteretic, inhomogeneous porous media: A matric head-based formulation and a numerical model. Water Resour. Res. 18:489–498.

Milly, P.C.D. 1992. Potential evaporation and soil moisture in general circulation models. J. Clim. 5:209–226.

Milly, P.C.D., and P.S. Eagleson. 1980. The coupled transport of water and heat in a critical soil column under atmospheric excitation. Technical report No. 258. R.M. Parsons Laboratory, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Monteith, J.L. 1958. The heat balance beneath crops. Climatology and microclimatology. UNESCO Arid Zone Res. 11:123–128.

Nassar, I.N., A.M. Globus, and R. Horton. 1992a. Simultaneous soil heat and water transfer. Soil Sci. 154:465–472.

Nassar, I.N., and R. Horton. 1989a. Determination of the apparent thermal diffusivity of a nonuniform soil. Soil Sci. 147:238–244.

Nassar, I.N., and R. Horton. 1989b. Water transport in unsaturated nonisothermal, salty soil: 1. Experimental results. Soil Sci. Soc. Am. J. 53:1323–1329.

Nassar, I.N., and R. Horton. 1989c. Water transport in unsaturated nonisothermal, salty soil: 2. Theoretical development. Soil Sci. Soc. Am. J. 53:1330–1337.

Nassar, I.N., and R. Horton. 1990. Composition of soil apparent thermal diffusivity from multiharmonic temperature analysis for nonuniform soils. Soil Sci. 149:125–130.

Nassar, I.N., and R. Horton. 1992. Simultaneous transfer of heat, water, and solute in porous media: I. Theoretical development. Soil Sci. Soc. Am. J. 56:1350–1356.

Nassar, I.N., and R. Horton. 1997. Heat, water, and solute transfer in unsaturated porous media: I. Theory development and transport coefficient evaluation. Transp. Porous Media 27:17–38.

Nassar, I.N., R. Horton, and A.M. Globus. 1992b. Simultaneous transfer of heat, water, and solute in porous media: II. Experiment and analysis. Soil Sci. Soc. Am. J. 56:1357–1365.

Neary, D.G., C.C. Klopatek, L.F. DeBano, and P.F. Ffolliott. 1999. Fire effects on belowground sustainability: A review and synthesis. For. Ecol. Manage. 122:51–71.

Nix, G.H., G.W. Lowery, R.I. Vachon, and G.E. Tanger. 1967. Direct determination of thermal diffusivity and conductivity with a refined line-source technique. *In* G. Heller (ed.) Progress in astronautics and aeronautics. Vol. 20. Thermophysics of spacecraft and planetary bodies. Academic Press, New York.

Nix, G.H., R.I. Vachon, G.W. Lowery, and T.A. McCurry. 1969. The line-source method: Procedure and iteration scheme for combined determination of conductivity and diffusivity, p. 999–1008. *In* C.Y. Ho and R.E. Taylor (eds.) Thermal conductivity. Proceedings of the Eighth Conference, October 7–10, 1968. Thermophysical Properties Research Center, Purdue University, West Lafayette, IN.

Noborio, K., K.J. McInnes, and J.L. Heilman. 1996a. Two-dimensional model for water, heat, and solute transport in furrow-irrigated soil: I. Theory. Soil Sci. Soc. Am. J. 60:1001–1009.

Noborio, K., K.J. McInnes, and J.L. Heilman. 1996b. Two-dimensional model for water, heat, and solute transport in furrow-irrigated soil: II. Field evaluation. Soil Sci. Soc. Am. J. 60:1010–1021.

Novak, M.D. 1993. Analytical solutions for two-dimensional soil heat flow with radiation surface boundary conditions. Soil Sci. Soc. Am. J. 57:30–39.

Ochsner, T.E. 2008. Measuring soil temperature, p. 235–251. *In* S.D. Logsdon et al. (eds.) Soil science: Step-by-step field analyses. SSSA, Madison, WI.

Ochsner, T.E., and J.M. Baker. 2008. In situ monitoring of soil thermal properties and heat flux during freezing and thawing. Soil Sci. Soc. Am. J. 72:1025–1032.

Ochsner, T.E., R. Horton, G.J. Kluitenberg, and Q. Wang. 2005. Evaluation of the heat pulse ratio technique for measuring soil water flux. Soil Sci. Soc. Am. J. 69:757–765.

Ochsner, T.E., R. Horton, and T. Ren. 2001. A new perspective on soil thermal properties. Soil Sci. Soc. Am. J. 65:1641–1647.

Ochsner, T.E., R. Horton, and T. Ren. 2003. Use of the dual-probe heat-pulse technique to monitor soil water content in the vadose zone. Vadose Zone J. 2:572–579.

Ochsner, T.E., T.J. Sauer, and R. Horton. 2006. Field tests of the soil heat flux plate method and some alternatives. Agron. J. 98:1005–1014.

Ochsner, T.E., T.J. Sauer, and R. Horton. 2007. Soil heat capacity and heat storage measurements in energy balance studies. Agron. J. 99:311–319.

Ohmura, A., and H. Gilgen. 1993. Re-evaluation of the global energy balance, p. 93–110. *In* Interactions between the global climate subsystems: The legacy of Hann. Geophysical Monograph No. 75. American Geophysical Union, Washington, DC.

Oleson K.W., G.-Y. Niu, Z.-L. Yang, D.M. Lawrence, P.E. Thornton, P.J. Lawrence, R. Stockli, R.E. Dickinson, G.B. Bonan, and S. Levis. 2007. CLM3.5 Documentation. http://www.cgd.ucar.edu/tss/clm/distribution/clm3.5/CLM3_5_documentation.pdf

Olivella, S., J. Carrera, A. Gens, and E.E. Alonso. 1994. Nonisothermal multiphase flow of brine and gas through saline media. Transp. Porous Media 15:271–293.

Parlange, M.B., A.T. Cahill, D.R. Nielsen, J.W. Hopmans, and O. Wendroth. 1998. Review of heat and water movement in field soils. Soil Till. Res. 47:5–10.

Parton, W.J. 1984. Predicting soil temperatures in a shortgrass steppe. Soil Sci. 138:93–101.

Persaud, N., and A.C. Chang. 1984. Analysis of the stochastic component in observed soil profile temperature. Soil Sci. 138:326–334.

Philip, J.R., and D.A. de Vries. 1957. Moisture movement in porous materials under temperature gradients. Trans. Am. Geophys. Union 38:222–232.

Pikul, J.L., Jr., R.R. Allmaras, and S.E. Waldman. 1985. Late season heat flux and water distribution in summer-fallowed Haploxerolls. Soil Sci. Soc. Am. J. 49:1517–1522.

Pruess, K., and J.S.Y. Wang. 1987. Numerical modeling of isothermal and nonisothermal flow in unsaturated fractured rock: A review, p. 11–21. *In* D.D. Evans and T.J. Nicholson (eds.) Flow and transport through unsaturated fractured rock, Geophysical Monograph No. 42. American Geophysical Union, Washington, DC.

Ren, T., G.J. Kluitenberg, and R. Horton. 2000. Determining soil water flux and pore water velocity by a heat pulse technique. Soil Sci. Soc. Am. J. 64:552–560.

Ren, T., K. Noborio, and R. Horton. 1999. Measuring soil water content, electrical conductivity, and thermal properties with a thermo-TDR probe. Soil Sci. Soc. Am. J. 63:450–457.

Richard, G., and P. Cellier. 1998. Effect of tillage on bare soil energy balance and thermal regime: An experimental study. Agronomie 18:163–181.

Ross, P.J., J. Williams, and R.L. McCown. 1985. Soil temperature and the energy balance of vegetative mulch in the semi-arid tropics. II. Dynamic analysis of the total energy balance. Aust. J. Soil Res. 23:515.

Saito, H., J. Simunek, and B.P. Mohanty. 2006. Numerical analysis of coupled water, vapor, and heat transport in the Vadose zone. Vadose Zone J. 5:784–800.

Sauer, T.J. 2002. Heat flux density, Chapter 5-5. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis: Part 1, Physical methods. 3rd edn. ASA, Madison, WI.

Sauer, T.J., J.L. Hatfield, and J.H. Prueger. 1997. Over-winter changes in radiant energy exchange of a corn residue-covered surface. Agric. For. Meteorol. 85:279–287.

Sauer, T.J., J.L. Hatfield, J.H. Prueger, and J.M. Norman. 1998. Surface energy balance of a corn residue-covered field. Agric. For. Meteorol. 89:155–168.

Sauer, T.J., and R. Horton. 2005. Soil heat flux. *In* J.L. Hatfield and J.M. Baker (eds.) Micrometeorology in agricultural systems. ASA monograph. ASA, Madison, WI.

Sauer, T.J., D.W. Meek, T.E. Ochsner, A.R. Harris, and R. Horton. 2003. Errors in heat flux measurement by flux plates of contrasting design and thermal conductivity. Vadose Zone J. 2:580–588.

Sauer, T.J., T.E. Ochsner, and R. Horton. 2007. Soil heat flux plates: Heat flow distortion and thermal contact resistance. Agron. J. 99:304–310.

Scanlon, B.R., and P.C.D. Milly. 1994. Water and heat fluxes in desert soils 2. Numerical simulations. Water Resour. Res. 30:721–733.

Selker, J.S., L. Thevenaz, H. Huwald, A. Mallet, W. Luxemburg, N. van de Giesen, M. Stejskal, J. Zeman, M. Westhoff, and M.B. Parlange. 2006. Distributed fiber-optic temperature sensing for hydrologic systems. Water Resour. Res. 42:W12202.

Shao, M., R. Horton, and D.B. Jaynes. 1998. Analytical solution for one-dimensional heat conduction–convection equation. Soil Sci. Soc. Am. J. 62:123–128.

Shiozawa, S., and G.S. Campbell. 1990. Soil thermal conductivity. Remote Sens. Rev. 5:301–310.

Smerdon, J.E., H.N. Pollack, J.W. Enz, and M.J. Lewis. 2003. Conduction-dominated heat transport of the annual temperature signal in soil. J. Geophys. Res. 108:2431.

Smerdon, J.E., and M. Stieglitz. 2006. Simulating heat transport of harmonic temperature signals in the Earth's shallow subsurface: Lower-boundary sensitivities. Geophys. Res. Lett. 33:L14402. doi:10.1029/2006GL026816.

Soil Survey Staff. 1999. Soil taxonomy. 2nd edn. Agricultural handbook No. 436. USDA-NRCS, Washington, DC.

Sophocleous, M. 1979. Analysis of water and heat flow in unsaturated-saturated porous media. Water Resour. Res. 15:1195–1206.

Steele-Dunne, S.C., M.M. Rutten, D.M. Krzeminska, M. Hausner, S.W. Tyler, J. Selker, T.A. Bogaard, and N.C. van de Giesen. 2010. Feasibility of soil moisture estimation using passive distributed temperature sensing. Water Resour. Res. 46:W03534.

Szeicz, G., C.H.M. van Bavel, and S. Takami. 1973. Stomatal factor in the water use and dry matter production by sorghum. Agric. Meteorol. 12:361–389.

Takakura, T., K.A. Jordan, and L.L. Boyd. 1971. Dynamic simulation of plant growth and environment in the greenhouse. Trans. ASAE 14:964–971.

Tarara, J.M., and J.M. Ham. 1997. Measuring soil water content in the laboratory and field with dual-probe heat-capacity sensors. Agron. J. 89:535–542.

Tarnawski, V.R., and B. Wagner. 1992. A new computerized approach to estimating the thermal properties of unfrozen soils. Can. Geotech. J. 29:714–720.

Uchijima, Z. 1976. Maize and rice, p. 33–64. *In* J.L. Monteith (ed.) Vegetation and the atmosphere. Vol. 2. Academic Press, New York.

Van Bavel, C.H.M., and D.I. Hillel. 1975. A simulation study of soil heat and moisture dynamics as affected by a dry mulch, p. 815–821. Proc. 1975 Summer Computer Simulation Conf., San Francisco, CA. Simulation Councils Inc., La Jolla, CA.

Van Bavel, C.H.M., and D.I. Hillel. 1976. Calculating potential and actual evaporation from a bare soil surface by simulation of concurrent flow of water and heat. Agric. Meteorol. 17:453–476.

van Haneghem, I.A., J. Schenk, and H.P.A. Boshoven. 1983. An improved nonsteady-state probe method for measurement in granular materials. Part II: Experimental results. High Temp. High Press. 15:367–374.

van Wijk, W.R., and D.A. de Vries. 1963. Periodic temperature variations in a homogenous soil, p. 102–143. *In* W.R. van Wijk (ed.) Physics of plant environment. North-Holland Publishing Co., Amsterdam, the Netherlands.

Wang, Q., T.E. Ochsner, and R. Horton. 2002. Mathematical analysis of heat pulse signals for soil water flux determination. Water Resour. Res. 38:1288.

Wierenga, P.J., and C.T. de Wit. 1970. Simulation of heat transfer in soils. Soil Sci. Soc. Am. Proc. 34:845–848.

Wierenga, P.J., R.M. Hagan, and D.R. Nielsen. 1970. Soil temperature profiles during infiltration and redistribution of cool and warm irrigation water. Water Resour. Res. 6:230–238.

Wierenga, P.J., D.R. Nielsen, and R.M. Hagan. 1969. Thermal properties of soil based upon field and laboratory measurements. Soil Sci. Soc. Am. Proc. 33:354–360.

Wiltshire, R.J. 1982. Solutions of the heat conduction equation in a nonuniform soil. Earth Surf. Processes Landforms 7:241–252.

Wiltshire, R.J. 1983. Periodic heat conduction in nonuniform soil. Earth Surf. Processes Landforms 8:547–555.

WMO. 2006. Measurement of temperature. WMO guide to meteorological instruments and methods of observation: WMO No. 8. Draft 7th edn. World Meteorological Organization, Geneva, Switzerland.

Wythers, K.R., W.K. Lauenroth, and J.M. Paruelo. 1999. Bare-soil evaporation under semi-arid field conditions. Soil Sci. Soc. Am. J. 63:1341–1349.

Yamanaka, T., A. Takeda, and J. Shimada. 1998. Evaporation beneath the soil surface: Some observational evidence and numerical experiments. Hydrol. Processes 12:2193–2203.

Soil Spatial Variability

Ole Wendroth
University of Kentucky

Sylvia Koszinski
Leibniz Centre for Agricultural Landscape Research (ZALF)

Vicente Vasquez
University of Kentucky

10.1 Sampling a Soil Process

One of the parameters of soil ecosystems are soil processes (Gregorich et al., 2001). Process is based on the Latin word "procedere—to proceed, to advance, to go on." Each individual stage of a process depends on the previous stage in either a causal (deterministic) manner or a probabilistic (stochastic) manner. A soil process can be considered in space and time. A spatial soil process is the change of a variable or a state vector consisting of several variables across a spatial domain caused by underlying effects. The spatial process of soil water content considered across a landscape can be influenced by spatial changes in soil type, topography, vegetation, rainfall, evapotranspiration, management, etc. Similarly, a temporal process is the change of a soil variable or state vector observed at the same location over time caused by underlying effects. The temporal process of soil water content is influenced by temporal occurrences of precipitation, evapotranspiration, drainage, capillary rise, etc. Hence, subsequent stages in either space or time do not vary randomly but depend on each other. If in fact subsequent stages of a process do not depend on each other but rather vary randomly, the monitoring of a sequence of stages would not allow us to derive any causal or statistical relationships between various ongoing changes. In other words, the spacing or time increments between observations would be inappropriate to identify the process.

Sampling a soil process in either the spatial or the temporal domain is different from what soil scientists usually try to achieve when they aspire a "good" average, or a representative mean value for a plot that received a specific treatment, which assures that the status of this plot is different from the one that received a different treatment. Investigating a spatial or temporal process does not require the a priori installation of two or more treatments but is rather focused on measuring gradual changes with a series of observations of a state variable or an ensemble of state variables, called state vector. Monitoring gradual changes implies that changes in the observed variable are smaller between subsequent and neighboring observations than over a long spatial distance or time interval. If changes over short distances or time intervals are indeed smaller than changes occurring over long intervals, the variation is called structured, and a range of representativity pertaining to an individual measurement can be derived. In this case, we can interpolate a value located between two measured points at an unobserved point. If on the other hand the variance observed over the shortest sampling interval is in the magnitude of the population variance, this variation is called random, and any interpolation between two points would not be justifiable. In this chapter, considerations are given to the sampling of a spatial or temporal process, to derive spatial or temporal relationships between different variables, apply coregionalization, develop simple spatial or temporal time series models, and quantify cyclic variation components or fluctuation patterns within data series. Appropriate soil sampling schemes in which investigators aimed to detect variability structure have been widely discussed in the literature. Burgess and Webster (1980), Warrick and Myers (1987), Brus and de Gruijter (1997), Boyer et al. (1996), Buscaglia and Varco (2003) are a few examples from different ecological disciplines. When applied to soil, agronomic and ecological investigations,

none of the analytical approaches presented here requires an experiment with treatments, but on the other hand, does not exclude treatment experiments. A wide variety of geostatistical applications in treatment experiments exists in the literature (e.g., Bhatti et al., 1991a; Mulla et al., 1992; Ball et al., 1993; Mulla, 1993; Bishop and Lark, 2006). Some of these are focused on the removal of underlying spatial soil trends whose impact was large enough to obscure treatment effects (e.g., Wilkinson et al., 1983; Mulla et al., 1990; Bhatti et al., 1991b). And among many others, Raftery (1993) and Van Es and Van Es (1993) have demonstrated the application of state-space models and semivariography, respectively, in experiments with treatments. On the other hand, beneficial experiments in agriculturally related disciplines, especially soil science, do not have to depend on replicated randomized treatments. As pointed out by Collis-George and Davey (1960), though argued in some respects by Barley (1964), plant–soil interactions may be understood better in a few experiments focused on understanding the biogeochemical and environmental controls on biological response. With the study of landscape processes and the application of statistical analyses that do not depend on spatial or temporal independence, alternative management practices can well be identified even without treatment experiments (Nielsen, 1987). Inherent variations of data series can be analyzed regardless of the size or extent of the spatial or temporal domain. For example, techniques are applicable to large scale investigations such as remote sensing or data observed over a decade as well as to the small scale of an x-ray image of a soil thin section or the change of soil water storage over time increments in the order of seconds or less.

10.2 Autocorrelation

The autocorrelation function is a special expression of autocovariance. The latter describes the average product of differences of two observations, $A_i(x_i)$ and $A_i(x_i + h)$, respectively, from the mean of a series $\overline{A}$, while the two observations are separated by a distance, the so-called lag h (Salas et al., 1988):

$$\text{cov}[A_i(x), A_i(x+h)] = \frac{1}{N} \sum_{i=1}^{N-h} [A_i(x_i) - \overline{A}][A_i(x_i + h) - \overline{A}]. \quad (10.1)$$

Normalizing the autocovariance function yields the autocorrelation function (Shumway, 1988):

$$r(h) = \frac{\text{cov}[A_i(x), A_i(x+h)]}{\sqrt{\text{var}[A_i(x)]}\sqrt{\text{var}[A_i(x+h)]}}, \quad (10.2)$$

where the two variance terms in the denominator describe the variance of both arrays of data $A_i(x_i)$ and $A_i(x_i + h)$. For large time series, the product in the denominator equals the variance s_A^2.

For a data set with structured variation, the autocorrelation coefficient $r(h)$ decreases with increasing separation distance between observations until it reaches zero. A data set without any variability structure exhibits zero autocorrelation for all lag intervals

(Warrick et al., 1986). This condition underlies all randomized sampling designs and associated statistical analyses. Regardless whether a data set manifests structured or random variation, at zero lag, $r(h) = 1$, because an observation is perfectly related with itself. Nielsen and Wendroth (2003) illustrated the calculation scheme for the autocorrelation function in detail. Nielsen and Alemi (1989) determined the autocorrelation function not only for static variables but also for a rate variable, that is, water flow rate.

In view of sampling a spatial soil process, the purpose of this paragraph is to apply autocorrelation analysis for the evaluation of soil water content measurements along a 440 m transect in a northeast German moraine catena. Gravimetric soil water content was measured at 40–60 cm soil depth based on auger samples (Wendroth et al., 2006; Robinson et al., 2008). The spatial resolution of measurements is varied in the analysis of measurements. First, samples taken at 40 m distances equally spaced along the transect are considered (Figure 10.1a). No obvious trend exists in the data. In some cases, fluctuations in soil water content from one location to the next are small—about 0.01 g g^{-1} and in other cases, they are large—greater than 0.03 g g^{-1}.

The resulting autocorrelation function is depicted in Figure 10.1b. The immediate drop of $r(h)$ at $h > 0$ reflects the random or unstructured variation of soil water content measurements. Regardless of the length of the transect over which we examine spatial differences in soil water content, these differences remain random. Because the autocorrelation coefficients are nil, a spatial range of representativity does not exist for this sampling scenario. This result implies that soil water content sampled at 40 m distances cannot be spatially interpolated because no structured variation exists. Moreover, soil water content measurements cannot be applied to explain the variation of any other variable, since the spatial domain represented by an individual water content measurement is unknown.

If the samples taken at 40 m intervals along this 440 m transect become part of a longer series of soil water content measurements, that is, over several kilometers or a hundred kilometer or any distance over which large scale influences on soil water content, such as varying soil type etc., become dominant, a structured variation of soil water content may perhaps be identified (Wendroth and Robinson, 2008). In such a case, the autocorrelation length of water content observations would be a multiple of the 40 m lag interval. However, to quantify a correlation length within the domain of the 440 m transect, samples need to be taken at distances closer than 40 m, or the volume of an individual sample would have to be increased substantially or both.

Reducing the sampling distance to one half, that is, 20 m, results in the water content data shown in Figure 10.1c exhibiting a magnitude of local fluctuations close to the magnitude of the average difference between neighboring observations. The corresponding autocorrelation behavior is illustrated in Figure 10.1d. Still, the spatial variation remains random. Hence, even over distances of 20 m, sampling was not close enough to allow spatial interpolation. When the distance between samples was further reduced to 10 m, a more continuous picture of fluctuations in soil water content can be observed (Figure 10.1e). The former peak water content at ~160 m distance in Figure 10.1a is now
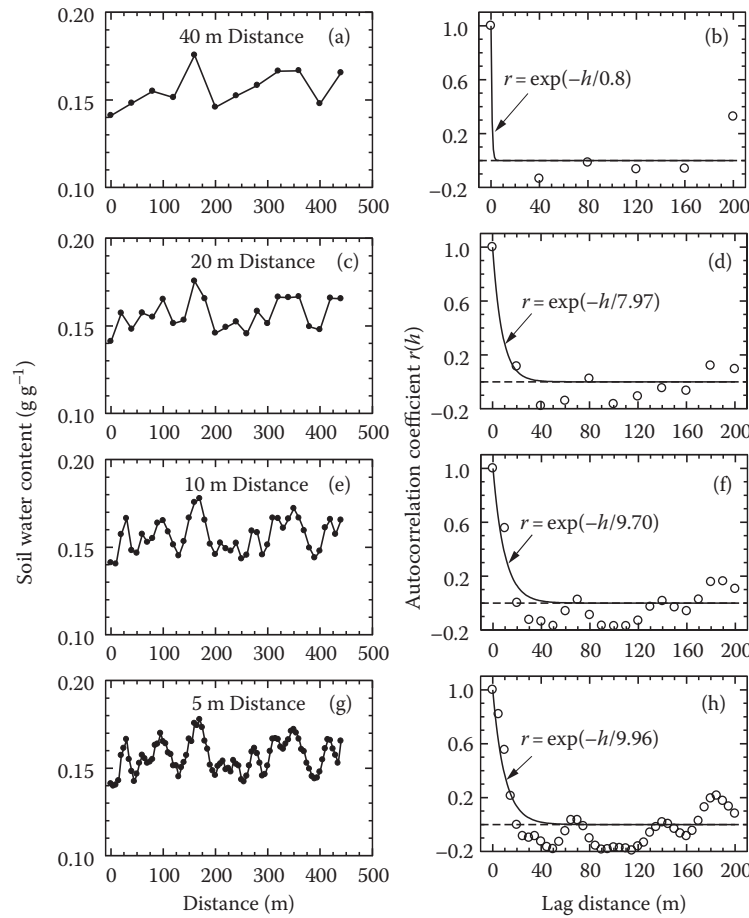
FIGURE 10.1 Soil water content measured at the 40–60 cm soil depth across a moraine agricultural landscape in northeast Germany. Measurements are shown for 40 m (a), 20 m (c), 10 m (e), and 5 m (g) sampling intervals. Respective autocorrelation functions $r(h)$ are displayed in (b), (d), (f), and (h).

supported by several values increasing toward that peak value and others decreasing when that peak location is passed (Figure 10.1e). From Figure 10.1a, the investigator could not tell whether this large value at a distance of 160 m was a measurement error, or in fact, a large value belonging to the continuity of the spatial water content process shown in Figure 10.1e. This locally large value gains reliability and represents the process in the context of previous and subsequent observations. The corresponding autocorrelation function depicted in Figure 10.1f manifests the structure of observed variations to a neighboring distance of 1 lag (10 m). By reducing the sampling distance still further to 5 m, the spatial process of soil water content becomes even more obvious (Figure 10.1g), and the autocorrelation function manifests an obvious variance structure. The autocorrelation functions depicted in Figure 10.1f and h imply that samples taken every 10 or 5 m would clearly support spatial interpolation through kriging or other techniques. Samples taken at intervals greater than 10 m would not allow regionalization, at least not in the limited size of this specifically sampled domain.

This brief introduction to the impact of sampling distance on the identifiability of a spatial process is not only relevant for obtaining insight into the representativity of a soil water content

measurement that was generally pointed out by McBratney and Webster (1983), but spatial structure indicated in the autocorrelation function (Figure 10.1f and h) also qualifies a soil water content observation as a covariate for another state variable. This statement is explained in the following paragraph.

10.3 Cross-Correlation

Once the spatial or temporal variation structure and a range of representativity of a variable are identified, its spatial relationship or cross-correlation behavior to another variable can be quantified. This is accomplished through the cross-correlation function $r_c(h)$, which quantifies the relationship between two variables sampled at the same location or time and at different locations with increasing separation distances. Knowing the spatial or temporal variation of two sets of measurements, we can determine from the cross-correlation function how far apart from each other samples can be taken to yield values that remain related with each other (Cassel et al., 2000). This information is especially important in situations where we cannot place two sensors or take soil samples at exactly the same location. The cross-correlation function is also relevant in situations where

two variables cover different spheres of influence, for example, the investigator wants to know whether a point measurement of soil texture is related to an average crop yield measurement taken over an area of 10 m by 10 m.

The cross-correlation function is defined as the covariance of variable A_i sampled at location x_i and variable B_i sampled at a lag interval apart, that is, at $(x_i + h)$, divided by the square root of the product of the variances of both variables (Shumway, 1988):

$$r_c(h) = \frac{\text{cov}[A_i(x_i), B_i(x_i + h)]}{\sqrt{\text{var}[A_i(x_i)]}\sqrt{\text{var}[B_i(x_i + h)]}} . \qquad (10.3)$$

In the same field where soil water content samples had been taken, soil clay content and geoelectrical resistivity ρ were measured at the same 40–60 cm soil depth. Data are presented in Figure 10.2a for a sampling distance of 40 m. From the spatial process of both data series, an inverse linear relationship is derived with a correlation coefficient of $r = -0.505$, given in Table 10.1. Progressively reducing the spatial sampling distances along the 440 m transect from 40 to 20 m, to 10 m, and to 5 m (Figure 10.2a and b, showing only the 40 and 5 m data) not only changes the correlation coefficient between clay content and ρ (Table 10.1), it also has a strong effect on the intercept and the slope of the linear regression (Table 10.1).
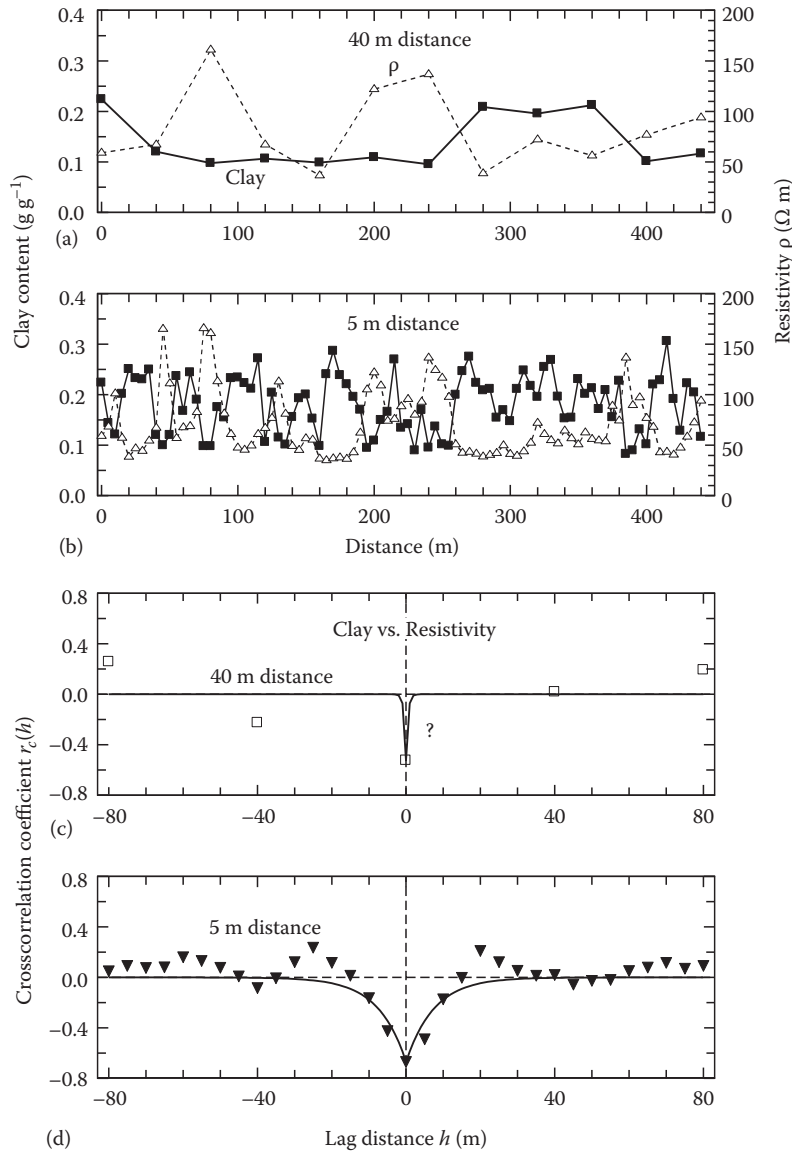
FIGURE 10.2 Soil clay content and geoelectrical resistivity ρ measured in a moraine agricultural landscape in northeast Germany at 40 m (a) and 5 m (b) sampling intervals. The cross-correlation functions $r_c(h)$ corresponding to the two sampling scenarios are shown in (c) and (d), respectively, where the question mark in (c) manifests the unknown cross-correlation length.

TABLE 10.1 Linear Regression Equation Coefficients for the Soil Clay Content—Geoelectrical Resistivity ρ Relationship at 40–60 cm Soil Depth and for Different Spatial Resolutions (Sampling Distances *s*)

Sampling Distance *s* (m)	Intercept *a*	Slope *b*	Correlation Coefficient *r*
40	0.183	-5.98×10^{-4}	-0.505
20	0.224	-9.14×10^{-4}	-0.597
10	0.264	-1.318×10^{-3}	-0.679
5	0.266	-1.223×10^{-3}	-0.669

The cross-correlation behavior between both data sets differs for different spatial sampling distances. For the scenario of 40 m sampling distance, observations of clay content and geoelectrical resistivity sampled at 40 m are not spatially cross-correlated (Figure 10.2c, with its "?" manifesting the unknown cross-correlation length) while those sampled every 5 m are clearly cross-correlated (Figure 10.2d). The latter cross-correlation function indicates that a clay content measurement is related to ρ over a distance of 5 to 10 m but not beyond that distance (Figure 10.2d). Notice that the results for different sampling distances are considered over the same domain or extent. The left-hand side of the cross-correlation function reflects clay content leading the ρ series, and the right-hand side ρ leading the clay content series. Leading refers to the way in which both series are shifted against each other for each lag in cross-correlation analysis (see scheme in Nielsen and Wendroth, 2003), that is, lag $h = -1$ refers to the clay series being one position ahead of the ρ series. Before the spatial relationship between two series is examined in cross-correlation analysis, their individual autocorrelation structure needs to be evident. If one or both series are not autocorrelated, their inclusion in a cross-correlation analysis is ineffective.

Analyses obtained for the same pair of variables sampled at different resolution manifests a classical problem in soil science: When soil scientists obtain a regression relationship for two variables over a large scale, such as a database for a pedotransfer function in order to derive soil hydraulic property parameters based on soil texture, the applicability of that information to estimate spatial processes at a smaller scale or resolution is flawed as long as the small scale spatial covariance and cross-covariance remain unknown, ignored, or not even considered. In other words, because the regression relationships changed from coarse to fine measurement resolution, applying the regression equation obtained from the 40 m scale to a smaller scale is invalid. Moreover, the lack of cross-correlation as obvious from Figure 10.2c prohibits any coregionalization of these two variables. Only if the spatial range is known over which one variable is related with the other, a meaningful interpolation or estimation of one variable based on the other can be made for a smaller scale (Lascano and Hatfield, 1992). It is strongly suggested here to include spatial covariance and cross-covariance in the concept of pedotransfer functions, to clearly define data requirements and identify their representativity, validity, and limitation at various scales.

10.4 Semivariogram

The discussion in this section is focused on 1D soil water content data measured along a transect. Indeed, semivariography is not limited to one dimension. It can also be applied in two or three dimensions with directional anisotropy being quantified and included in the analysis and interpretation. What is being said here is valid for studies and data sets that also extend over more than one domain.

In this paragraph, the semivariogram γ(*h*) that is inversely related to the autocovariance is introduced and discussed. Similar to the autocorrelation function *r*(*h*), the relationship between observations taken at locations separated by a lag distance *h* is quantified. Unlike *r*(*h*), the semivariogram is not a normalized form of covariance but is based on absolute differences in the original units in which the variable was measured. The semivariogram is defined as (Vieira et al., 1983; Isaaks and Srivastava, 1989; Journel and Huijbregts, 1991)

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [A_i(x_i) - A_i(x_i + h)]^2, \qquad (10.4)$$

where the sum of squared differences between observations separated by a common lag distance *h* is averaged and divided by the specific number of pairs of observations *N*(*h*) in each lag class and divided by 2. In other words, the semivariogram γ(*h*) manifests the average variance between pairs of observations within a given lag distance class.

At this point, we note that in the literature, the term semivariogram is frequently used interchangeably with variogram. Oftentimes, book indices refer to "see variogram" when the reader looks for "semivariogram" (e.g., Isaaks and Srivastava, 1989; Deutsch and Journel, 1997) or vice versa (e.g., Nielsen and Wendroth, 2003). Others, such as Davis (1986) use "semivariogram" exclusively. Cressie (1991) mentions the historical wording and usages of the concept and provides a detailed analysis of various equations used to estimate alternative formulations. Here, we only point out a reason for the difference between the two terms that are related to variance and semivariance owing to the manner in which they are calculated. The variance is the average squared deviation of individual observations from the mean, while the semivariance is calculated from Equation 10.4. The semivariance does not consider the deviation from the mean but is based on the average of squared differences between observations that are separated by a given lag distance. When this average squared difference is divided by 2, it becomes comparable to the variance, since the deviation of either one of the two observations from the mean—as typically calculated for the variance—is half as much as their difference from each other (see Nielsen and Wendroth, 2003).

Hence, the magnitude of semivariance at unstructured variation is similar to the population variance. Although the magnitudes of semivariance and variance may be compared with each other, they are nevertheless based on a different squared deviation or difference. Probably for this reason, the *y*-axis of

a semivariogram is never named "variance" but always "semi-variance" in order to avoid confusion and to refer to its specific manner of calculation. From here on, we shall use the terms semivariance and semivariogram.

In cases where observations show structured variation, $\gamma(h)$ increases with increasing lag distance until it reaches a plateau the so-called sill. The lag distance at which the plateau is reached is called the range. In Figure 10.3a, soil water content observations are shown based on gravimetric soil samples that were taken along a 470 m transect every 5 m in the upper 10 cm of a farmer's field soil in Kentucky. The resulting experimental semivariogram is depicted in Figure 10.4a. Notice, except for $\gamma(h)$ at $h = 5$ m, the semivariance values increase with increasing lag distance until the variance reaches a plateau somewhat above 70 m lag distance. With increasing lag distance, the number of

pairs of neighboring observations separated by lag distance h decreases as shown with the open circles in Figure 10.4d.

Typically, we expect the semivariance to increase with increasing lag distance, and at some distance the semivariance reaches a plateau. Bell et al. (1993), who consider the semivariogram an ecologically relevant measure of environmental structure, question this exception and state that environmental variance may further increase with distance so that environments can be regarded neither as uniform at small scales nor as random at large spatial scales.

Various types of semivariogram models, shown in detail in Isaaks and Srivastava (1989), Deutsch and Journel (1997), and Nielsen and Wendroth (2003), can be fitted to experimental semivariograms. Typically, a bounded semivariogram model (Nielsen and Wendroth, 2003) consists of a nugget variance C_0 that is the theoretical intercept at $h = 0$, a structural component C, and the
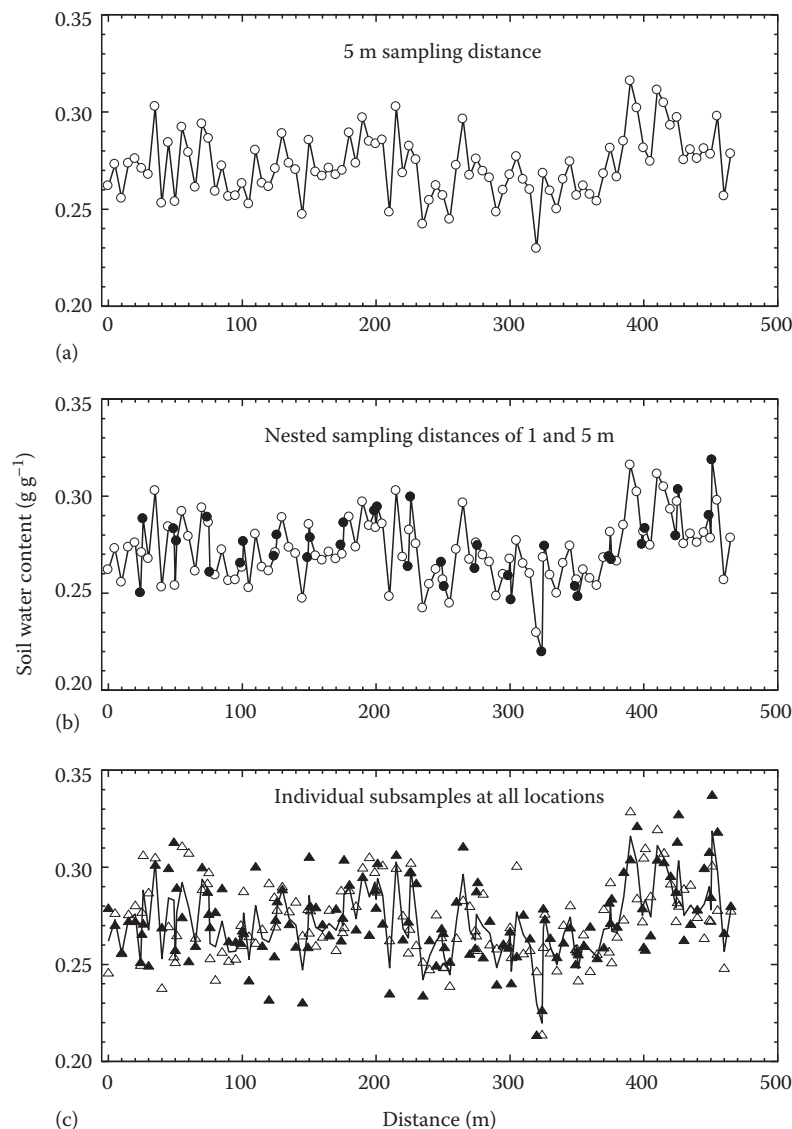
FIGURE 10.3   Soil water content sampled in the upper 10 cm of a silt loam farmer's field soil in Kentucky. Data are shown sampled at every 5 m (a), and in addition at 1 m before and behind the 5 m sampling locations at every 25 m (nested, b). Individual subsamples for all locations are shown in (c) together with the average line from (b).
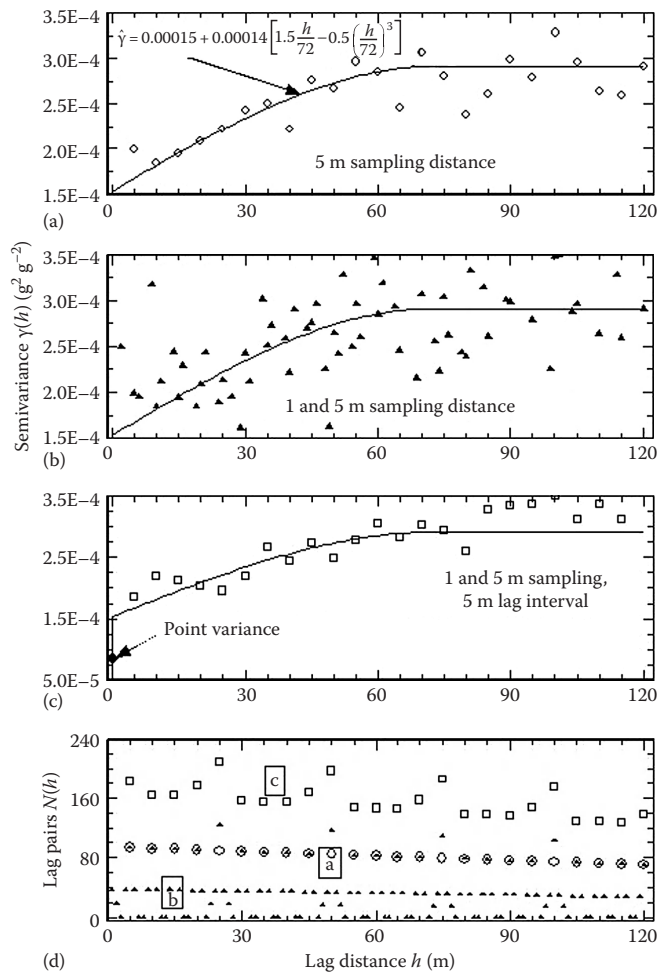
FIGURE 10.4 Spatial variability behavior quantified in semivariograms. Experimental semivariogram and semivariogram model for water content measurements taken at 5 m intervals depicted in Figure 10.3a (a); associated numbers of semivariogram pairs $N(h)$ are shown with open circles in (d). Experimental semivariogram for water content measurements shown in Figure 10.3b at 5 m intervals and at 1 m intervals in nests for 1 m lag intervals (b); associated numbers of semivariogram pairs $N(h)$ are shown with triangles in (d). Experimental semivariogram for water content measurements shown in Figure 10.3b at 5 m intervals and at 1 m intervals in nests but here for 5 m lag intervals (c); associated numbers of semivariogram pairs $N(h)$ are shown with open squares in (d). The filled diamond symbol in (c) refers to the average point variance between duplicate samples taken at each location. Notice the y-axis scale in (c) differing from (a) and (b).

range a, which is the extension of spatial or temporal dependence of observations. The nugget variance represents short-scale variation, sampling, and measurement errors (Isaaks and Srivastava, 1989; Mulla and McBratney, 1999). Nugget and structural semivariance add up to the sill. A spherical semivariogram model was fitted to the experimental semivariogram in Figure 10.4a, where the nugget and sill semivariances are 0.00015 and 0.00014 $g^2\ g^{-2}$, respectively, and the range is 72 m. Most frequently, semivariogram models are used for spatial interpolation using kriging or cokriging. However, semivariogram model parameters provide

the opportunity to compare various data sets with respect to their semivariance behavior, for example, based on the nugget–sill ratio (Cambardella et al., 1994). Notice, when fitting a semivariogram model to an experimental semivariogram for kriging purposes, it is important to obtain a good fit for those lags that carry a substantial weight in the kriging equations (Webster, 1985).

Had the autocorrelation function been calculated for the series of soil water content measurements shown in Figure 10.3a, it would have approximated a vertical mirror image of the semivariogram shown in Figure 10.4a. Unlike the autocorrelation function, the semivariogram displays the semivariance in terms of the square of the original units of measurements. Hence, the investigator is enabled to quantify variance behavior for various separation distances, sampling domain sizes, and sample volumes as well as to compare the semivariance of observations with the variance or uncertainty associated with the measuring device.

Having noted the relatively large nugget semivariance shown in Figure 10.4a to be about 50% of the sill semivariance, the investigator chose a narrower sampling distance to hopefully better define the variability structure at distances <5 m. But it is oftentimes not affordable to regularly sample a long transect at short, for example, 1 m intervals. In this case, every 25 m along the transect two additional soil samples were taken at 1 m distance away from the sampling location within the regular 5 m scheme, as shown with the filled circles in Figure 10.3b. The resulting semivariogram is shown in Figure 10.4b together with the same model as in Figure 10.4a. Since sampling in the nested approach provides 1 m lag intervals in addition to the regular 5 m intervals, $\gamma(h)$ can be calculated for a finer spatial resolution and larger number of lag classes at almost every 1 m interval. Except for multiples of 5 m lag intervals, the number of observed water content pairs present in each lag class is drastically reduced compared to the 5 m sampling scheme to even no pairs of observations separated by specific lag classes, for example, $h = 3, 7, 8, 12, 13$ m, etc. (Figure 10.4d, triangle symbols). The semivariogram depicted in Figure 10.4b is noisier than the one in Figure 10.4a, and its general trend is similar to the one calculated for the 5 m lag intervals. In this case, the nested sampling did not support a more precise quantification of the nugget variance, but confirmed the one obtained from sampling at 5 m distances. One useful conclusion would be that measurements obtained every 5 m could be spatially interpolated or downscaled to 1 m sampling intervals.

Thus, far in this analysis, lag classes were multiples of 1 or 5 m because of either 1 or 5 m distances between sampling locations. Alternatively, a range of lag distances can be included in the same lag class if an upper and a lower bound is defined accordingly. This procedure is especially helpful in designs with nonuniform sampling intervals and for 2D or 3D data sets. Based on the data shown in Figure 10.3b, the semivariogram shown in Figure 10.4c was recalculated for 5 m lag intervals while including all separation distances in-between the 5 m intervals. Hence, the first lag class included all neighboring measurements separated by 1, 2, 4, and 5 m. Increasing the size of lag class intervals is associated with a larger number of pairs of observations per lag class (Figure 10.4d, open square symbols). Again, the same semivariogram

model as in Figure 10.4a and b represents the general trend of the experimental semivariogram with the latter behaving much smoother than in the former cases.

In addition, the investigators evaluated the variance of soil water content at an even smaller scale. Each of the individual samples at every 5 or 1 m, respectively, was composed of two subsamples taken a few centimeters apart from each other. The values of both individual subsamples of soil water content are shown in Figure 10.3c with the same line connecting average water contents at each location as shown in Figure 10.3b. The local point variance existing for a soil water content composed of the two subsamples is illustrated in the semivariogram plot in Figure 10.4c as the filled diamond at $h = 0$. Indeed, this point variance between subsamples is much smaller than the nugget semivariance extrapolated from $\gamma(h)$ at lag distances >1 m. Notice the scale of the ordinate in Figure 10.4c is different than that in Figure 10.4a and b.

A decrease in local semivariance compared to the semivariance prevailing over short distances or from bulk samples is not always observed. Wendroth et al. (2001) and Giebel et al. (2006) found scale-variant behavior of soil mineral nitrogen. Their measurements based on individual small auger samples resulted in a semivariance larger than the nugget semivariance, and only mixing four individual samples taken at the same point and using this bulk sample in spatial analysis allowed identifying a spatial structure of soil mineral nitrogen.

In contrast to the above example, with the exception of taking two or more individual samples at the local scale, had soil water content not been based on gravimetric samples but had been measured with a sensor, the same spatial design and concept to determine variance or error components could have been used. With a sensor used in an access tube or being installed at a fixed point in the soil, readings taken at the same location and depth within sufficiently short-time intervals to exclude temporal soil water content changes could be used to derive the local spatial semivariance of soil water content.

Studying the variance behavior of any state variable, not just that of soil water content, over different distances in space or time, determining the impact of data aggregation on the mean value and its associated variance, and quantifying the instrumental contribution to variance in the way outlined in this example is in general a widely ignored or neglected concept for any type of measurement (Robinson et al., 2008). It is important, however, to make these considerations part of the experimental protocol for determining a spatial or temporal process, for scale transfers of one or several variables, and for spatial or temporal relationships between different variables.

10.5 Kriging

As in the previous section, transect data are used to illustrate the concepts of kriging and cokriging even though these interpolation and coregionalization methods are more frequently applied to 2D and 3D data sets. Kriging is a technique to estimate a value of an observation $A_0^\star(x_0)$ at an unsampled location based on values of the same variable $A_i(x_i)$ observed in the neighborhood of

the unsampled point (Vieira et al., 1981; Trangmar et al., 1985). Based on the spatial relationship between neighboring observations and the distance of these neighboring locations relative to the unsampled location, a weight λ_i is assigned to each of the neighboring observations to contribute to the estimate. The estimated value is the weighted sum of neighboring observations:

$$A_0^\star(x_0) = \sum_{i=1}^n \lambda_i A_i(x_i), \qquad (10.5)$$

where

$$\sum_{i=1}^n \lambda_i = 1. \qquad (10.6)$$

In the kriging algorithm, the weight coefficients are calculated while the difference between the estimated and the true value is zero and the variance between both is at a minimum through

$$\sum_{j=1}^n \lambda_j \gamma(x_i - x_j) + \mu = \gamma(x_i - x_0), \qquad (10.7)$$

where μ is the Lagrangian multiplier. The semivariance values for the respective distance between sampling locations x_i and y_i and between sampling locations and the unsampled location x_i and x_0, respectively, are obtained from a semivariogram model, which is the line fitted to the experimental semivariogram. The minimized kriging variance is calculated with

$$s_k^2(x_0) = \sum_{j=1}^n \lambda_j \gamma(x_j - x_0) + \mu. \qquad (10.8)$$

The set of equations in the kriging matrix is solved for obtaining the weights λ_i and the Lagrangian multiplier μ. Kriging is called the best linear unbiased estimator. Illustrative examples for the setup of the kriging matrix are given in Warrick et al. (1986), Yates and Warrick (2002), and Nielsen and Wendroth (2003).

Kriging, generally used as an interpolation tool, effectively smoothes the process between observed points based on the observed variability structure and its spatial extent. Hence, the kriged distribution of a soil property or a univariate kriged map can support visual assessment of spatial relationships between the variable of interest and any other underlying soil process. On the other hand, unlike other interpolation techniques, such as inverse distances, kriging relies on a spatial variability structure between observations. If that structure is not accurately derived in the semivariogram, subsequent spatial interpolation cannot be trusted. If the method of inverse distances being less rigid than kriging is used in absence of spatial structure, it can lead to fallacious estimates and misleading maps of spatial distributions (Gotway et al., 1996; Kravchenko, 2003) because of lack of underlying variability structure.

Yost et al. (1982a, 1982b) were among the first who derived spatial variability structure of phosphate sorption distribution across
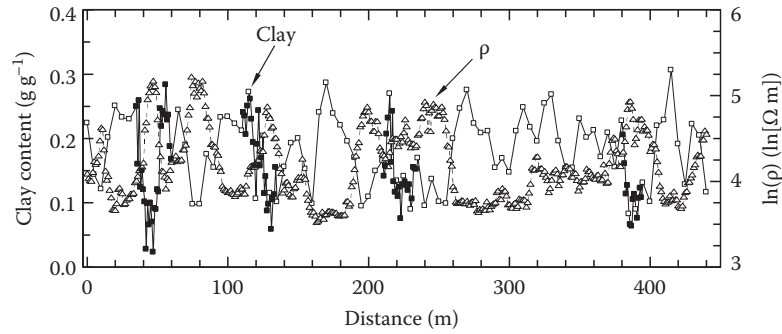
FIGURE 10.5 Soil clay content and geoelectrical resistivity ρ measured along a 440 m transect in a northeast German moraine catena at the 40–60 cm soil depth. Open square symbols depict clay content measured at every 5 m and closed square symbols measured clay content within four nests at 1 m sampling distance. The measurements in the nests were disregarded for the kriging and cokriging estimation scenarios.

a large area and then interpolated unsampled locations based on spatial variability structure in a kriging procedure. Liebhold et al. (1991) mapped moth insect populations based on kriging estimates of spatial distributions. Goovaerts and Chiang (1993) mapped field nitrogen status based on factorial kriging analysis. In a similar fashion, Ghidey and Alberts (1999) mapped soil nitrate concentration based on topography, depth to claypan, and several other underlying variables. Sampling fields without randomized treatments, Donald (1994) mapped weed distributions across landscapes using kriging and pointed out opportunities for new research hypotheses on weed biology and control across landscapes. When applying geostatistical analysis and kriging to attributes that vary over orders of magnitude, indicator kriging can be applied in which spatial distributions of the probability of

an attribute to be below a certain threshold value are interpolated (e.g., Ritzi, 2000). In other applications of indicator kriging, Smith et al. (1993) and Halvorson et al. (1995) combined various indices to map soil quality. Recently, in several approaches, kriging was applied not only in a spatial domain but simultaneously in a spatio-temporal extent (Jost et al., 2005; Cichota et al., 2006).

The soil clay content data considered above from a different perspective (Figure 10.2) are used here again for spatial interpolation by kriging with a slight difference. Besides observations at regular 5 m intervals (open square symbols in Figure 10.5), clay content is known also at 1 m intervals within four nests (closed square symbols in Figure 10.5).

In Figure 10.6a, the experimental semivariogram and the fitted model indicate a range of spatial dependence of 10 m.

$$\hat{\gamma} = 0.0008 + 0.0029 \left[1.5 \frac{h}{10} - 0.5 \left(\frac{h}{10} \right)^3 \right]$$

(a)

(b)
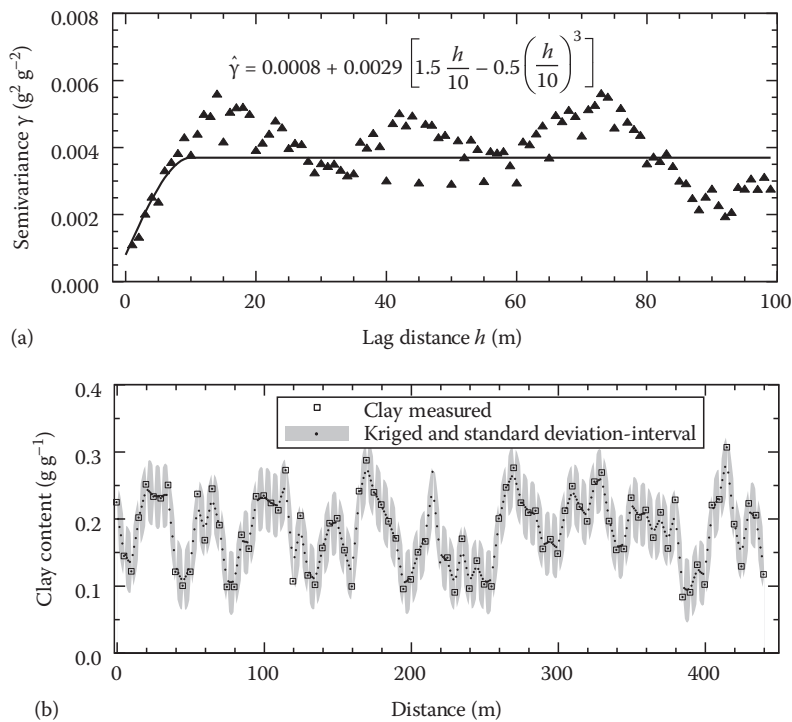
FIGURE 10.6 Kriging scenario: Experimental semivariogram and semivariogram model (nugget + spherical) for soil clay content at 40–60 cm depth (a). Clay content measured at 5 m distances and kriged for 1 m distances (b) along a 440 m transect in a northeast German catena.

The sill proceeds in the order of magnitude of the total variance. It could be argued that the range should extend to a lag distance h of ~15 m, and the sill be somewhat higher. However, in the kriging matrix, only the closest neighbors to the location of estimate are germane. Observations farther than 10 m away from the location of interest do not make a sizeable contribution to the kriged estimate.

The kriging result is depicted in Figure 10.6b. Clay content observations at 5 m intervals were the only information provided to estimate the clay content at 1 m intervals. The filled circles denote the kriged estimates and the gray shaded area the ±1 standard deviation interval around the kriged estimate. As expected, the smooth process of kriged estimates between sampling locations is noticeable. Moreover, because the kriging standard deviation interval increases with increasing distances from sampling locations, it manifests relative maximum values midway between two measured points. We learn from this example that kriging merely interpolates values at unsampled locations based on measurements at surrounding locations. Depending on the change between the observed values—increase or decrease or no change—kriging estimates between two observed values either consistently increase or decrease or do not change at all, respectively, but do not reveal fluctuations at unsampled locations that might in fact exist at a scale smaller than the one chosen for the sampling.

10.6 Cokriging

In many cases of soil surveys and environmental contamination studies, an attempt is made to estimate soil textural or contaminant distribution based on the distribution of another variable that is expected to indirectly reflect the distribution of a particular textural fraction, but is easier and cheaper to monitor. In case of soil clay content, geoelectrical resistivity can be applied as a proxy variable because both variables are spatially related. In spatial studies of contaminants, the distribution of soil pH can help to estimate the distribution of heavy metals in an area (Van Groenigen et al., 1997). McBratney and Webster (1983) cokriged various soil textural particle size fractions among different depths with each other. In various other applications, existing soil landscape models were improved based on using auxiliary variables especially in cases with limited observations (e.g., Vauclin et al., 1983; Alemi et al., 1988; Ben-Jemaa et al., 1994; Chaplot et al., 2000). In many cases of environmental studies in which a variable of interest is coregionalized using cokriging, not all variables are collected at each location, causing only a limited number of common sampling locations contributing to the cross-variogram. In this case, pseudo-cross-semivariograms are a compromise to overcome limited data support for variograms (Zhang et al., 1992, 1999).

The theory and setup of the cokriging matrix are similar to that of the kriging algorithm. Cokriging requires semivariogram models for both individual variables involved in the estimation and the cross-semivariogram $\gamma_{AB}(h)$, which manifests the variance behavior that both variables A_i and B_i have in common:

$$\gamma_{AB}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [A_i(x_i) - A_i(x_i + h)][B_i(x_i) - B_i(x_i + h)].$$

(10.9)

In general, the same types of models usually fitted to semivariograms can also be used for describing the cross-semivariogram. The cokriged estimate of a variable $A_0^\star(x_0)$ at an unsampled location is obtained from observations at known locations of both variables $A_i(x_i)$ and $B_j(x_j)$ by

$$A_0^\star(x_0) = \sum_{i=1}^{n} \lambda_{Ai} A_i(x_i) + \sum_{j=1}^{m} \lambda_{Bj} B_j(x_j).$$

(10.10)

The weighting factors for both variables are determined to meet the conditions for A_i

$$\sum_{i=1}^{n} \lambda_{Ai} = 1$$

(10.11)

and for B_j

$$\sum_{j=1}^{m} \lambda_{Bj} = 0.$$

(10.12)

The cokriging matrix is designed as (Yates and Warrick, 2002)

$$\gamma_A(x_{Ai} - x_0) = \sum_{i=1}^{n} \lambda_{Ai} \gamma_A(x_{Ai} - x_A) + \sum_{j=1}^{m} \lambda_{Bj} \gamma_{AB}(x_{Aj} - x_B) + \mu_A$$

$$\gamma_{AB}(x_{Bi} - x_0) = \sum_{i=1}^{n} \lambda_{Ai} \gamma_{AB}(x_{Bi} - x_A) + \sum_{j=1}^{m} \lambda_{Bj} \gamma_{BB}(x_{Bj} - x_B) + \mu_B.$$

(10.13)

The cokriging variance is

$$\sigma_{ck}^2 = \gamma_A(0) + \mu_A - \sum_{i=1}^{n} \lambda_{Ai} \gamma_A(x_{Ai} - x_0) - \sum_{j=1}^{m} \lambda_{Bj} \gamma_{AB}(x_{Bj} - x_0).$$

(10.14)

When fitting semivariogram models with common covariance structure to the two experimental semivariograms γ_A and γ_B, and the cross-semivariogram γ_{AB},

$$\begin{bmatrix} \gamma_{A,1}(h) & \gamma_{AB,1}(h) \\ \gamma_{BA,1}(h) & \gamma_{B,1}(h) \end{bmatrix} = \begin{bmatrix} c_{A,1} & c_{AB,1} \\ c_{BA,1} & c_{B,1} \end{bmatrix} \begin{bmatrix} \gamma_1(h) & 0 \\ 0 & \gamma_1(h) \end{bmatrix}$$

(10.15a)

$$\begin{bmatrix} \gamma_{A,2}(h) & \gamma_{AB,2}(h) \\ \gamma_{BA,2}(h) & \gamma_{B,2}(h) \end{bmatrix} = \begin{bmatrix} c_{A,2} & c_{AB,2} \\ c_{BA,2} & c_{B,2} \end{bmatrix} \begin{bmatrix} \gamma_2(h) & 0 \\ 0 & \gamma_2(h) \end{bmatrix}$$

$$(10.15b)$$

$$\begin{bmatrix} \gamma_{A,3}(h) & \gamma_{AB,3}(h) \\ \gamma_{BA,3}(h) & \gamma_{B,3}(h) \end{bmatrix} = \begin{bmatrix} c_{A,3} & c_{AB,3} \\ c_{BA,3} & c_{B,3} \end{bmatrix} \begin{bmatrix} \gamma_3(h) & 0 \\ 0 & \gamma_3(h) \end{bmatrix}$$

$$(10.15c)$$

the following restrictions apply for the linear model of coregionalization remains positive definite (Isaaks and Srivastava, 1989):

$$c_{Ai} > 0 \quad \text{and} \quad c_{Bi} > 0 \quad \text{for } i = 1,2,\ldots,m \qquad (10.16a)$$

$$c_{Ai}c_{Bi} > c_{ABi}c_{BAi} \quad \text{for } i = 1,2,\ldots,m. \qquad (10.16b)$$

In fact, this restriction causes simultaneous semivariogram and cross-semivariogram model fitting to become a major challenge in the cokriging procedure. In Figure 10.7, the two experimental semivariograms and the common cross-semivariogram for clay content and geoelectrical resistivity $\ln(\rho)$ are shown. Unlike demonstrated in Figure 10.6, where a spherical semivariogram model for soil clay content was derived for kriging, a common
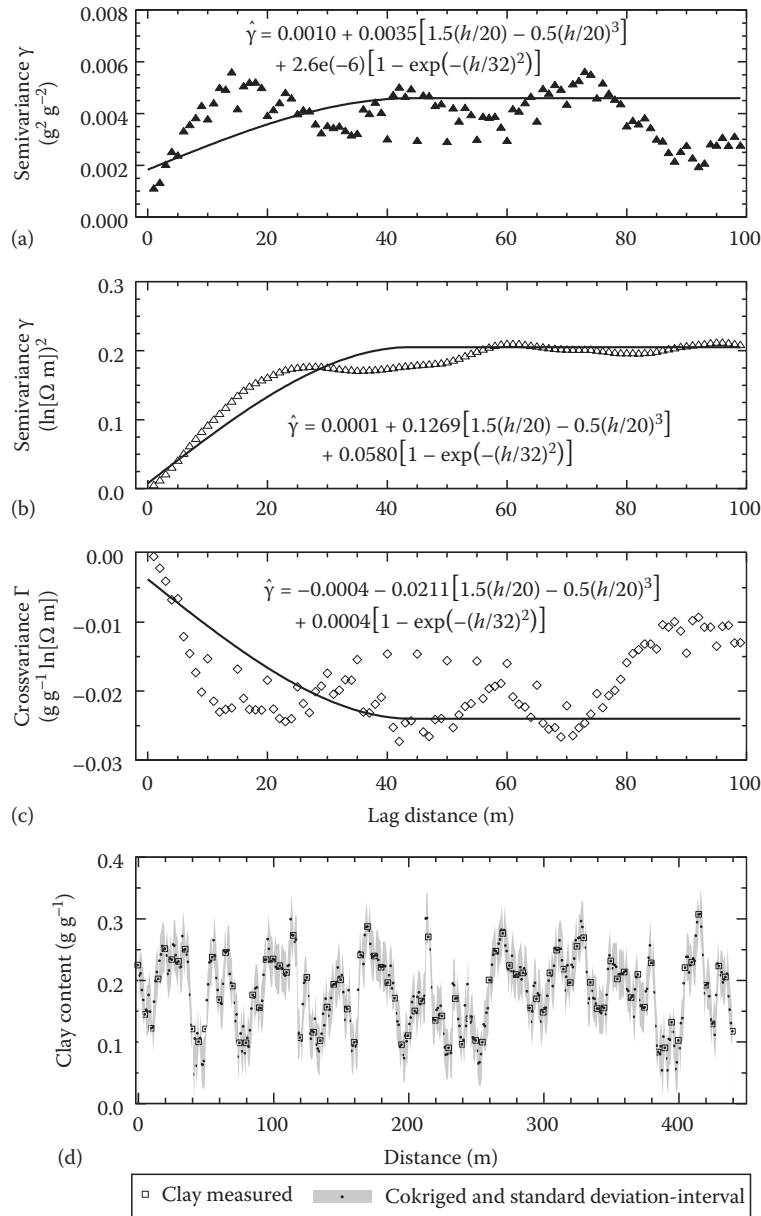
FIGURE 10.7 Cokriging scenario: Experimental semivariogram and common linear model of coregionalization (nugget + spherical + Gaussian) for soil clay content (a), geoelectrical resistivity (b) and cross-semivariogram of both variables (c) measured in a 440 m transect in a northeast German catena. Measured (5 m) and cokriged (1 m) clay content (d).

spherical semivariogram model for clay content and $\ln(\rho)$ that would adequately represent the variability structure of both variables at least for short lag intervals was not identifiable. Instead, a combination of a spherical with a Gaussian semivariogram model consisting of a nugget

$$\begin{bmatrix} c_{A,1} & c_{AB,1} \\ c_{BA,1} & c_{B,1} \end{bmatrix} = \begin{bmatrix} 0.00100 & -0.00037 \\ -0.00037 & 0.00014 \end{bmatrix}, \qquad (10.17a)$$

a spherical component with a spatial range $a_s = 20\,\text{m}$

$$\begin{bmatrix} c_{A,2} & c_{AB,2} \\ c_{BA,2} & c_{B,2} \end{bmatrix} = \begin{bmatrix} 0.00351 & -0.02110 \\ -0.02110 & 0.12687 \end{bmatrix}, \qquad (10.17b)$$

and a Gaussian component with a range $a_G = 32\,\text{m}$

$$\begin{bmatrix} c_{A,3} & c_{AB,3} \\ c_{BA,3} & c_{B,3} \end{bmatrix} = \begin{bmatrix} 2.6e^{-6} & 0.00039 \\ 0.00039 & 0.05797 \end{bmatrix} \qquad (10.17c)$$

resulted in the best common fit of the experimental covariance behavior. The complete set of semivariogram and cross-semivariogram model equations is given in Figure 10.7a through c. These models were applied together with clay content measured at every 5 m (Figure 10.7d) and $\ln(\rho)$ at every 1 m to estimate clay content. Hence, with regard to clay content input data the estimation scenario is the same as in the above kriging procedure (Figure 10.6b), but in addition to clay content, $\ln(\rho)$

and the common covariance behavior between both variables are taken into account in this coregionalization. In general, the standard deviation of the estimated clay content is ~25% smaller in the cokriging compared to the kriging scenario. Unlike kriging, which mainly resulted in a smoothed process of clay content between known locations, fluctuations over short distances become evident in the cokriging results based on the second variable, $\ln(\rho)$, involved in the estimation (Figure 10.7d).

Besides the general process of the series along the 440 m transect estimated with kriging or cokriging, the result is especially interesting with respect to the clay estimation within the four nests, where clay content at every 1 m was known but not incorporated in the estimation procedure. How well did $\ln(\rho)$ capture the clay content process compared to actual measurements? This comparison is illustrated for the first nest, that is, a section of the transect between 35 and 60 m in Figure 10.8.

The smooth process of kriged clay content and the relatively large kriging standard deviation are depicted in Figure 10.8a together with measured clay content. On the other hand, cokriging resulted in pronounced fluctuations of estimated clay content. This did not necessarily improve the agreement between measured and estimated clay content from 35 to 39 m, but in most cases between 42 and 49 m, and definitely between 51 and 59 m. In summary, $\ln(\rho)$ and its spatial relationship with clay content helped to improve the clay estimation compared to univariate interpolation, that is, kriging. However, at the finest resolution of 1 m being investigated here, fluctuations in measured clay content do not perfectly coincide with fluctuations of $\ln(\rho)$. This result can be explained by the different domain of influence of an individual measurement of both variables. Clay
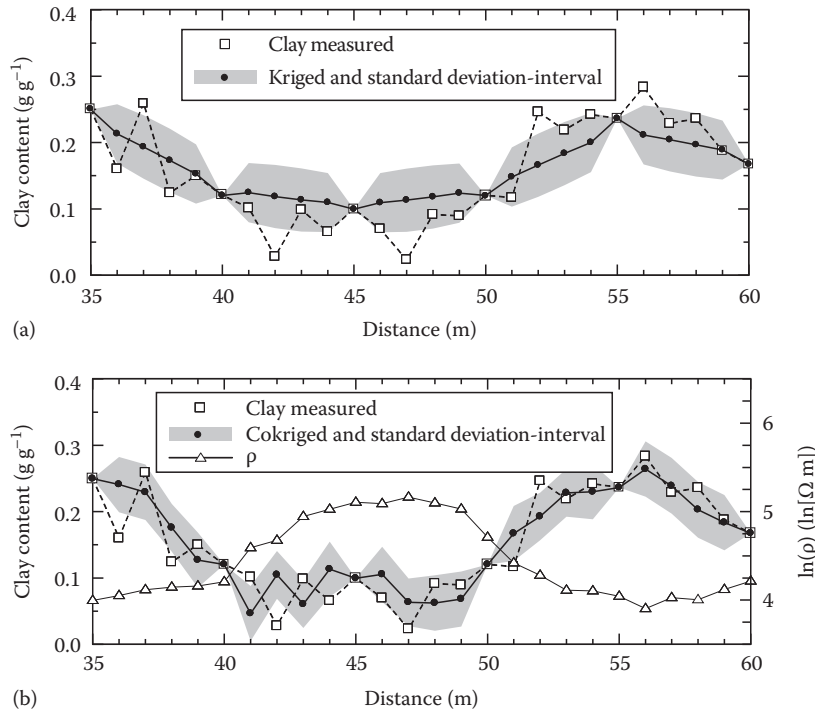
FIGURE 10.8 Detailed results of kriging and cokriging scenarios between 35 and 60 m distance depicted in Figures 10.6b and 10.7d.

content samples were pulled with an auger 2 cm in diameter at a depth increment of 20 cm. A single value of geoelectrical resistivity results from the field measurement of apparent geoelectrical resistivity undergoing a numerical inversion, and represents a horizontal domain of ~50 cm, causing the smooth behavior of the ln(ρ) series. Despite lack of sensitivity of clay content fluctuations occurring over 1 m distance, clay content measurements taken every 5 m were successfully downscaled in this coregionalization procedure based on small scale measurements of ln(ρ).

10.7 Autoregressive State-Space Modeling

Requirements for identifying a spatial or temporal process were presented in the previous paragraphs. It became obvious that understanding the spatial process of an observed state variable whose spatial covariance structure is successfully identified strongly depends on its association with other spatial processes. In other words, the question that needs to be addressed is: What other variables coincide with the process of interest across a spatial or temporal domain and how can the association between the variables included in the state vector be described?

Coregionalization using cokriging is a method that is restricted to a common covariance model, that is, a set of semivariograms with common covariance structure. Spatial association between soil variables observed along transects has been derived from autoregressive analysis (Morkoc et al., 1985; Shumway, 1985), a technique that was adapted from time series analysis for spatial data series. Lebron et al. (2007) described the spatial process of soil hydraulic conductivity across a Pinyon-Juniper ecohydrological gradient. In autoregressive models, a variable at a location or a point in time is described based on antecedent observations of the same variable and in case of multivariate autoregressive models on other related variables observed at the previous sampling location (e.g., Wendroth et al., 1992; Nielsen et al., 1994). For a first-order autoregression model, each of the variables measured at the previous location $i - 1$ is linked to the variable at location i through an autoregression coefficient φ or a matrix of autoregression coefficients in case of several variables. A second-order model utilizes antecedent measurements at locations $i - 1$ and $i - 2$, while an n-order model is based on measurements at n previous locations. Coefficients of autoregression are usually estimated in the same way as ordinary statistical regression coefficients. In the particular case of autoregressive state-space models, however, autoregression coefficients are derived in a different way including estimations of measurement and model errors. The set of observations of one or several variables, that is, the observed vector Y_i, does not have to be assumed absolutely correct but consists of the true status of observed variables Z_i and a measurement error υ_i in the observation equation

$$Y_i = M_i Z_i + \upsilon_i, \qquad (10.18)$$

where M_i is a measurement or design matrix.

In the state equation, that is, in this case an autoregression type model, the state vector at location i is related to the state vector at $i - 1$ through the matrix of autoregression coefficients Φ and a model uncertainty ω_i with

$$Z_i = \Phi Z_{i-1} + \omega_i. \qquad (10.19)$$

The advantage of the state-space form of autoregressive models compared to ordinary autoregressive models is the filtering of observational noise, and the fact that the chosen model type and order and the observation vector result in model uncertainty. In other words, the model does not completely capture the spatial or temporal process of the variable of interest. The simultaneous optimization of autoregression coefficients, measurement, and model variance used here is based on the Kalman filter (Kalman, 1960; Kalman and Bucy, 1961) and encoded in the EM algorithm described by Shumway and Stoffer (1982). In the filtering scheme, state estimations and state variance are updated whenever in the spatial or temporal series an observation is available (Katul et al., 1993). This update depends on the model and measurement error and is performed based on a Kalman gain matrix and a recursive smoothing matrix. The filtering scheme is described in more detail in Shumway (1988), Shumway and Stoffer (2000), Katul et al. (1993), Nielsen and Wendroth (2003), and in Webster and Heuvelink (2006). Not only autoregressive but other kinds of equations can be implemented in state-space models. For example, Katul et al. (1993) and Wendroth et al. (1993) used a state-space scheme in connection with a physically based equation describing soil water transport, to estimate unsaturated hydraulic conductivity coefficients from an internal drainage experiment. Parlange et al. (1993) applied a similar approach to estimate soil water diffusivity. Assouline (1993) modeled the hydrologic budget of Lake Kinneret based on state-space analysis of individual water balance components, Skaggs and Mohanty (1998) applied the Kalman filter scheme to describe water table fluctuations, Cahill et al. (1999) estimated field-effective soil hydraulic property coefficients using a physically based state-space model, and Wu et al. (2001) developed a state-space scheme for modeling the water and salinity regime in a soil with a shallow groundwater table. Bierkens et al. (2001) described the space-time development of water table depth based on a Kalman filter approach, and Walker et al. (2002) estimated 3D soil water storage based on surface soil water content measurements implemented in a state-space model. For further details on the state-space concept used in connection with physically based models, the reader is referred to Nielsen and Wendroth (2003).

Autoregressive state-space analysis is used here to analyze measured solute concentration data in a field leaching experiment conducted in a Maury silt loam soil in Kentucky. Solute concentration was measured in two subsequent depths (20–30 and 30–40 cm) along a 16 m transect at 25 cm intervals. Logarithmic transforms of solute concentration measured in both depths are shown in Figure 10.9a. Prior to autoregressive state-space analysis, spatial covariance structure of both variables was quantified
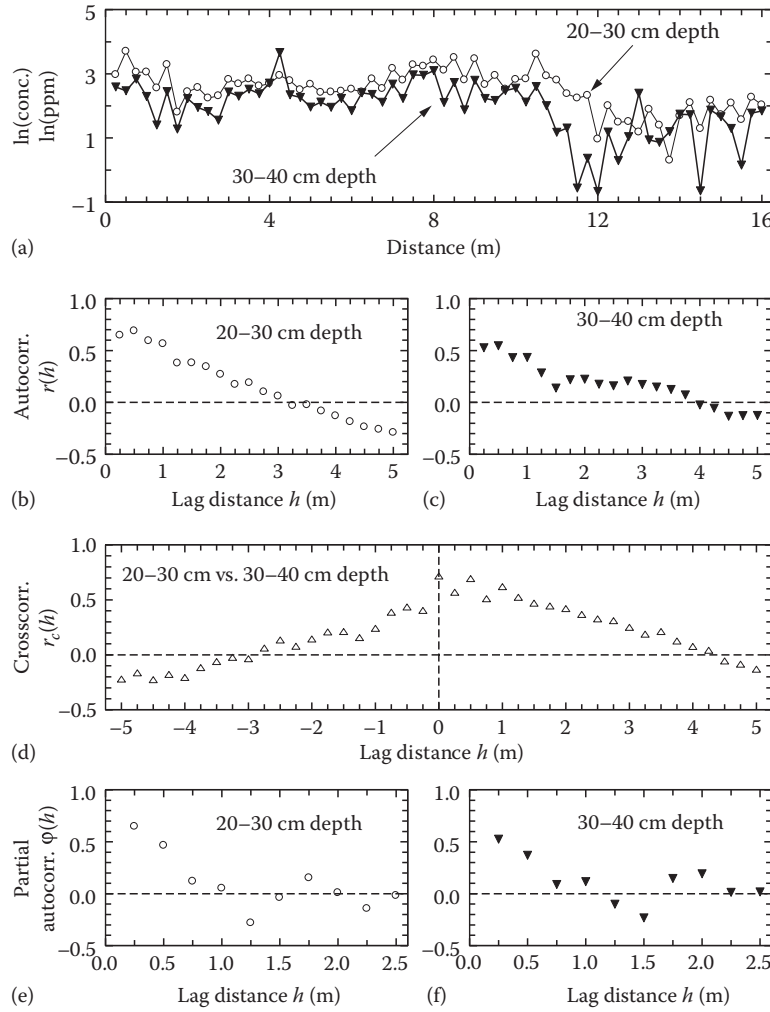
FIGURE 10.9 Ion concentration measured at 20–30 and 30–40 cm soil depth in a Kentuckian Maury silt loam (a), autocorrelation functions $r(h)$ at 20–30 cm (b) and 30–40 cm soil depth (c), cross-correlation function $r_c(h)$ between ion concentrations at both depths (d), and partial autocorrelation functions $\varphi(h)$ for the 20–30 cm (e), and the 30–40 cm depth (f).

via auto- and cross-correlation. With both series being spatially autocorrelated (Figure 10.9b and c) and cross-correlated (Figure 10.9d), measurements of solute concentration based on 10 cm long soil samples, ~2 cm in diameter taken every 25 cm provide ample opportunity to monitor the spatial process of solute concentration in both soil layers.

In order to identify the model order of the autoregressive state-space model, the partial autocorrelation function φ_{ii} of order p is calculated with the Durbin–Levinson algorithm (Shumway and Stoffer, 2000)

$$\varphi_{ii} = \frac{r(i) - \sum_{j=1}^{i} \varphi_{pj} r(i-j)}{1 - \sum_{j=1}^{i} \varphi_{pj} r(j)}, \quad i > 1 \tag{10.20}$$

and

$$\varphi_{11} = r(1). \tag{10.21}$$

For $i \geq 2$,

$$\varphi_{ij} = \varphi_{i-1,j} - \varphi_{ii} \varphi_{i-1,i-j}, \quad j = 1, 2, \dots, n-1. \tag{10.22}$$

The resulting partial autocorrelation functions φ_{ii} where the order is expressed as lag distance with $\varphi(h)$ for the solute concentration data of both depths are depicted in Figure 10.9e and f. With the 95% significance level at 0.245, solute concentration data series can be described with a model of the order up to $p = 2$, that is, solute concentration in both layers at location i can be described as a function of solute concentration in both layers at preceding locations $i - 1$ and $i - 2$.

For numerical stability, prior to state-space analysis, observations y_i are normalized (ysc_i) with (Nielsen and Wendroth, 2003)

$$ysc_i = \frac{y_i - (\mu_y - 2\sigma_y)}{4\sigma_y}, \tag{10.23}$$

where μ_y and σ_y denote the mean and standard deviation, respectively, of the particular series.
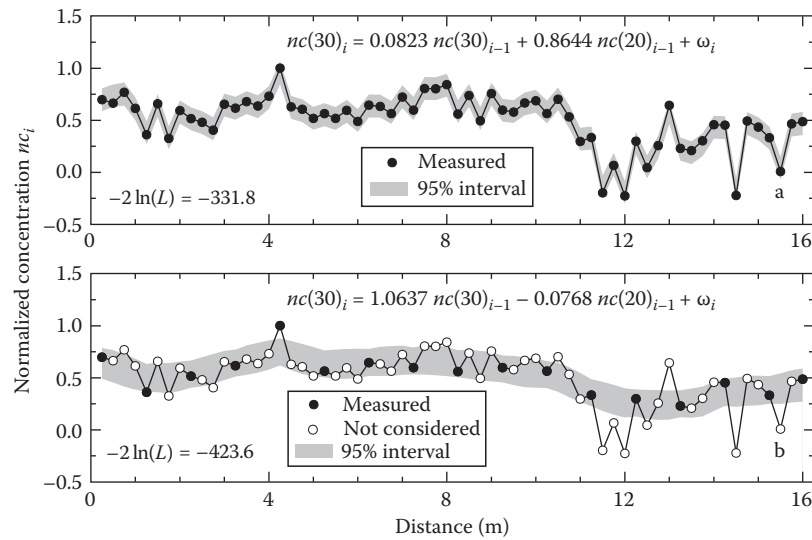
FIGURE 10.10 First-order autoregressive state-space model of normalized ion concentration at the 30–40 cm soil depth $nc(30)_i$ along a 16 m transect in a Kentucky Maury silt loam soil. Ion concentrations are shown for the 30–40 cm depth [$nc(30)_i$], and concentrations of ions in the 20–30 cm depth [$nc(20)_i$] are included in the estimation. A scenario where all observations are considered is shown in (a) and a scenario where all $nc(20)_i$ observations are included but only every fourth $nc(30)_i$ observation is considered in (b).

Model equations and results for the process of normalized solute concentration at the 30–40 cm layer $nc(30)_i$ are presented in Figure 10.10a for a first-order model in a scenario, where all normalized observations of $nc(30)_i$ and those in the 20–30 cm layer $nc(20)_i$ are included in the model calculations.

Notice that autoregressive models including state-space models depend on equally spaced distances between observations. On the other hand, samples do not need to be taken at each location. Scenarios where one or several variables are observed at each location and one or several other variables at a lower density at, for example, every other location or even less frequent are common. For these missing locations, estimates are obtained in the state-space procedure as shown in the following example.

When only every fourth observation of $nc(30)_i$ is included in the model calculations, the estimated series proceeds smoother (Figure 10.10b) and the 95% interval of estimation is larger than in the previous scenario. Although the value measured at a distance of 4.25 m was considered in the calculations, its extremely high fluctuation was not captured in the model. The second-order model shown in Figure 10.11 results in a better representation of the point-to-point fluctuations than the first-order model. Additionally, the confidence interval of estimation remains narrower for the second-order model compared to that of the first order. For both second-order scenarios, the coefficients in the autoregression equations clearly emphasize the larger weight on $i − 2$ neighboring observations compared to the $i − 1$ neighbors. In both scenarios, the log likelihood [$−2\ln(L)$] indicates a better model description for the models of second rather than of first order.

10.8 Spectral Analysis

Managing soils and crops with machinery in spatially regular paths may cause repetitive or cyclic patterns in the direction perpendicular to the management operation. Moreover, observing ecologically relevant soil state variables over time often results in sinusoidal patterns of variation, for example, the daily or the annual soil temperature cycle, the annual water cycle, and related variables such as microbial respiration, crop water use, biomass production, transpiration, drainage fluxes, etc. Such a cyclic fluctuation of observations can be described with a sine or cosine wave function or a combination of both. Each wave can be characterized by its magnitude in signal, that is, the amplitude. The amplitude is the distance between the baseline and the maximum of the wave, or it is half of the difference between maximum and minimum of the wave. Furthermore, the length of the wave λ is an important characteristic, which is the inverse of its frequency f. The number of cycles or waves per basic time or spatial distance interval is another important measure to describe a cyclic-data series. Moreover, the start point of a wave is flexible relative to another wave or another series. This shift of the starting point is called phase. Relative to each other, two series can be in phase or out of phase. For further details, see Davis (1986).

From Jean Baptiste Fourier we learn that every continuous single-valued function can be represented by a series of sinusoids. In other words, any observed series can be considered as a combination of many independent sinusoidal curves, all with different cyclic characteristics. Spectral analysis according to Fourier allows filtering the different waves and their characteristics that contribute to the observed series, similar to a laboratory analyzer that identifies the compounds of a mixed chemical. Similar to a
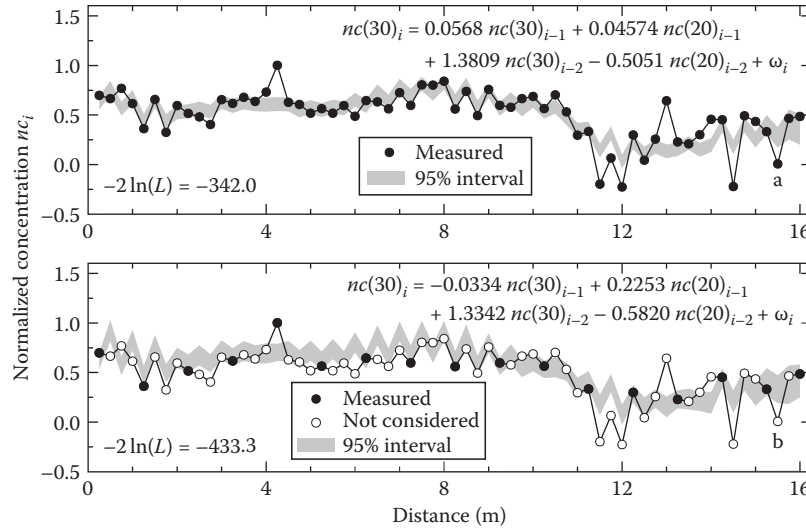
FIGURE 10.11 Second-order autoregressive state-space model of normalized ion concentration at the 30–40 cm soil depth $nc(30)_i$ along a 16 m transect in a Kentucky Maury silt loam soil. Ion concentrations are shown for the 30–40 cm depth [$nc(30)_i$], and concentrations of ions in the 20–30 cm depth [$nc(20)_i$] are included in the estimation. A scenario where all observations are considered is shown in (a) and a scenario where all $nc(20)_i$ observations are included but only every fourth $nc(30)_i$ observation is considered in (b).

chromatogram that is obtained from such an analysis in the so-called power spectrum, the individual variance components of the observed series are depicted. Unlike an ANOVA where the statistical influences and significance of different factors and their interactions are separated, spectral analysis decomposes the individual cyclical variance components of a series, their wavelength or frequency, and their contribution to the overall variance. The power spectrum $S(f)$ is defined as

$$S(f) = 2 \int_{0}^{\infty} r(h) \cos(2\pi\, f h)\, dh, \qquad (10.24)$$

which is an integrated form of the autocorrelation function. Hence, in a power spectrum, the presence of a cyclic component is manifested by a peak corresponding to its specific wavelength or frequency. The lag distance corresponding to the full length of its cycle is the corresponding wavelength in the power spectrum. For more details, see Bazza et al. (1988), Shumway (1988), and Nielsen and Wendroth (2003). Sisson and Wierenga (1981) decomposed the spatial variance of steady-state infiltration rates into cyclic components. Folorunso and Rolston (1985) analyzed cyclic relationships between soil properties and denitrification gas fluxes using spectral analysis. Kachanoski et al. (1985a, 1985b) identified frequency dependence of soil microtopography and thickness of A and B horizons. Nielsen et al. (1983) derived how different micrometeorological impacts that would inhibit a uniform interpretation of plant canopy temperature and soil water content measurements can be decomposed by cospectral analysis. Böttcher et al. (1997) showed in a German pine stand that the spatial variability of groundwater chemistry was related to a cyclic solute input pattern at the surface. Timlin et al. (1998) identified similar frequencies for curvature and grain yield spectra in a

farmer's field grown to corn. Wendroth and Nielsen (2002) found frequency-based variations of soil water content and concomitant leaching of a chloride tracer associated with soil tillage operations. The effect of tillage- and traffic-induced changes in soil structure was investigated by Roseberg and McCoy (1992) using spectral techniques. Domsch and Wendroth (1997) discovered cyclic variation of soil structure and penetrometer resistance caused by the management effect of soil tillage with varying intensity.

In Figure 10.12a and b, two cyclic series are given, both based on sine or cosine waves. Series x_1 (Figure 10.12a) is a wavelet consisting of waves of different length and amplitudes, which will be explained in the next paragraph. Series x_2 (Figure 10.12b) consists of 3.5 sine waves, each being 64 observations long. The sum of both series results is series x_3, shown in Figure 10.12c.

The power spectrum of series x_3 as a function of frequency f is shown in Figure 10.13. Inasmuch as frequency is inversely related to wavelength or lag distance of a particular cycle, as the position of a peak in the graph moves from left to right, it represents a variance component based on a cycle of decreasing wavelength or lag distance. Hence, the peak observed at a frequency $f = 0.016$ corresponds to a wavelength of 64, which reflects the sine wave of series x_2 shown in Figure 10.12b whose length is 64 observation points. The other three peaks corresponding to a wavelength of 6.7, 5.5, and 4.5, respectively, are caused by the cyclic components of series x_1. Notice that none of the peaks based on series x_1 approach the magnitude of the peak based on series x_2. The sine wave is the dominating process, whereas the cyclic components of process x_1 cause fluctuations smaller in magnitude. During spectral analysis, the resolution of the power spectrum, that is, the number of wavelengths for which the power spectrum is calculated has to be determined. Depending on the resolution of the frequency range, the number of spectral estimates must be chosen as a power of 2. The statistical significance of a peak in
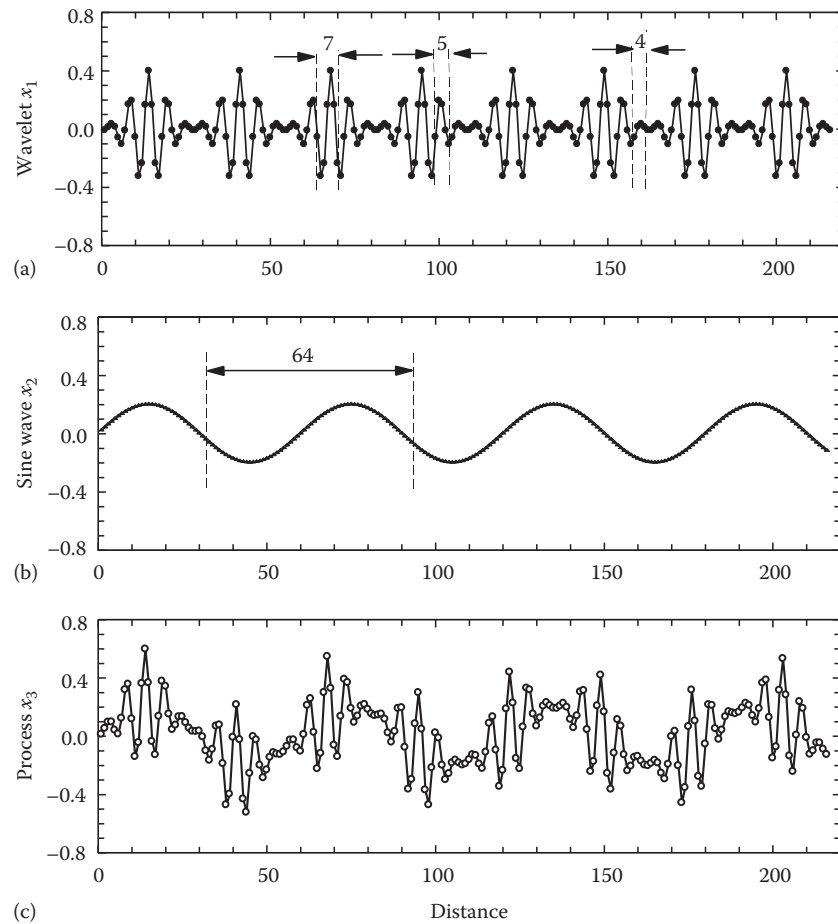
FIGURE 10.12 Series of Morlet wavelets x_1 (a), of sine waves x_2 (b), and the sum of both denoted as process x_3 (c).
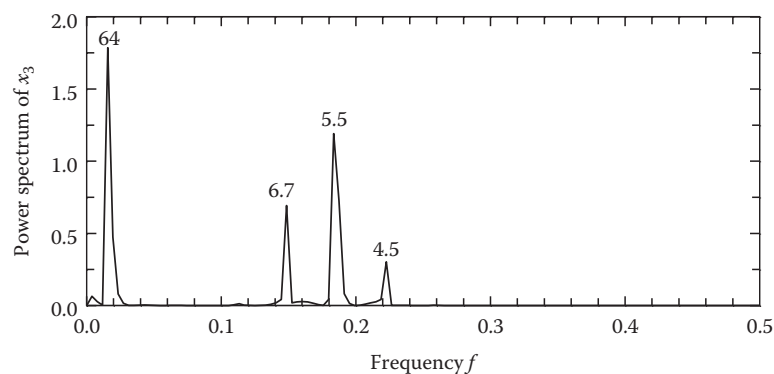
FIGURE 10.13 Power spectrum for process x_3 depicted in Figure 10.12, which is the sum of a set of Morlet wavelets and a sine wave.

the power spectrum can be quantified based on the *F*-statistics of the spectral variance decomposition, which is not described here. For further details see Bazza et al. (1988), Shumway (1988), and Wendroth and Nielsen (2002).

Two sets of data measured in a farmer's wheat field in Kentucky are shown in Figure 10.14. Wheat grain yield was obtained with an automated combine harvester yield mapping system, and the normalized difference vegetation index (NDVI) was measured in spring time with an active laser-induced canopy reflectance

sensor, known as Greenseeker. Both data series were measured along a transect, and sensor data aggregated over a width of 10 m and a length of 5 m. The most pronounced cyclic variation components underlying the wheat yield and the NDVI data occur at a frequency $f = 0.035$, corresponding to a wavelength of 29 observations. The distance of 29 observations taken at 5 m intervals matches almost perfectly to 150 m. This distance is the wavelength at which nitrogen fertilizer was applied at continuously varying rates between 0 and 168 kg ha^{-1}—not randomized—in a
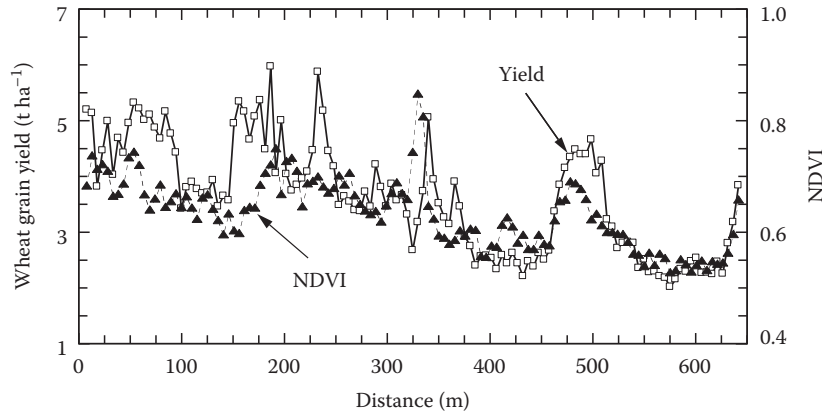
FIGURE 10.14 Wheat grain yield and normalized difference vegetation index (NDVI) measured in a 650 m long strip in a farmer's field in Western Kentucky.

sinusoidal pattern in plots of 30 m width and 15 m length along the farmer's field, similar to a study of Shillito et al. (2009). At a band of smaller frequencies $f = 0.016$–0.023 or wavelengths between 43 and 64 observations, that is, 215–320 m whose center is depicted at $f = 0.02$ in Figure 10.15a, a less pronounced cyclic

FIGURE 10.15 Power spectra for wheat grain yield (a) and NDVI (b) measured in a farmer's field in Western Kentucky and presented in Figure 10.14. Cross- and quad-spectra are shown in (c) and the coherency spectrum in (d).

variation component can be identified. Soil differences reaching especially from 180 to 480 m along the transect coupled with long-term cropping history cause this long trend in variation. The peak at $f = 0.0625$, corresponding to a wavelength of 16 observations or 80 m is more pronounced in the NDVI (Figure 10.15b) than in the yield spectrum (Figure 10.15a).

In order to better understand and relate different spatial processes occurring in this field, spectral relationships between two different variables can be identified with joint analysis of both data series using cospectral analysis.

10.9 Cospectral Analysis

As the spectrum is an integrated form of the autocorrelation function, the cross-spectrum $Co(f)$ is an integrated form of the cross-correlation function $r_c(h)$:

$$Co(f) = 2\int_0^\infty r_c(h)\cos(2\pi fh)\,dh. \tag{10.25}$$

For cospectral analysis, the left- and the right-hand side of the cross-correlation function needs to be averaged. In case of the cross-spectrum, this average accomplished by

$$r_c(h) = 0.5[r_c(h < 0) + r_c(h > 0)] \tag{10.26}$$

yields an even function when both sides of the cross-correlation function are equal. The resulting cosine wave with a maximum at 0, decreases to the minimum at a lag equal to half of the wavelength and increases toward a lag corresponding to the full length of the cosine wave. Hence, the cross-spectrum emphasizes cyclic variations described by a cosine wave. The second component of cospectral analysis is the so-called quad-spectrum. It is based on an averaging of positive and negative lags of the cross-correlation function with

$$r_c'(h) = 0.5[r_c(h < 0) - r_c(h > 0)]. \tag{10.27}$$

The quad-spectrum defined as

$$Q(f) = 2\int_0^\infty r_c'(h)\sin(2\pi f h)\mathrm{d}h, \qquad (10.28)$$

identifies those common cyclic variations between two series of observations that are delayed or shifted against each other by a quarter of a wavelength. This odd function emphasizes cyclic variations that are described by a sine wave.

For two series that yield a symmetric cross-correlation function at both sides of zero lag, that is, that are described by a cosine function, the cross-spectrum reflects common cyclic variations whereas the quad-spectrum is zero. On the other hand, common cyclic variation components of two series that are delayed against each other by a quarter period or 90° can be described by a sine function, and the quad-spectrum shows their common frequency components while the cross-spectrum is zero.

The extent of the delay against each other or the phase lag is quantified in the phase spectrum $h_\varphi(f)$ with

$$h_\varphi(f) = \frac{1}{2\pi f}\tan^{-1}\left[\frac{Q(f)}{Co(f)}\right]. \qquad (10.29)$$

The cross-spectrum shown in Figure 10.15c clearly identifies the two strongly pronounced frequencies that the wheat yield and NDVI data have in common. Their physical interpretations have been given above. For the same pair of variables, the quad-spectrum is merely different from zero for all frequencies as both series move along with any lag relative to each other. Hence, no matter what the shift between two series is against each other, the cross- or the quad-spectrum identifies common cyclic variance components. The shift against each other is quantified in the phase spectrum.

Similar to the coefficient of determination in regression analysis of two variables, the value of the coherency spectrum $Coh(f)$ is close to 1 for every frequency at which the two variables $A_i(x_i)$ and $B_i(x_i)$ have common cyclic variation components, regardless of any shift between them. The coherency spectrum is based on the cross- and quad-spectrum and the two univariate spectra:

$$Coh(f) = \frac{Co^2(f) + Q^2(f)}{S_A(f)S_B(f)}. \qquad (10.30)$$

The coherency spectrum for yield and NDVI series is presented in Figure 10.15d and clearly reflects the common cyclic variance behavior of wheat yield and NDVI over the entire range of frequencies. Since both NDVI and wheat yield evidently show common covariance and coherence, their statistical relationship can be used to predict yield based on NDVI measurements.

It is important to notice that for spectral analysis, sampling distances need to be uniform. The sampling frequency needs to be designed large enough to sufficiently cover the shortest wave component that is expected to cause variability in an experiment, that is, at least four observations per cycle should be taken to avoid under sampling the spatial process.

10.10 Wavelet Analysis

As illustrated and discussed in the previous paragraph, spatial and temporal data series can consist of fluctuations over short or long distances and reflect trends, periodic fluctuations, etc. (Grinsted et al., 2004). Very often, different components of fluctuations underlie each other. In the previous paragraph, periodic oscillations that can occur in a repetitive manner were identified for uni- and bivariate data sets using spectral and cospectral analysis, that is, frequency domain-based techniques. Another analytical technique that decomposes cyclic variations of a data series is wavelet analysis. Unlike spectral analysis, wavelet analysis detects segments within a data set that consist of cyclic fluctuations but that do not necessarily repeat regularly over the entire data set. A wavelet spectrum therefore indicates where in the data set such a fluctuation occurs and over what scale it extends.

Basic ideas and simple applications of wavelet analysis are discussed in this paragraph. The more interested reader is referred to Torrence and Compo (1998), Grinsted et al. (2004), Lau and Weng (1995), and especially the work of Shu et al. (2008), Si (2003), and Si and Farrell (2004) who applied uni- and bivariate wavelet analysis to identify spatial relationships between soil textural and soil hydraulic property information. Wavelet analysis was applied by Lark and Webster (1999) to detect spatial changes in soil type. In this analysis, the continuous wavelet transform is used. Alternatively, the discrete wavelet transform may be applied, which can be considered as a subsampling version of the continuous wavelet transform (Shu et al., 2008). To be applicable in wavelet analysis, the data set needs to consist of equally spaced observations in space or time. Prior to wavelet analysis, a basic or "mother" wavelet shape is chosen, for example, Haar or Daubechies (Nievergelt, 1999; Walker, 1999; Boggess and Narcowich, 2001), Morlet, Paul, or Derivative of Gaussian (DOG) (Torrence and Compo, 1998), and Mexican Hat (Lark and Webster, 1999). The choice of the basic wavelet depends on the nature of the data and the kind of information expected to be extracted or filtered by the analysis. In the wavelet estimation algorithm, this basic shape of the wavelet is moved across the data series and fitted to individual segments of the series, while it is "stretched" in length and amplitude to fit the subset of data as close as possible. Hence, the amplitude or magnitude of the signal and the length scale over which a given wavelet fits a section of a data series are both flexible. Unlike spectral analysis, the result of wavelet analysis is not a periodogram that manifests repetitive patterns of the data in one peak at a given frequency or frequency band. A wavelet spectrum displays the segment of the data series exhibiting wavelet-like fluctuation. It shows where within the data series the fluctuation occurs, over what length scale it extends, and how much it contributes to the overall variance. Hence, the *x*-axis of the wavelet spectrum is the same as

the *x*-axis of the data series, that is, distance in time or space. The *y*-axis represents the scale given in the original sampling units over which the fluctuation expands.

From the various wavelet shapes, a Morlet wavelet is chosen here, which is a plane wave modulated by a Gaussian (Torrence and Compo, 1998) because it allows a good balance between space and frequency localization (Shu et al., 2008)

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2}, \tag{10.31}$$

where

η is the dimensionless time parameter
ω_0 the frequency, here equal to 6

Expressing the complex exponential term containing the imaginary number $i = \sqrt{-1}$ as

$$e^{i\omega_0\eta} = \cos(\omega_0\eta) + i\sin(\omega_0\eta) \tag{10.32}$$

the Morlet wavelet is the sum of a cosine curve (its real part) and a sine curve (its imaginary part).

In Figure 10.16a, a series x_4 consisting of five wavelets is given. Each of the first four wavelets is extended over a length of 27 points. The fifth wavelet reaches from 109 to 216 distance, and its maxima and minima are reduced by a factor of 3 compared to the first four wavelets.

Here, only the continuous wavelet transform $W_n(s)$ is presented. Based on the inverted discrete Fourier transform, the wavelet transform of a series X_i ($i = 1, \ldots, N$) with uniform sampling distance δx is defined as the convolution of X_i as (Torrence and Compo, 1998; Grinsted et al., 2004; Shu et al., 2008)

$$W_i(s) = \sqrt{\frac{\delta x}{s}} \sum_{j=1}^{N} X_i \psi \left[(j-1)\frac{\delta x}{s} \right], \tag{10.33}$$

where the wavelet scale *s* denotes a complex conjugate. At each scale, the wavelet is normalized in order to be comparable to wavelets at other scales and in other series. Further details on the normalization of wavelet functions are explained in Torrence and Compo (1998). With increasing scale the length of the base wavelet increases. Hence, the wavelet coefficient, that is, the weight over a wavelet domain depends on the location and the length of the wavelet. Since the sum of these weights over the wavelet domain always adds up to unity, the wavelet coefficients reflect the variation of samples. The larger the scale of a wavelet becomes, the higher is the edge artifact. For this reason, the cone of influence (COI) is introduced (Grinsted et al., 2004).

Similar to cospectral analysis, the cross-wavelet spectrum reflects areas of high common fluctuations at the same location and across the same scale. The cross-wavelet spectrum is defined as

$$\left| W_i^{XY}(s) \right| = \left| W_i^Y \overline{W_i^X(s)} \right|, \tag{10.34}$$

with

$\overline{W_i^X}$ being the complex conjugate of W_i^X
W_i^{XY} the local relative phase between X_i and Y_i

The normalized form of the cross-wavelet spectrum is the wavelet coherence, indicating the measure of correlation of two series at different resolutions. In its squared form and applying a smoothing operator *S*, the wavelet coherence is defined as (Shu et al., 2008)

(a)

(b)
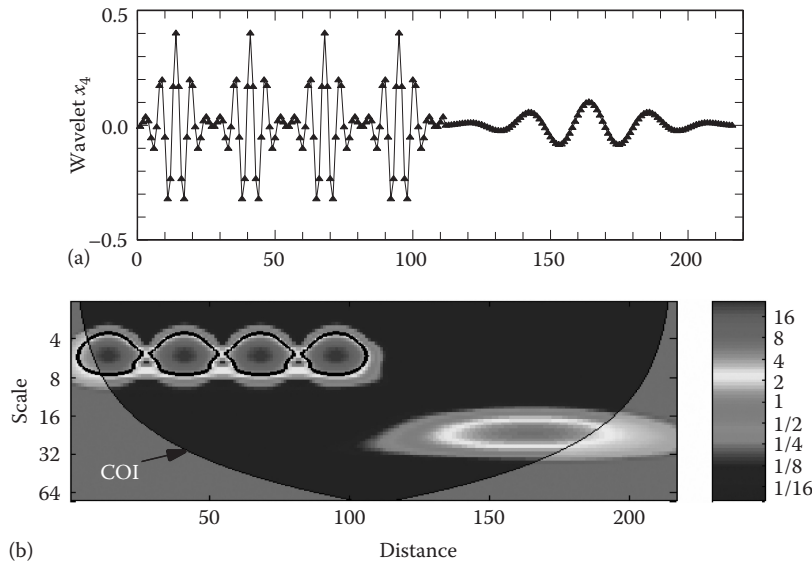
FIGURE 10.16 Series x_4 consisting of four short Morlet wavelets between distances of 0 and 115, and one stretched Morlet wavelet reaching from 120–210 distance with a smaller amplitude (a). The corresponding wavelet spectrum is presented in (b) with contours reflecting data fluctuations, their scale (period), and their location along the series. The COI is depicted.
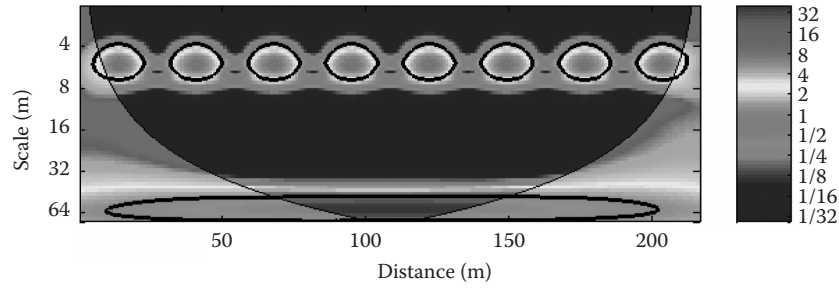
FIGURE 10.17   Wavelet spectrum for process x_3, shown in Figure 10.12c.

$$R_i^2(s) = \frac{\left|S(W_i^{XY}(s)s^{-1})\right|^2}{S\left(\left|W_i^X(s)\right|^2 s^{-1}\right) S\left(\left|W_i^Y(s)\right|^2 s^{-1}\right)}. \qquad (10.35)$$

The smoothing operator depends on the base form of the wavelet. For further details see Torrence and Webster (1999) and Si and Zeleke (2005).

The continuous wavelet transform of the process x_4 is shown in Figure 10.16b. As expected for this illustrative example, in the six-point wavelength band, the four basic wavelets cause pronounced signal contours between distances of 0 and 110. At a different wave band reaching over an extent between 16 and 32 points, a less pronounced peak or signal contour manifests the smaller variance of the individual longer wavelet. The contour of the signal is stretched according to the more extended range of the basic Morlet wavelet in the observed series. The area covered with intensive colors marks the above-mentioned COI (Figure 10.16b). Outside this cone, the detection of variance components may be affected by edge artifacts.

The wavelet transform of process x_3 (Figure 10.12c), emphasizes the diagnostic capabilities of wavelet transform analysis as shown in Figure 10.17. At the six-point wavelength band, the eight base wavelets contribute to the variation at their respective locations. At the 64-point band, another pronounced variation component is reflected in the wavelet spectrum. The variation at this band clearly identifies the sine waves, whose wavelength was given with 64 points (Figure 10.12b).

The wavelet spectra for both wheat yield and NDVI data (Figure 10.14) exhibit some common variability features at the 30 m band, at approximately 330 m distance in the transect referring to the transition zone between soil types and cropping history (Figure 10.18a and b). For the 150 m wavelength of nitrogen fertilizer application pattern, a strong variability influence is evident from the wavelet spectra of both, wheat yield and NDVI. This feature is more pronounced for the NDVI spectrum (Figure 10.18b) than for the yield spectrum (Figure 10.18a), as obvious from the original data.

The cross-wavelet spectrum (Figure 10.18c) emphasizes the common locations of fluctuations especially at the 30–60 m band at ~300–360 m along the transect, the zone where a transition from higher to lower soil quality occurs. In this cross-wavelet
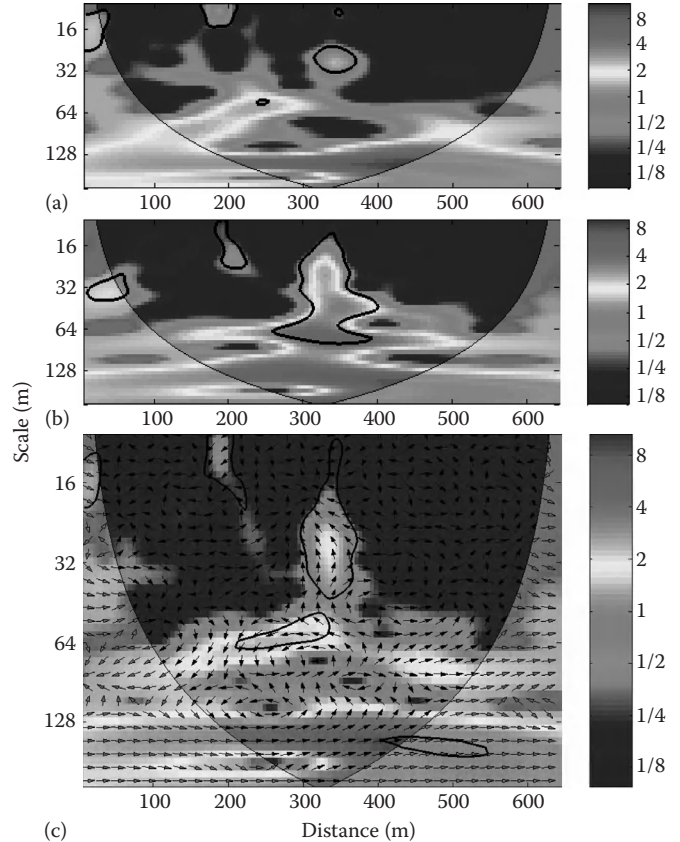
FIGURE 10.18   Wavelet spectrum for wheat yield (a) and NDVI (b) data, both shown in Figure 10.14. The cross-wavelet spectrum is presented in (c).

spectrum, arrows to the right indicate in-phase behavior, and arrows to the left out-of-phase behavior between the wavelet spectra. In Figure 10.19, the coherency wavelet spectrum reflects the strong association between both variables at the 16–30 m band between 300 and 360 m. Moreover, the common variation pattern is present along the entire transect at the 150 m band, manifesting the impact of cyclically varying nitrogen application.

Notice that the wavelet spectrum indicates the local presence of individual variation components, their wavelength band or scale, and their contribution to the variance at that scale. If the base wavelet shape is chosen appropriately, the
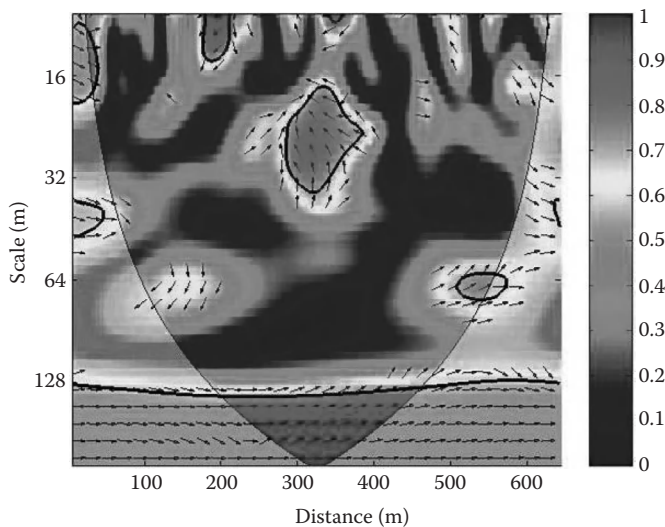
FIGURE 10.19 Coherency wavelet for wheat yield and NDVI data shown in Figure 10.14.

wavelet spectrum is a tool eligible to detect a pronounced fluctuation in a data series, its intensity, extent, and local relationship to the fluctuation of another process. On the other hand, if cyclically repeating variation components of one or several series and their phase shift relative to each other need to be identified, power spectrum and cross-spectral analysis are the analytical tools of choice.

Appendix

Data Analysis Software

The computer programs for continuous wavelet transform and wavelet coherence, based on MATLAB®, were provided by A. Grinsted and are available at http://www.pol.ac.uk/home/research/waveletcoherence. Geostatistical analyses for semivariograms, kriging, and cokriging were accomplished with ISATIS (Geovariances). Time series analysis, that is, auto-, cross-, and partial autocorrelation, autoregressive state-space analysis, spectral and cross-spectral analysis were conducted with WINASTSA.

Acknowledgments

References

Alemi, M.H., M.R. Shahriari, and D.R. Nielsen. 1988. Kriging and cokriging of soil water properties. Soil Technol. 1:117–132.

Assouline, S. 1993. Estimation of lake hydrologic budget terms using simultaneous solution of water, heat, and salt balances and a Kalman filtering approach: Application to Lake Kinneret. Water Resour. Res. 29:3041–3048.

Ball, S.T., D.J. Mulla, and C.F. Konzak. 1993. Spatial heterogeneity affects variety trial interpolation. Crop Sci. 33:931–935.

Barley, K.P. 1964. The utility of field experiments—A comment on a paper by Collis-George and Davey (1960). Soils Fertil. 27:267–269.

Bazza, M., R.H. Shumway, and D.R. Nielsen. 1988. Two-dimensional spectral analyses of soil surface temperature. Hilgardia 56:1–28.

Bell, G., M.J. Lechowicz, A. Appenzeller, M. Chandler, E. De Blois, L. Jackson, B. MacKenzie, R. Preziosi, M. Schallenberg, and N. Tinker. 1993. The spatial structure of the physical environment. Oecologia 96:114–121.

Ben-Jemaa, F., M.A. Mariño, and H.A. Loaiciga. 1994. Multivariate geostatistical design of ground-water monitoring networks. J. Water Resour. Plann. Manage. 120:505–522.

Bhatti, A.U., D.J. Mulla, and B.E. Frazier. 1991a. Estimation of soil properties and wheat yields on complex eroded hills using geostatistics and thematic mapper images. Remote Sens. Environ. 37:181–191.

Bhatti, A.U., D.J. Mulla, F.E. Koehler, and A.H. Gurmani. 1991b. Identifying and removing spatial correlation from yield experiments. Soil Sci. Soc. Am. J. 55:1523–1528.

Bierkens, M.F., M. Knotters, and T. Hoogland. 2001. Space-time modeling of water table depth using a regionalized time series model and the Kalman filter. Water Resour. Res. 37:1277–1290.

Bishop, T.F.A., and R.M. Lark. 2006. The geostatistical analysis of experiments at the landscape scale. Geoderma 133:87–106.

Boggess, A., and F.J. Narcowich. 2001. A first course in wavelets with Fourier analysis. Prentice Hall, Upper Saddle River, NJ.

Böttcher, J., S. Lauer, O. Strebel, and M. Puhlmann. 1997. Spatial variability of canopy throughfall and groundwater sulfate concentrations under a pine stand. J. Environ. Qual. 26:503–510.

Boyer, D.G., R.J. Wright, C.M. Feldhake, and D.P. Bligh. 1996. Soil spatial variability relationships in a steeply sloping acid soil environment. Soil Sci. 161:278–287.

Brus, D.J., and J.J. de Gruijter. 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma 80:1–59.

Burgess, T.M., and R. Webster. 1980. Optimal interpolation and isarithmic mapping of soil properties I. The semivariogram and punctual kriging. J. Soil Sci. 31:315–331.

Buscaglia, H.J., and J.J. Varco. 2003. Comparison of sampling designs in the detection of spatial variability of Mississippi Delta soils. Soil Sci. Soc. Am. J. 67:1180–1185.

Cahill, A.T., F. Ungaro, M.B. Parlange, M. Mata, and D.R. Nielsen. 1999. Combined spatial and Kalman filter estimation of optimal soil hydraulic properties. Water Resour. Res. 35:1079–1088.

Cambardella, C.A., T.B. Moorman, J.M. Novak, T.B. Parkin, D.L. Karlen, R.F. Turco, and A.E. Konopka. 1994. Field-scale variability of soil properties in Central Iowa soils. Soil Sci. Soc. Am. J. 58:1501–1511.

Cassel, D.K., O. Wendroth, and D.R. Nielsen. 2000. Assessing spatial variability in an agricultural experiment station field: Opportunities arising from spatial dependence. Agron. J. 92:706–714.

Chaplot, V., C. Walter, P. Curmi, and A. Hollier-Larousse. 2000. The use of auxiliary geophysical data to improve a soil-landscape model. Soil Sci. 165:961–970.

Cichota, R., A.L. Berreta Hurtado, and Q. de Jong van Lier. 2006. Spatio-temporal variability of soil water tension in a tropical soil. Geoderma 133:231–243.

Collis-George, N., and B.G. Davey. 1960. The doubtful utility of present-day field experimentation and other determinations involving soil–plant interactions. Soils Fertil. 23:307–310.

Cressie, N.A.C. 1991. Statistics for spatial analysis. Wiley Interscience, New York.

Davis, J.C. 1986. Statistics and data analysis in geology. 2nd edn. John Wiley & Sons, New York.

Deutsch, C.V., and A.G. Journel. 1997. GSLIB—Geostatistical software library and user's guide. Oxford University Press, New York.

Domsch, H., and O. Wendroth. 1997. On-site diagnosis of soil structure for site specific management, p. 95–102. *In* J.V. Stafford (ed.) Proc. First Europ. Conf. Prec. Agric. BIOS Scientific Publishers, Oxford, U.K.

Donald, W.W. 1994. Geostatistics for mapping weeds with a Canada thistle (*Cirsium arvense*) patch as a case study. Weed Sci. 42:648–657.

Folorunso, O.A., and D.E. Rolston. 1985. Spatial and spectral relationships between field-measured denitrification gas fluxes and soil properties. Soil Sci. Soc. Am. J. 49:1087–1093.

Ghidey, F., and E.E. Alberts. 1999. Temporal and spatial patterns of nitrate in a claypan soil. J. Environ. Qual. 28:584–594.

Giebel, A., O. Wendroth, H.I. Reuter, K.C. Kersebaum, and J. Schwarz. 2006. How representatively can we sample soil mineral nitrogen? J. Plant Nutr. Soil Sci. 169:52–59.

Goovaerts, P., and C.N. Chiang. 1993. Temporal persistence of spatial patterns for mineralizable nitrogen and selected soil properties. Soil Sci. Soc. Am. J. 57:372–381.

Gotway, C.A., R.B. Ferguson, G.W. Hergert, and T.A. Peterson. 1996. Comparison of kriging and inverse-distance methods for mapping soil parameters. Soil Sci. Soc. Am. J. 60:1237–1247.

Gregorich, E.G., L.W. Turchenek, M.R. Carter, and D.A. Angers. 2001. Soil and environmental science dictionary. Canadian Society of Soil Science. CRC Press, Boca Raton, FL.

Grinsted, A., J.C. Moore, and S. Jevrejeva. 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes Geophys. 11:561–566.

Halvorson, J.J., J.L. Smith, H. Bolton, and R.E. Rossi. 1995. Evaluating shrub-associated spatial patterns of soil properties in a shrub-steppe using multiple-variable geostatistics. Soil Sci. Soc. Am. J. 59:1476–1487.

Isaaks, E.H., and R.M. Srivastava. 1989. Applied geostatistics. Oxford University Press, New York.

Jost, G., G.B.M. Heuvelink, and A. Papritz. 2005. Analysing the space-time distribution of soil water storage of a forest ecosystem using spatio-temporal kriging. Geoderma 128:258–273.

Journel, A.G., and Ch.J. Huijbregts. 1991. Mining geostatistics. Academic Press, London, U.K.

Kachanoski, R.G., E. de Jong, and D.E. Rolston. 1985a. Spatial and spectral relationships of soil properties and microtopography. II. Density and thickness of B-horizon. Soil Sci. Soc. Am. J. 49:812–816.

Kachanoski, R.G., D.E. Rolston, and E. de Jong. 1985b. Spatial and spectral relationships of soil properties and microtopography. I. Density and thickness of A-horizon. Soil Sci. Soc. Am. J. 49:804–812.

Kalman, R.E. 1960. A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. 8:35–45.

Kalman, R.E., and R.S. Bucy. 1961. New results in linear filtering and prediction theory. Trans. ASME J. Basic Eng. 83:95–108.

Katul, G.G., O. Wendroth, M.B. Parlange, C.E. Puente, and D.R. Nielsen. 1993. Estimation of in situ hydraulic conductivity function from nonlinear filtering theory. Water Resour. Res. 29:1063–1070.

Kravchenko, A.N. 2003. Influence of spatial structure on accuracy of interpolation methods. Soil Sci. Soc. Am. J. 67:1564–1571.

Lark, R.M., and R. Webster. 1999. Analysis and elucidation of soil variation using wavelets. Eur. J. Soil Sci. 50:185–206.

Lascano, R.J., and J.L. Hatfield. 1992. Spatial variability of evaporation along two transects of a bare soil. Soil Sci. Soc. Am. J. 56:341–346.

Lau, K.M., and H. Weng. 1995. Climate signal detection using wavelet transform: How to make a time series sing. Bull. Am. Meteorol. Soc. 76:2391–2402.

Lebron, I., M.D. Madsen, D.G. Chandler, D.A. Robinson, O. Wendroth, and J. Belnap. 2007. Ecohydrological controls on soil moisture and hydraulic conductivity within a pinyon-juniper woodland. Water Resour. Res. 43:W08422. doi: 10.1029/2006WR005398.

Liebhold, A.M., X. Zhang, M.E. Hohn, J.S. Elkington, M. Ticehurst, G.L. Benzon, and R.W. Campbell. 1991. Geostatistical analysis of gypsy moth (Lepidoptera: Lymantriidae) egg mass populations. Environ. Entomol. 20:1407–1417.

McBratney, A.B., and R. Webster. 1983. How many observations are needed for regional estimation of soil properties. Soil Sci. 135:177–183.

Morkoc, F., J.W. Biggar, D.R. Nielsen, and D.E. Rolston. 1985. Analysis of soil water content and temperature using state-space approach. Soil Sci. Soc. Am. J. 49:798–803.

Mulla, D.J. 1993. Mapping and managing spatial patterns in soil fertility and crop yield, p. 15–26. *In* P.C. Robert, R.H. Rust, and W.E. Larson (eds.) Site specific crop management. ASA/CSSA/SSSA, Madison, WI.

Mulla, D.J., A.U. Bhatti, M.W. Hammond, and J.A. Benson. 1992. A comparison of winter wheat yield and quality under uniform versus spatially variable fertilizer management. Agric. Ecosyst. Environ. 38:301–311.

Mulla, D.J., A.U. Bhatti, and R. Kunkel. 1990. Methods for removing spatial variability from field research trials. Adv. Soil Sci. 13:201–213.

Mulla, D.J., and A.B. McBratney. 1999. Soil spatial variability, p. A321–A352. *In* M.E. Sumner (ed.) Handbook of soil science. 1st edn. CRC Press, Boca Raton, FL.

Nielsen, D.R. 1987. Emerging frontiers in soil science. Geoderma 40:267–273.

Nielsen, D.R., and M.H. Alemi. 1989. Statistical opportunities for analyzing spatial and temporal heterogeneity of field soils. Plant Soil 115:285–296.

Nielsen, D.R., G.G. Katul, O. Wendroth, M.V. Folegatti, and M.B. Parlange. 1994. State-space approaches to estimate soil physical properties from field measurements, p. 61–85. Proc. 15th Conf. ISSS. Vol. 2a. July 10–16, 1994. Acapulco, Mexico.

Nielsen, D.R., P.M. Tillotson, and S.R. Vieira. 1983. Analyzing field-measured soil water properties. Agric. Water Manage. 6:93–109.

Nielsen, D.R., and O. Wendroth. 2003. Spatial and temporal statistics—Sampling field soils and their vegetation. Catena, Reiskirchen, Germany.

Nievergelt, Y. 1999. Wavelets made easy. Birkhäuser, Boston, NJ.

Parlange, M.B., G.G. Katul, M.V. Folegatti, and D.R. Nielsen. 1993. Evaporation and the field scale soil water diffusivity function. Water Resour. Res. 29:1279–1286.

Raftery, A.E. 1993. Change point and change curve modeling in stochastic processes and spatial statistics. J. Appl. Stat. Sci. 1:403–423.

Ritzi, R.W. 2000. Behavior of indicator variograms and transition probabilities in relation to the variance in lengths of hydrofacies. Water Resour. Res. 36:3375–3381.

Robinson, D.A., C.S. Campbell, J.W. Hopmans, B.K. Hornbuckle, S.B. Jones, R. Knight, F. Ogden, J. Selker, and O. Wendroth. 2008. Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. Vadose Zone J. 7:358–389.

Roseberg, R.J., and E.L. McCoy. 1992. Tillage- and traffic-induced changes in macroporosity and macropore continuity: Air permeability assessment. Soil Sci. Soc. Am. J. 56:1261–1267.

Salas, J.D., J.W. Delleur, V. Yevjevich, and W.L. Lane. 1988. Applied modeling of hydrologic time series. Water Resources Publications, Littleton, CO.

Shillito, R.M., D.J. Timlin, D. Fleisher, V.R. Reddy, and B. Quebedeaux. 2009. Yield response of potato to spatially patterned nitrogen application. Agric. Ecosyst. Environ. 129:107–116.

Shu, Q., Z. Liu, and B. Si. 2008. Characterizing scale- and location-dependent correlation of water retention parameters with soil physical properties using wavelet techniques. J. Environ. Qual. 37:2284–2292.

Shumway, R.H. 1985. Time series in soil science: Is there life after kriging. *In* D.R. Nielsen and J. Bouma (eds.) Soil spatial variability. Proc. Workshop ISSS/SSSA, Las Vegas, NV.

Shumway, R.H. 1988. Applied statistical time series analysis. Prentice Hall, Englewood Cliffs, NJ.

Shumway, R.H., and D.S. Stoffer. 1982. An approach to time series smoothing and forecasting using the EM algorithm. J. Time Ser. Anal. 3:253–264.

Shumway, R.H., and D. Stoffer. 2000. Time series analysis and its applications. Springer, New York.

Si, B.C. 2003. Spatial and scale-dependent soil hydraulic properties: A wavelet approach, p. 163–178. *In* Y. Pachepsky, D.E. Radcliffe, and H.M. Selim (eds.) Scaling methods in soil physics. CRC Press, Boca Raton, FL.

Si, B.C., and R.E. Farrell. 2004. Scale- dependent relationships between wheat yield and topographic indices: A wavelet approach. Soil Sci. Soc. Am. J. 68:577–588.

Si, B.C., and T.B. Zeleke. 2005. Wavelet coherency analysis to relate saturated hydraulic properties to soil physical properties. Water Resour. Res. 41:W11424. doi: 10.1029/2005WR004118.

Sisson, J.B., and P.J. Wierenga. 1981. Spatial variability of steady-state infiltration rates as a stochastic process. Soil Sci. Soc. Am. J. 45:699–704.

Skaggs, T.H., and B.P. Mohanty. 1998. Water table dynamics in tile-drained fields. Soil Sci. Soc. Am. J. 62:1191–1196.

Smith, J.L., J.L. Halvorson, and R.I. Papendick. 1993. Using multiple-variable indicator kriging for evaluating soil quality. Soil Sci. Soc. Am. J. 57:743–749.

Timlin, D.J., Ya. Pachepsky, V.A. Snyder, and R.B. Bryant. 1998. Spatial and temporal variability of corn grain yield on a hillslope. Soil Sci. Soc. Am. J. 62:764–773.

Torrence, C., and G.P. Compo. 1998. A practical guide to wavelet analysis. Bull. Am. Meteorol. Soc. 79:61–78.

Torrence, C., and P. Webster. 1999. Interdecadal changes in the ESNO–monsoon system. J. Clim. 12:2679–2690.

Trangmar, B.B., R.S. Yost, and G. Uehara. 1985. Application of geostatistics to spatial studies of soil properties. Adv. Agron. 38:45–94.

Van Es, H.M., and C.L. Van Es. 1993. Spatial nature of randomization and its effect on the outcome of field experiments. Agron. J. 85:420–428.

Van Groenigen, J.W., A. Stein, and R. Zuurbier. 1997. Optimization of environmental sampling using interactive GIS. Soil Technol. 10:83–97.

Vauclin, M., S.R. Vieira, G. Vachaud, and D.R. Nielsen. 1983. The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J. 47:175–184.

Vieira, S.R., Hatfield, J.L., Nielsen, D.R., and Biggar, J.W. 1983. Geostatistical theory and application to variability of some agronomic properties. Hilgardia 51:1–75.

Vieira, S.R., D.R. Nielsen, and J.W. Biggar. 1981. Spatial variability of field-measured infiltration rate. Soil Sci. Soc. Am. J. 45:1040–1048.

Walker, J.S. 1999. A primer on wavelets and their scientific applications. CRC Press, Boca Raton, FL.

Walker, J.P., G.R. Willgoose, and J.D. Kalma. 2002. Three-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: Simplified Kalman filter covariance forecasting and field application. Water Resour. Res. 38:1301.

Warrick, A.W., and D.E. Myers. 1987. Optimization of sampling locations for variogram calculations. Water Resour. Res. 23:496–500.

Warrick, A.W., D.E. Myers, and D.R. Nielsen. 1986. Geostatistical methods applied to soil science, p. 53–82. *In* A. Klute (ed.) Methods of soil analysis. Part 1. 2nd edn. Agronomy Monograph No. 9. ASA and SSSA, Madison, WI.

Webster, R. 1985. Quantitative spatial analysis of soil in the field. Adv. Soil Sci. 3:1–70.

Webster, R., and G.B.M. Heuvelink. 2006. The Kalman filter for the pedologist's tool kit. Eur. J. Soil Sci. 57:758–773.

Wendroth, O., A.M. Al-Omran, C. Kirda, K. Reichardt, and D.R. Nielsen. 1992. State-space approach to spatial variability of crop yield. Soil Sci. Soc. Am. J. 56:801–807.

Wendroth, O., P. Jürschik, K.C. Kersebaum, H. Reuter, C. van Kessel, and D.R. Nielsen. 2001. Identifying, understanding, and describing spatial processes in agricultural landscapes—Four case studies. Soil Tillage Res. 58:113–128.

Wendroth, O., G.G. Katul, M.B. Parlange, C.E. Puente, and D.R. Nielsen. 1993. A nonlinear filtering approach for determining hydraulic conductivity functions in field soils. Soil Sci. 156:293–301.

Wendroth, O., S. Koszinski, and E. Pena-Yewtukhiv. 2006. Spatial association between soil hydraulic properties, soil texture and geoelectric resistivity. Vadose Zone J. 5:341–355.

Wendroth, O., and D.R. Nielsen. 2002. Time and space series, p. 119–137. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis, Part 4. 3rd edn., Chapter 1, Soil sampling and statistical procedures. Soil Science Society of America Book Series No. 5. Soil Science Society of America Inc., Madison, WI.

Wendroth, O., and D.A. Robinson. 2008. Scaling processes in watersheds, p. 1024–1028. *In* S.W. Trimble, B.A. Stewart, and T.A. Howell (eds.) Encyclopedia of water science. 2nd edn. Taylor & Francis, New York.

Wilkinson, G.N., S.R. Eckert, T.W. Hancock, and O. Mayo. 1983. Nearest neighbor (NN) analysis of field experiments (including discussion). J. R. Stat. Soc. B. 45:151–211.

Wu, L., T.H. Skaggs, P.J. Shouse, and J.E. Ayars. 2001. State space analysis of soil water and salinity regimes in a loam soil underlain by shallow groundwater. Soil Sci. Soc. Am. J. 65:1065–1074.

Yates, S.R., and A.W. Warrick. 2002. Geostatistics, p. 81–118. *In* J.H. Dane and G.C. Topp (eds.) Methods of soil analysis. Part 4. 3rd edn., Chapter 1, Soil sampling and statistical procedures. Soil Science Society of America Book Series No. 5. Soil Science Society of America Inc., Madison, WI.

Yost, R.S., G. Uehara, and R.L. Fox. 1982a. Geostatistical analysis of soil chemical properties of large land areas. 1. Semivariograms. Soil Sci. Soc. Am. J. 46:1028–1032.

Yost, R.S., G. Uehara, and R.L. Fox. 1982b. Geostatistical analysis of soil chemical properties of large land areas. II. Kriging. Soil Sci. Soc. Am. J. 46:1033–1037.

Zhang, R., D.E. Myers, and A.W. Warrick. 1992. Estimation of the spatial distribution of soil chemicals using pseudo-cross-variograms. Soil Sci. Soc. Am. J. 56:1444–1452.

Zhang, R., P. Shouse, and S. Yates. 1999. Estimates of soil nitrate distributions using cokriging with pseudo-crossvariograms. J. Environ. Qual. 28:424–428.

II

Soil Chemistry

Donald L. Sparks
University of Delaware

SINCE THE FIRST EDITION of the *Handbook of Soil Science* was published in 2000, many changes have taken place in the area of soil chemistry. Global environmental challenges such as climate change, water quality, nutrient management, soil contamination, ecosystem health, and energy sustainability have brought soil chemistry research to the forefront. Because these are complex challenges, they must be tackled in a highly interdisciplinary/multidisciplinary manner over a range of spatial and temporal scales. Thus, soil chemists are increasingly interacting with a wide array of scientists and engineers in multiple fields such as chemistry, biology, physics, geochemistry, engineering, economics, ecology, and environmental policy. The use of highly sophisticated analytical tools, particularly those that are in situ and synchrotron-based, and development of advanced speciation and computational models have provided soil chemists with tools that are indispensable in understanding soil chemical reactions and processes at multiple scales. Accordingly, significant advances have been made in understanding, as never before, the chemistry of soil organic matter, soil solution and solid phase speciation, the kinetics and mechanisms of soil chemical reactions, soil colloidal chemistry, oxidation–reduction, ion exchange, adsorption and precipitation processes and complexes, abiotic catalysis, and soil acidity. Advances in these topics are comprehensively covered in the following chapters.

I am extremely grateful to the authors for their excellent contributions and to the referees who made many helpful suggestions for improvement.

11

Soil Organic Matter

Jeffrey A. Baldock
Commonwealth Scientific and Industrial Research Organisation

Kris Broos
Flemish Institute for Technological Research

11.1 Introduction

Research pertaining to the organic fraction of soils can be traced back in excess of 200 years. Achard (1786) isolated a dark amorphous precipitate upon acidification of an alkaline extract from peat. The effect of organic matter on soil N fertility (von Liebig, 1840), studies on the use of animal manures for maintaining soil fertility (Lawes, 1861), and the influence of soil and tree species on the development of humus form (Muller, 1887) all demonstrated the importance of organic matter in soil processes. The advancement of organic chemical methodologies and confirmation of the presence of various chemical structures in soil organic matter (SOM) lead to the development of theories that SOM was composed of a heterogeneous mixture of dominantly colloidal organic substances containing acidic functional groups and N. More recently, the polyphenol theory was proposed in which quinone structures of lignin and microbial origin polymerize in the presence of N-containing groups (amino acids, peptides, and proteins) to produce nitrogenous polymers (Flaig et al., 1975).

Early research pertaining to SOM has been reviewed by Stevenson (1994). While alkaline extraction of SOM is still practiced, modern analytical techniques, including solid-state ^{13}C nuclear magnetic resonance (^{13}C NMR) spectroscopy, infrared (IR) spectroscopy, pyrolysis gas chromatography/mass spectroscopy (Py-GC/MS), and x-ray adsorption (XAFS), allow selective probing of SOM chemistry within samples of whole soil. Application of these technologies avoids problems of incomplete extraction, artifact synthesis, and lack of biological significance often ascribed to alkaline extraction procedures. The combination of these techniques with SOM fractionation approaches that are capable of identifying biologically important SOM components has significantly advanced our knowledge of the organic fraction of soils and its dynamics over the last 30 years.

Despite such a long history of research and new methodological and technological advancements, many questions related to the genesis and chemical composition of SOM and its impacts on soil fertility, soil pedogenesis, and soil physical and chemical properties persist today. Many excellent texts and review papers have been written on the topic of SOM. Some of the more recent of these are given in Table 11.1. Table 11.1 is not comprehensive, but rather attempts to provide a starting point for anyone interested in understanding the nature and roles of organic matter in soils.

An examination of terms used to describe SOM and its components in the literature revealed a lack of precise and consistent definitions of what SOM is and what its various component fractions represent. Such a problem exists because of the heterogeneity of SOM with respect to its source, chemical and physical composition, diversity of function, and its dynamic character. As a result, the term SOM has been used to describe all organic

TABLE 11.1 List of Some of the Texts and Review Articles Pertaining to the Study of SOM Released Since 1995

Texts

SOM management for sustainable agriculture (Lefroy et al., 1995)

Humic substances of soils and general theory of humification (Orlov, 1995)

Carbon forms and functions in forest soils (McFee and Kelly, 1995)

The role of nonliving organic matter in the earth's carbon cycle (Zepp and Sonntag, 1995)

Driven by nature: Plant litter quality and decomposition (Cadisch and Giller, 1997)

Soil carbon management: Economic, environmental, and societal benefits (Kimble et al., 2007)

Soil carbon sequestration under organic farming in the Mediterranean environment (Mainari and Caporali, 2008)

Review articles

The chemical composition of SOM in classical humic compound fractions and in bulk samples—A review (Beyer, 1996)

Carbon in primary and secondary organo-mineral complexes (Christensen, 1996a)

Applications of NMR to SOM analysis—History and prospects (Preston, 1996)

Analytical pyrolysis and computer modeling of humic and soil particles (Schulten et al., 1998)

Life after death—Lignin–humic relationships reexamined (Shevchenko and Bailey, 1996)

Stabilization and destabilization of SOM—Mechanisms and controls (Sollins et al., 1996)

Soil organic carbon/SOM (Baldock and Skjemstad, 1999)

Role of the soil matrix and minerals in protecting natural organic materials against biological attack (Baldock and Skjemstad, 2000)

The supramolecular structure of humic substances: A novel understanding of humus chemistry and implications in soil science (Piccolo, 2002)

Indications for SOM quality in soils under different managements (von Lützow et al., 2002)

Importance of mechanisms and processes of the stabilization of SOM for modeling carbon turnover (Krull et al., 2003)

Cycling and composition of organic matter in terrestrial and marine ecosystems (Baldock et al., 2004)

Soil mineral–organic matter–microorganism interactions: Fundamentals and impacts (Huang, 2004)

The depth distribution of soil organic carbon in relation to land use and management and the potential of carbon sequestration in subsoil horizons (Lorenz and Lal, 2005)

Root effects on SOM decomposition (Cheng and Kuzyakov, 2005)

Mechanisms and regulation of organic matter stabilization in soils (Kögel-Knabner et al., 2005)

Labile organic matter fractions as central components of the quality of agricultural soils: An overview (Haynes, 2005)

Soil minerals and organic components: Impact on biological processes, human welfare, and nutrition (Haider and Guggenberger, 2005)

The soil carbon dioxide sink (Smith and Ineson, 2007)

Impacts of climate change on forest soil carbon: Principles, factors, models, uncertainties (Reichstein, 2007)

Composition and cycling of organic carbon in soils (Baldock, 2007)

An integrative approach of organic matter stabilization in temperate soils: Linking chemistry, physics, and biology (Kogel-Knabner et al., 2008a)

Physical carbon-sequestration mechanisms under special consideration of soil wettability (Bachmann et al., 2008)

Soil-carbon preservation through habitat constraints and biological limitations on decomposer activity (Ekschmitt et al., 2008)

Storage and stability of organic matter and fossil carbon in a luvisol and phaeozem with continuous maize cropping: A synthesis (Flessa et al., 2008)

Contribution of dissolved organic matter to carbon storage in forest mineral soils (Kalbitz and Kaiser, 2008)

Organo-mineral associations in temperate soils: Integrating biology, mineralogy, and organic matter chemistry (Kogel-Knabner et al., 2008b)

Comparison of two quantitative soil organic carbon models with a conceptual model using data from an agricultural long-term experiment (Ludwig et al., 2008)

How relevant is recalcitrance for the stabilization of organic matter in soils? (Marschner et al., 2008)

Stabilization mechanisms of organic matter in four temperate soils: Development and application of a conceptual model (von Lützow et al., 2008)

For additional earlier references consult Hedges and Oades (1997) and Baldock and Nelson (2000).

materials found in soil (Stevenson, 1994), all organic materials excluding charcoal (Oades, 1988), or all organic materials excluding nondecayed plant and animal tissues, their partial decomposition products, and the living soil biomass (MacCarthy et al., 1990). As suggested by MacCarthy et al. (1990), it is most important that readers establish how particular authors apply the various terms to fully understand and assess the implications of research findings. The definitions of SOM and its components to be used in this chapter have been derived from several sources (Oades, 1988; MacCarthy et al., 1990; Stevenson, 1994; Baldock and Nelson, 2000). The term SOM is used in this work to refer to the sum of all naturally derived organic materials present, and a series of further terms are proposed to define specific components of SOM (Table 11.2).

With the advent of modern elemental analyzers, quantification of the organic fraction of a soil is now often completed by measuring the amount of organic carbon present in a soil. Although this chapter is devoted to the composition, chemistry, and functions of SOM as well as the factors that define SOM content, where particular research studies have focused on the measurement of

TABLE 11.2 Definitions of SOM and Its Components

Component	Definition
SOM	The sum of all natural and thermally altered biologically derived organic materials found in the soil or on the soil surface irrespective of its source, whether it is living or dead, or stage of decomposition, but excluding the aboveground portion of living plants
Living components	
Phytomass	Living tissues of plant origin. Standing plant components, which are dead (e.g., standing dead trees), are included in phytomass
Microbial biomass	Organic matter associated with cells of living soil microorganisms
Faunal biomass	Organic matter associated with living soil fauna
Nonliving components	
Surface plant residues	Organic debris located on the soil surface typically dominated by pieces of plant residues (also referred to as litter)
Buried plant residues	Pieces of organic debris >2 mm found in the soil matrix collected by sieving
DOM	Water soluble organic compounds found in the soil solution, which are <0.45 μm by definition
POM	Pieces of organic debris 53–2000 μm in size with a recognizable cellular structure that are collected on a 53 μm sieve after complete dispersion of a soil (typically dominated by plant residues)
HUM	Organic materials <53 μm remaining after removal of POM and DOM
Resistant organic matter	Highly carbonized organic materials including charcoal, charred plant materials, graphite, and coal with long turnover times
Additional fractions of organic matter	
Macroorganic matter	Fragments of organic matter >20 μm or >50 μm contained within the soil matrix and typically isolated by sieving a dispersed soil
Light fraction	Organic materials isolated from soils by flotation of dispersed suspensions on water or heavy liquids of densities 1.5–2.0 Mg m^{-3}
Nonhumic biomolecules	Organic biopolymers including polysaccharides and sugars, proteins and amino acids, fats, waxes and other lipids, and lignin
Humic substances	Organic molecules with chemical structures, which do not allow them to be placed into the category of nonhumic biomolecules
Humic acid	Organic molecules soluble in alkaline solution but precipitate on acidification of the alkaline extracts
Fulvic acid	Organic molecules soluble in alkaline solution and remain soluble on acidification of the alkaline extracts
Humin	Organic materials that are insoluble in alkaline solution

soil organic carbon (SOC), the discussion will refer to the results obtained in terms of SOC and not SOM. In most instances, SOC and SOM can be used interchangeably; however, in some instances, it is more appropriate to refer to the organic fraction of soils as SOM rather than SOC. For example, when referring to the nutrient content of the soil organic fraction, it would be most appropriate to use SOM since, by definition, SOC in itself only refers to the carbon present and not the associated nutrients.

11.2 Composition of Soil Organic Matter

Soil organic matter is composed of a vast array of different organic residues ranging in size from simple monomeric molecules through complex polymeric compounds to pieces of plant residue >1 cm in size in agricultural systems and potentially >1 m in forests. Independent of size, individual components of SOM can exist at any point along a decomposition continuum ranging from fresh unaltered materials to highly altered materials bearing little chemical or physical similarity to their original source. Additionally, SOM can exist in a free state or be associated to varying degrees with surfaces of mineral particles or buried within aggregations of mineral particles. Such diversity in physical and chemical properties when combined with the spatial heterogeneity of its distribution within the soil matrix at size scales ranging from micrometer to meter makes studies of SOM composition and function challenging.

Often analytical techniques are applied to the whole organic component of soils resulting in the acquisition of weighted average compositional or behavioral data. Such data may or may not provide an adequate assessment of the potential involvement of organic matter in soil processes. For example, the presence of a small amount of biologically labile organic matter could significantly alter short-term measurements of soil carbon mineralization but not be related to a measure of the total amount of organic carbon present. To better understand the nature of SOM and its roles in soil processes, various schemes have been developed to divide SOM into more homogeneous components and to subsequently conduct separate analyses on the isolated fractions.

11.2.1 Elemental Composition

Carbon accounts for the majority of the mass of SOM; however, oxygen, hydrogen, nitrogen, phosphorus, and sulfur can also make significant contributions. Contents of organic matter in

soils are typically quantified by measuring the gravimetric contents of organic carbon and/or total nitrogen. The traditional factor used to convert gravimetric contents of SOC into SOM is 1.72 (Equation 11.1), which assumes a carbon content of approximately 58% by weight. To obtain estimates of SOM content based on the gravimetric content of nitrogen requires the use of a C:N ratio expressed on a mass basis and typically assigned a value of 12 for agricultural soils (Equation 11.2). Spain et al. (1983) summarized the C:N ratios across 3652 surface horizons from Australian soils and obtained ratios ranging from 5 to 35 across all soil types under a variety of management strategies. Snowdon et al. (2005) obtained values ranging from 6 to 51 for Australian forest soils. Higher C:N ratios would be expected where the organic matter in a soil is dominated by less decomposed pieces of plant debris rather than well-decomposed organic matter. Conversely, where well-decomposed materials dominate, lower C:N ratios would be expected. In a review of global C:N ratios for SOM, Cleveland and Liptzin (2007) obtained values ranging from 1.7 to 25.7 when expressed on a gravimetric basis. Average gravimetric C:N ratios for the entire data set (n = 146), grasslands (n = 75), and forests (n = 55) were 12.2, 11.8, and 12.4, respectively, consistent with the typically assigned gravimetric C:N ratio of 12:

$$SOM\,(g\,kg^{-1}\,soil) = 1.72 \times SOC\,(g\,C\,kg^{-1}\,soil), \qquad (11.1)$$

$$SOM\,(g\,kg^{-1}\,soil) = 1.72 \times C:N\,ratio$$
$$\times\,soil\,organic\,nitrogen\,(g\,N\,kg^{-1}\,soil). \quad (11.2)$$

Although C:N ratios have typically been expressed on a gravimetric basis for soils and plant materials, in aquatic and geochemical sciences, elemental ratios have tended to be expressed on a molar basis. Redfield (1958) recognized that marine plankton maintained an average molar C:N:P ratio of 106:16:1 similar to that in marine water. The presence of such a defined relationship for marine systems has been used to constrain marine net primary productivity (NPP) (e.g., Hecky and Kilham, 1988; Turner et al., 2003), and the recognition of defined elemental ratios in components of terrestrial systems is used to constrain nutrient dynamics within carbon and nutrient simulation models (e.g., the Century model; Parton et al., 1987, 1988). Such relationships have led to the development of the concept of ecological stoichiometry in which it is recognized that variations in molar elemental ratios fluctuate within defined boundaries based on environmental conditions and biochemical requirements of organisms. A trend toward expressing elemental ratios on a molar basis is emerging for SOM. Stevenson (1986) provided an estimate of the molar C:N:P:S ratio of SOM as 107:7.7:1:1. Cleveland and Liptzin (2007) defined average C:N:P ratios of soils obtaining values of 186:13:1 for SOM and 60:7:1 for soil microbial biomass with a tendency for grassland soils to be more nutrient rich and forest soils to be more carbon rich. Further quantification of the variance of these ratios for

different components of SOM (e.g., pieces of plant-residue and well-decomposed materials—see Section 11.1) and particular combinations of soil type and land use will aid the development of more accurate predictions of nutrient cycling and potential responses to alterations in management. In particular, improved definition of the net mineralization of nutrients (i.e., the balance between rates of gross mineralization and immobilization) and release into plant-available pools should be obtained.

Molar ratios of carbon, oxygen, and hydrogen can also provide useful information pertaining to chemical changes that occur during decomposition, heating, or other processes where transformations of SOM take place. One useful representation of C:O:H ratios that appears underutilized in SOM studies is a van Krevelen plot (van Krevelen, 1950) in which molar H:C is plotted against molar O:C. Where the elemental composition of SOM components (biomolecules) can be defined, van Krevelen plots can be used to define the possible compositional space in which a sample must reside and whether there is a prevalence of one component over another. Additionally, the trajectory on a van Krevelen plot associated with changes induced by a process of alteration can give an indication of the associated chemical changes (e.g., dehydration, loss of CO_2, or oxidation, which are all associated with different trajectories). Baldock and Smernik (2002) used a van Krevelen plot to define the extent of alteration of natural organic materials during a heating process. When using van Krevelen plots, it is essential that reliable values are obtained for the molar concentrations of O and H. For soils, the presence of hygroscopic water and alterations to soil mineral components during the heating phase of C, O, and H determinations may introduce errors in measurements of O and H concentrations.

Another useful property of SOM that can be calculated from its molar elemental composition is its oxidation state (C_{ox}). The value of C_{ox} for a given organic molecule of the form $C_xH_yO_zN_w$ can be calculated according to Equation 11.3 in which x, y, z, and w represent the molar percentages of C, H, O, and N, respectively, in the molecule or weighted average molecule in the case of SOM. The values for C_{ox} can vary from −4 to +4 but typically vary between −2.2 and +3 in organic molecules. For example, lipids, carbohydrates, and organic acids have C_{ox} values of −2 to −1, 0, and 0 to +3, respectively (Masiello et al., 2008). C_{ox} values of SOM and its components therefore contain information pertaining to molecular composition, biochemical synthesis pathway, extent of decomposition, and diagenetic history. For example, if C_{ox} values <−1 are obtained, then there is likely to be a preponderance of lipid/aliphatic structures present. Alternatively, if values >1 are obtained, then a high content of carboxylic acids is likely. C_{ox} values between −1 and +1 are less informative as a large range of combinations of different compounds could provide such values. However, C_{ox} values may still be used to constrain predictions of molecular composition of SOM and its components.

C_{ox} is mathematically related to the oxidative ratio (OR), which defines the molar ratio of O_2 consumption to CO_2 emission during decomposition. Organic molecules having positive C_{ox} values do not need as much oxygen for complete mineralization as those with negative C_{ox} values. Based on the assumption

that net photosynthesis occurs via Equation 11.4 with all N being obtained as ammonia, the OR can be calculated from C_{ox} according to Equation 11.5. If the required N is obtained as nitrate or through biological fixation of N_2, Equations 11.6 or 11.7, respectively, should be used to calculate OR (Masiello et al., 2008). Quantification of the OR (and thus C_{ox}) of terrestrial forms of organic matter, including that found in soils, can be used to define the fate of CO_2 emitted to the atmosphere by fossil fuel burning (Randerson et al., 2006):

$$C_{ox} = \frac{2z - y + 3w}{x}, \tag{11.3}$$

$$xCO_2 + \frac{1}{2}(y - 3w)H_2O + wNH_3$$

$$\rightarrow C_xH_yO_zN_w + \left(x + \frac{1}{4}(y - 3w) - \frac{z}{2} \right)O_2, \tag{11.4}$$
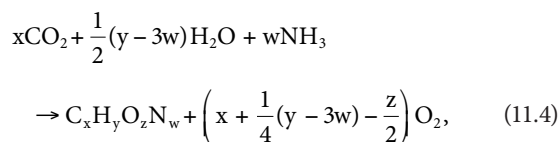
$$OR = 1 - \frac{C_{ox}}{4}, \tag{11.5}$$

$$OR = 1 - \frac{C_{ox}}{4} + \frac{2w}{x}, \tag{11.6}$$

$$OR = 1 - \frac{C_{ox}}{4} + \frac{3w}{4x}. \tag{11.7}$$

11.2.2 Chemical/Molecular Composition

A heterogeneous chemical structure and an ability to form strong associations with soil minerals make the chemical characterization of SOM difficult. Two approaches to define the chemical/molecular composition of SOM have been used: (1) chemical extraction or degradative methodologies and (2) modern spectroscopic techniques capable of analyzing SOM in situ. With both approaches, it is essential to gain a full understanding of the methodology being used and possible deficiencies in order to assess how selective and quantitative the applied procedures are and whether there is a potential for creating artifacts. Without such an understanding, it is not possible to make appropriate interpretations of the SOM compositional data obtained.

11.2.2.1 Chemical Extraction and Degradative Methods

Chemical extraction and degradative methods for characterizing SOM use aqueous solutions and organic solvents to liberate SOM components with particular chemical characteristics. One of the most common approaches uses an alkaline extraction followed by acidification (Schnitzer, 2000) of the extract to isolate humic acid, fulvic acid, and humin (see Table 11.2 for definitions). These three fractions should not be considered as discrete compounds, as each will contain a multitude of different chemical structures that can be further fractionated and purified. Criticisms pertaining to the use of alkaline extraction/acid

precipitation to separate soil organic and mineral components include the following:

1. Questions related to the ability of alkaline extractable material to be representative of the composition of the unextracted and entire SOM fraction
2. The apparent lack of a relationship between the biological functioning of SOM and its alkaline extractability based on C and N isotopic tracer studies (Oades, 1995)
3. Incomplete segregation and probable mixing of SOM with different molecular compositions and susceptibilities to decomposition
4. Creation of artifacts (alteration of molecular structure) during the extraction and precipitation procedures

It is acknowledged that in early studies of SOM, separation of soil organic and mineral components was essential to allow selective characterization of the SOM, and that alkaline extraction provided such a capability. However, with the advent of more selective degradation procedures and modern analytical instrumentation that can use intact soil samples, it is suggested that the use of alkaline extractants as a means of characterizing SOM should be avoided where possible.

Many additional methods have been developed based on various extraction or degradative methods considered to be "selective" for particular molecular components of SOM. Hydrolysis with 6 M HCl or methane sulfonic acid has been used to quantify the proportion of SOM associated with proteins, amino acids, and amino sugars (e.g., Friedel and Scheller, 2002; Martens and Loeffelmann, 2003; Appuhn et al., 2004). Hydrolysis reactions with sulfuric acid have been used to quantify the allocation of SOM to carbohydrate structures (e.g., Rovira and Vallejo, 2000; Martens and Loeffelmann, 2002). The proportion of lignin in SOM has been quantified using methods that attempt to either isolate the intact lignin molecule (e.g., Tuomela et al., 2000) or quantify the monomeric species released by breaking the macromolecular structure into its component monomeric species (e.g., Baldock et al., 1997b; Chefetz et al., 2002; Leifeld and Kögel-Knabner, 2005). Various solvent-based procedures have been developed and used to quantify the presence of lipids and lipid-like organic materials in soils (e.g., Poulenard et al., 2004; Rumpel et al., 2004).

Although more specific than alkaline extraction procedures, results derived from these more specific degradation procedures should be considered as approximate due to the potential for incomplete extraction and/or nonselective action. Using a combination of extraction techniques in a "proximate analysis approach" designed to isolate and quantify klason lignin, Preston et al. (1997) demonstrated the presence of a range of nonlignin forms of carbon. Additionally, the use of degradation procedures to quantify SOM components involved in biological processes within the soil matrix may be questionable. As an example, consider a polysaccharide molecule located in a pore on the external surface of an aggregate versus the same molecule buried within a matrix of clay particles. Both polysaccharides may be broken down by a sulfuric acid digestion procedure and

detected by subsequent analyses, but the availability of these two molecules to microbial oxidation as well as their relative contribution to soil function would be expected to vary significantly.

Attempts to use chemical fractionation procedures to allocate SOM to labile and recalcitrant fractions without defined molecular composition have also been made. Hydrolysis with HCl and permanganate oxidation procedures have been proposed to offer such capabilities. With HCl hydrolysis, the proportion of carbon or nitrogen entering the hydrolysates has been considered to be indicative of the biologically labile component of SOM. Both Leavitt et al. (1996) and Paul et al. (2001) found that carbon remaining in the nonhydrolysable fraction of soil subjected to HCl hydrolysis was older than that found in the total SOC fraction prior to hydrolysis. Quantifying the proportion of SOC oxidized in permanganate solutions of increasing concentrations has also been used to define SOC fractions with different biological liabilities (Blair et al., 1995). However, the existence of strong correlations between the amounts of SOC oxidized at each permanganate concentration and total SOC (Lefroy et al., 1993) question the selectivity of this approach toward identifying differentially labile SOC components (Blair et al., 1995; Mendham et al., 2002). Furthermore, Mendham et al. (2002) showed that permanganate oxidizable SOC had little relation to the labile pool of SOC respired over a 96 day incubation period, again questioning the existence of a link between chemical and biological reactivities. The absence of a strong linkage between chemical and biological reactivities and the lack of a clear definition of the chemical nature of the SOC components attacked by each permanganate solution can limit the ability of chemical fractionation techniques to be used as proxies for biologically meaningful fractions of SOC (Baldock, 2007).

11.2.2.2 Chemical Characterization with Analytical Instrumentation

A variety of analytical techniques exist that can be used to characterize the chemical composition of SOM. In this chapter, four techniques that can be used to quantify the chemical nature of SOM in situ will be examined: ^{13}C NMR spectroscopy, analytical pyrolysis, Fourier transform infrared spectroscopy (FTIR), and x-ray spectroscopy. Examples of the chemical information obtained by applying these analytical techniques to SOM are given in Figures 11.1 and 11.2. The forms of analytical instrumentation to be discussed are all complementary and, where possible, efforts should be made to utilize combinations of these methods to confirm and provide additional insight into the chemical composition of SOM.

11.2.2.2.1 *Solid-State Nuclear Magnetic Resonance Spectroscopy*

Solid-state ^{13}C NMR spectroscopy can be used to define the chemical environment around individual carbon atoms in an intact soil sample. The only pretreatments required are drying and then grinding to ensure homogeneity. A typical cross polarization ^{13}C NMR spectrum for a soil is presented in Figure 11.1a. Organic carbon found in different chemical environments can be differentiated on the basis of chemical shift (expressed in

units of parts per million of the applied magnetic field). Duncan (1987) has presented a comprehensive review of the chemical shift values associated with different types of carbon. Solid-state ^{13}C NMR spectra acquired for SOM are typically divided into chemical shift regions indicative of the major chemical forms that individual carbon atoms can take Figure 11.1a. Integration of the signal intensity within each region provides a quantitative indication of the amount of each form of carbon present in a sample, provided signal acquisition from all forms of carbon in a sample is quantitative. Methods now exist to define the level of quantitative detection of carbon within solid-state ^{13}C NMR analyses of SOM (Smernik and Oades, 2000a, 2000b); however, adoption has been limited. ^{13}C NMR detection efficiencies can be improved by the application of hydrofluoric acid (HF) pretreatments to concentrate carbon and remove paramagnetic species (Skjemstad et al., 1994; Schmidt and Gleixner, 2005).

Although the distribution of ^{13}C NMR signal intensity within various chemical shift regions has been used to infer the molecular composition of organic matter in a sample, ^{13}C NMR gives no direct information pertaining how the different types of carbon present in a sample are arranged into molecules. For example, the presence of signal intensity in the *O*-alkyl region (65–105 ppm) is often ascribed to carbohydrates. However, three of the 11–12 carbons found in lignin monomers will also resonate within the *O*-alkyl chemical shift region as will some carbon associated with proteins and lipid/wax materials. To extend the level of molecular information obtained from the application of NMR to natural organic materials, molecular mixing models have been defined (e.g., Hedges et al., 2002; Baldock et al., 2004; Nelson and Baldock, 2005). In this approach, it is assumed that the organic carbon present in a sample can be described as a mixture of biomolecules (endmembers) with defined representative chemical structures. The molecular composition of ^{13}C NMR observable organic carbon is then estimated by defining the mixture of endmembers that minimizes the sum of squares of differences between predicted and measured distributions of ^{13}C NMR signal intensity. Baldock et al. (2004) and Nelson and Baldock (2005) have used this approach to estimate the molecular composition and elemental stoichiometry of organic materials found in a range of terrestrial and aquatic systems.

Solid-state NMR has also been used to characterize the chemical composition of organic N in soils (Knicker et al., 1993; Clinton et al., 1995; Knicker and Skjemstad, 2000); however, the low NMR sensitivity and natural abundance of the ^{15}N nucleus make the acquisition of spectra with adequate signal-to-noise ratios difficult. The majority of organic N present in soils is thought to reside in protein (amide) or heterocyclic compounds. In most ^{15}N NMR studies, signals originating from amide structures dominate the acquired spectra, with heterocyclic N structures being detected as a small shoulder on the amide resonance (Mahieu et al., 2000). To improve spectral quality, ^{15}N-labeled plant materials have been used in decomposition and composting experiments (e.g., Knicker and Lüdemann, 1995) with the majority of N again being allocated to amide structures. However, the typical use of short contact times (≤1 ms) and high-field spectrometers (≥300 MHz) have been found to discriminate against nonprotonated heterocyclic N
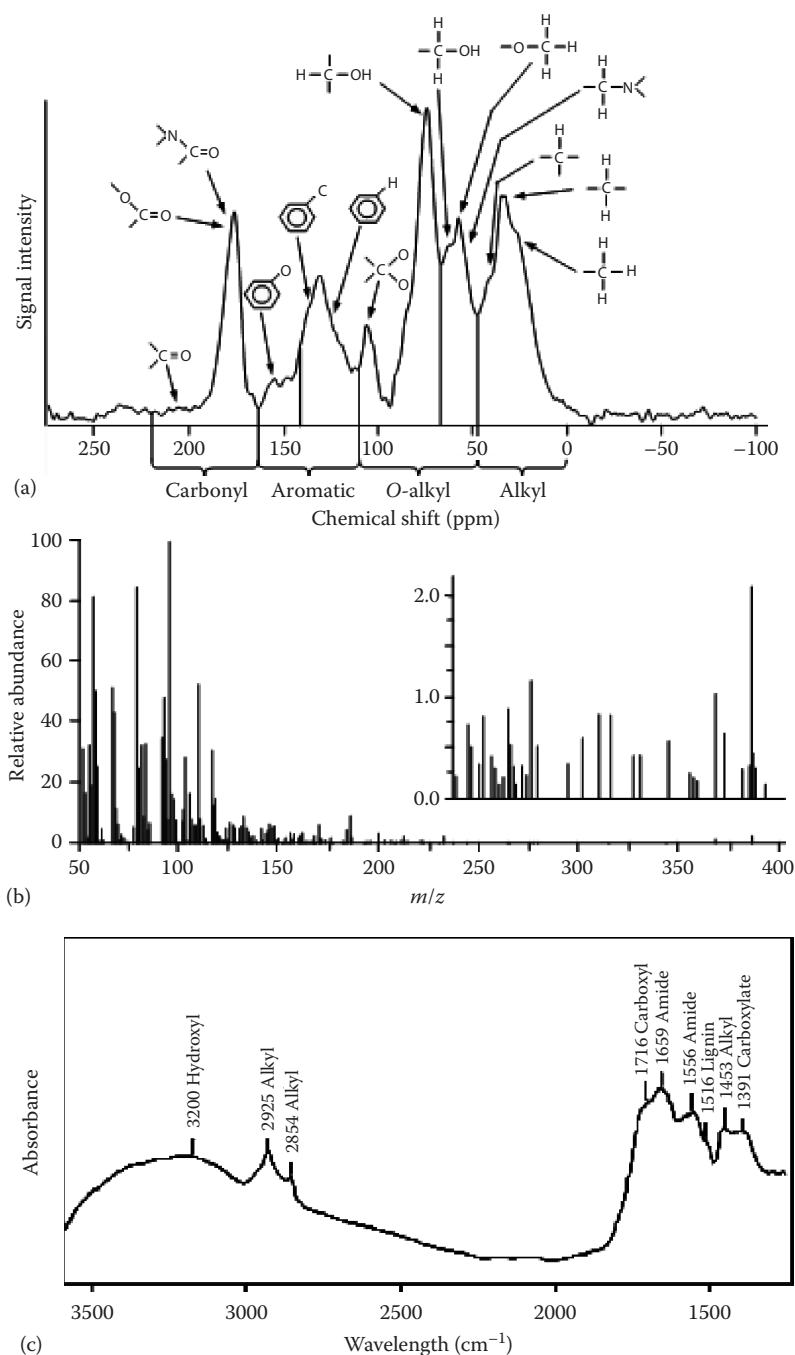
FIGURE 11.1 Examples of the data obtained from (a) a solid-state ^{13}C NMR analysis of soil humus, (b) a pyrolysis-field ionization mass spectrum of a humic acid (Schulten, 1987), and (c) a diffuse reflectance FTIR spectrum acquired for the 0–2.5 cm layer of a mineral soil (52 g C kg^{-1}) after subtraction of mineral-derived signals.

(Keleman et al., 2002). Smernik and Baldock (2005) applied a spin counting technique to ^{15}N NMR analyses of clay-associated organic matter and found that although the major form of N present was amide, up to half or more of the organic N present must have resided in a form that was insensitive to ^{15}N NMR, potentially heterocyclic N. Careful consideration of ^{15}N pulse programs and the use of spin counting techniques are required to obtain ^{15}N NMR results that are indicative of all N present in a sample of soil.

The ^{31}P nucleus is well suited to NMR due to its high natural abundance and relative NMR sensitivity. ^{31}P NMR analyses have been performed on solutions extracted from soils and on solid whole soils (see Toor et al., 2006). Solution-state ^{31}P NMR offers improved resolution and thus identification and quantification of P species; however, it suffers from the requirement to extract P-containing materials into solution and thus may not allow characterization of all P in a sample. Solid-state ^{31}P NMR can be
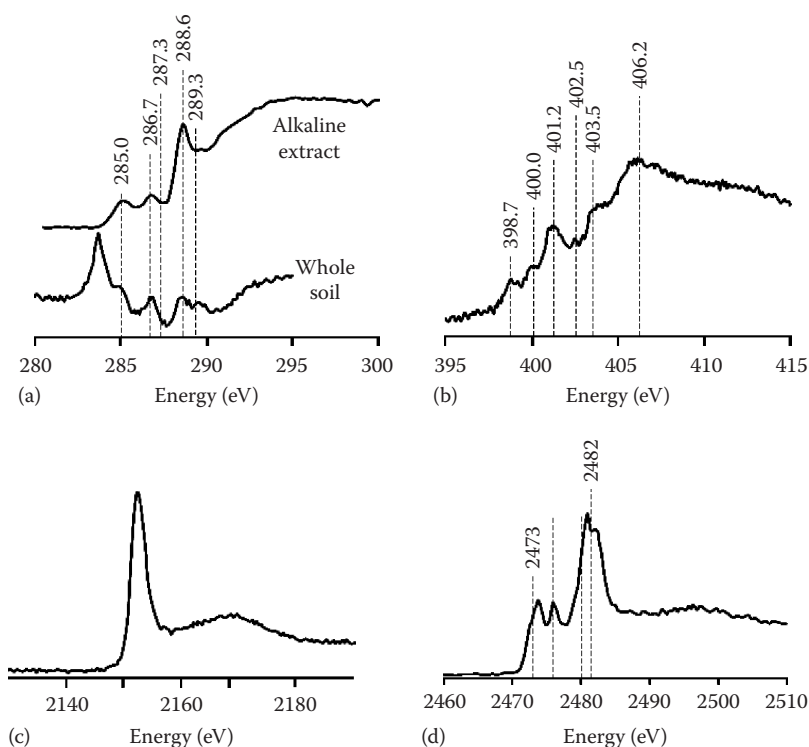
FIGURE 11.2 (a) Carbon K-edge XANES spectra acquired for the alkaline extractable organic carbon from a forest soil in western Kenya (Lehmann et al., 2008) and a whole soil from Elstow (Gillespie et al., 2009). (b) Nitrogen K-edge XANES spectrum acquired for a whole soil from Elstow (Gillespie et al., 2009). (c) Phosphorus K-edge XANES spectrum for topsoil from a German peat (Kruse and Leinweber, 2008). (d) Sulfur K-edge XANES spectrum acquired for whole soil taken from a podzol O horizon (Prietzel et al., 2007).

used to analyze a whole undisturbed sample without extraction; however, the association of soil P with Fe or Mn can render it invisible to ^{31}P NMR. Organic phosphates tend to remain unresolved and appear as a broad resonance between +1 and −1 ppm due to the wide variety of organic phosphate forms (e.g., phospholipids, nucleic acids, metabolic products) and their limited mobility. In solution-state ^{31}P NMR, rapid tumbling of extracted organic P-containing molecules leads to sharper signals. Toor et al. (2006) concluded that solution-state ^{31}P NMR is useful for characterizing the chemical composition of soil organic P, while solid-state ^{13}P NMR is more suited to the study of inorganic P components of soils.

11.2.2.2.2 Analytical Pyrolysis Mass Spectrometry

Analytical pyrolysis characterizes the chemical composition of products of controlled pyrolysis of a sample to obtain information pertaining to the nature of the organic matter present in the sample. The application of this technique to the analysis of organic matter in whole soils and soil extracts is given by Schnitzer and Schulten (1995). Several different approaches to using analytical pyrolysis exist including off-line pyrolysis, pyrolysis mass spectrometry, derivatization pyrolysis mass spectrometry, and Py-GC/MS. A detailed description of each technique is presented by Schulten and Leinweber (1996). In pyrolysis mass spectrometry, the mass/charge ratio (*m/z*) is used to differentiate

the pyrolysis products released during a rapid (Curie-point) or controlled heating of a sample. Several methods of ionization are available to promote the movement of pyrolysis products through the mass spectrometer: field ionization (FI), chemical ionization (CI), fast atom bombardment (FAB), and laser ionization (LI). Irrespective of the method of ionization used, distinctive patterns of pyrolysis products are produced (Figure 11.1b). After normalization of the *m/z* peak intensities, a qualitative assessment of the chemical nature of the organic C in a soil sample can be obtained. Incorporation of a gas chromatograph between the pyrolysis chamber and the mass spectrometer can further aid in the separation of similar pyrolysis fragments prior to detection and analysis by the mass spectrometer.

A cautionary note regarding the comparison of different *m/z* signal intensities is required. First, the intensities observed at any single *m/z* value may result from multiple pyrolysate fragments if they have the same *m/z* ratio. Second, the volatilization of different types of pyrolysis fragments varies (e.g., in pyrolysis FI MS, volatilization decreases with increasing polarity of the fragments). As a result, differences in signal intensities at various *m/z* values do not necessarily correlate with contents of the parent molecules present in the original soil samples. It is appropriate, however, to utilize variations in a given *m/z* signal intensity between samples run under a constant set of analytical conditions to infer compositional differences.

11.2.2.2.3 Fourier Transform Infrared Spectroscopy

FTIR can be used to determine the type of the atoms to which C is bound as well as the nature of the bond. Detailed identification of chemical structure is therefore possible. However, the application of FTIR to characterize the composition of SOM in whole soils is limited by two factors: the low organic matter content of most mineral soils and the presence of significant signals from soil mineral components. The use of "difference spectra" (spectrum of an untreated sample minus that of a pretreated sample) may provide more useful data. Subtraction of an FTIR spectrum obtained for a sample heated to 350°C to remove all organic matter from the FTIR spectrum obtained for the original soil can provide an indication of the nature of the organic components removed, provided no significant alteration to soil mineral components occurred during the heating process. Where significant alterations to soil mineralogy also occur on heating these changes will also be included in the difference spectrum. Figure 11.1c provides an example of a difference spectrum obtained for a soil.

11.2.2.2.4 X-Ray Absorption Spectroscopy

X-ray absorption spectroscopy (XAS) allows collection of atomic scale chemical information. A summary of the fundamentals of XAS techniques is given by Schulze and Bertsch (1995). In XAS, x-rays are absorbed by an atom at defined energies. In response, a higher-energy electron replaces the ejected electron, and excess energy is released in the form of fluorescence. The energy at which fluorescence occurs is characteristic for each electron within an atom and can be used to provide chemical information about a given type of atom (C, N, P, or S). XAS spectra are typically divided into two regions: the near-edge and extended region. X-ray absorption near-edge structure (XANES) spectroscopy examines the region before the absorption edge to around 100 eV past the edge. XANES spectra can also be referred to as near-edge x-ray absorption fine structure (NEXAFS) spectra. XANES spectra are sensitive to the oxidation state of the atom being examined. Resonances are identified by comparison with spectra acquired for standard compounds. In the extended region of x-ray absorption spectra, NEXAFS spectra provide information on coordination number, identity of nearest neighbors, and bond distances for the element being examined. XAS has been used to examine the chemical nature of organic C, N, P, and S found in soils (Figure 11.2).

Solomon et al. (2009) performed C K-edge XANES analysis of a suit of known organic C compounds indicative of those found in soil (carbohydrates, amino sugars, amino acids, phenols, quinine, benzenepolycarboxylic acid, and markers for black C). The acquired spectra showed distinct features and resonance positions for each form of carbon in the known compounds. Lehmann et al. (2008) provided the following broad classifications for XANES C resonances for alkaline extractable SOC: aromatic carbon ring structures (284–286 eV), phenolic/pyrimidine or imidazole carbon (286.4–287.4 eV), aliphatic carbon (287.3–287.8 eV), and carboxylic/amide carbon (288–289 eV). Working with whole soils (Figure 11.2a), Gillespie et al. (2009) identified

additional resonances and used XANES to quantify differences between bulk and rhizosphere soil. The C K-edge XANES spectra acquired by Gillespie et al. (2009) were complicated by the presence of residual C in the beamline and second-order O K-edge features that are typically absent from spectra acquired for pure organic compounds or samples with high (>50 mg C g^{-1} sample) carbon contents. XANES has also been combined with scanning transmission x-ray microscopy (STXM) to define spatial changes in carbon composition across microaggregates (Kinyangi et al., 2006; Lehmann et al., 2008) and black carbon particles (Lehmann et al., 2005; Liang et al., 2006, 2008). Solomon et al. (2009) concluded that C K-edge XANES may provide a means of fingerprinting complex organic C compounds of ecological importance; however, due to the broad spectral features found for organic carbon in soils (e.g., Lehmann et al., 2008), it is unlikely that XANES can be used on its own to provide an unambiguous chemical characterization of SOC.

Gillespie et al. (2009) performed N K-edge XANES on whole soils (Figure 11.2b) and attributed the following resonances to organic N structures: aromatic N in six-membered rings (398.7 eV), nitrilic and pyrazolic N (400.0 eV), amide (protein) N (401.2 eV), pyrrollic N (402.5 eV), nitroaromatic N (403.5 eV), and alkyl N (406.2 eV). Leinweber et al. (2007) presented a detailed systematic analysis of N K-edge XANES resonances obtained for a range of known organic N structures likely to occur in soils and concluded that N K-edge XANES can be used to reliably distinguish pyridinic and nitrile N (≤400 eV) from amide, nitro, and pyrrollic N (>400 eV). Jokic et al. (2004) used N K-edge XANES to detect the presence of heterocyclic N compounds in chemically unaltered whole soil samples.

Most applications of P XANES spectroscopy have focused on the K-edge; however, the speciation of P forms, particularly organic P, can be difficult or even impossible at the P K-edge. Kruse and Leinweber (2008) determined the nature of the P present in a series of inorganic and organic standard compounds as well as the topsoil (Figure 11.2c) and subsoil of a German peat using P K-edge XANES. The spectra obtained for the inorganic and organic standards showed only slight variations in the energy position of the main resonance and thereby offered little potential to distinguish inorganic from organic P. Additionally, little distinction was noted between the three organic P compounds (asolectin, ATP, and phytic acid). P XANES spectra collected at the L$_{2,3}$-edge instead of the K-edge contain more distinguishable features. Kruse et al. (2009) present a detailed analysis of the P L$_{2,3}$-edge XANES spectra obtained for a range of inorganic and organic forms of P. The P L$_{2,3}$-edge XANES spectra contained more spectral features than P K-edge spectra, particularly for the organic P compounds, leading Kruse et al. (2009) to suggest that quantitative differentiation of different types of P may be possible by applying linear combination fitting to P L$_{2,3}$-edge XANES spectra.

K-edge XANES can also be used to examine S speciation in soils (Prietzel et al., 2003; Solomon et al., 2005) (Figure 11.2d). S K-edge XANES resonances have been defined as 2472.6–2473.4 eV (reduced S in organic polysulfide, disulfide, monosulfide, and thiol structures), 2475.8–2481.3 eV (intermediate S

in sulfoxide, sulfone, and sulfonate structures), and 2482.5 eV (oxidized S in ester sulfate structures). Prietzel et al. (2007) used S K-edge XANES spectra to characterize the differences between S in whole soils with their corresponding humic extracts. In all soils examined by Prietzel et al. (2007), S was mainly present as organic mono- and disulfide (27%–52%), ester sulfate (14%–39%), and sulfone (15%–27%) with sulfoxide and sulfite having more minor contributes and an absence of sulfonate and inorganic sulfide. However, inorganic SO_4^{2-} cannot be differentiated from ester sulfate because both species have an equivalent oxidation state of +6. Prietzel et al. (2007) noted that the contribution of reduced forms of organic S (organic mono- and disulfides) increases and the contribution of ester sulfate and SO_4^{2-} decreases with increasing OC content and decreasing O_2 availability.

11.2.3 Physical Composition

The major input of organic carbon into soils results from the deposition of carbon captured by photosynthesis. Deposition within the soil matrix occurs in the form of root residues and the organic materials exuded during root growth. Shoot residues are first deposited onto the soil surface and enter the soil via leaching, in the case of dissolved organic materials (DOM), or through some form of mixing process that can be biotic (bioturbation), abiotic (opening of crack as swelling soils dry), or anthropogenic (cultivation). Approaches used to characterize the physical composition of SOM have been based on solubility, particle size, and density.

11.2.3.1 Dissolved Organic Materials

Dissolved organic materials represents a small proportion of the total organic material present in a soil. However, owing to its mobility and chemical properties, it can contribute significantly to soil processes through the provision of energy to microorganisms, complexation and transport of elements within the soil profile, and dissolution of soil minerals. DOM refers to the organic materials that do not settle out of the soil solution phase under the influence of gravity. It is operationally defined as the organic matter that passes through a 0.45 μm filter. At the upper limit of the DOM size, the distinction between DOM and particulate forms of SOM becomes ambiguous. DOM must be extracted from soils to allow quantitative and qualitative analysis. In the laboratory, DOM has been extracted in leachate collected at the base of undisturbed or disturbed soil columns or by suspension of soil in water or dilute salt solutions and collection of the <0.45 μm fraction (e.g., Dunnivant et al., 1992; Nelson et al., 1993). In the field, DOM can be collected using a variety of in situ devices, such as zero-tension or tension lysimeters using porous cups or plates (e.g., Weihermuller et al., 2007; Sanderman et al., 2008). It is important that the methods used to collect and isolate DOM are clearly specified and to assess the potential impacts that adsorption processes and filtering procedures could have on both the nature and amount of DOM isolated. Reviews pertaining to the collection, analysis, and fate of DOM in soils include Herbert and Bertsch (1995), Kalbitz et al. (2000), Neff and Asner (2001), and Kalbitz and Kaiser (2008).

DOM enters the soil as soluble material in water passing through vegetation, litter layers, or organic-rich soil surface horizons. It can also be generated in situ by adsorption/desorption reactions occurring on mineral surfaces, excretion of organic compounds from plant roots, excretion of metabolic waste products from soil organisms, and excretion and action of extracellular enzymes. DOM is lost through uptake and mineralization by soil organisms, sorption reactions with soil particles, precipitation reactions, and in leachate exiting the soil profile. The concentration of DOM in soil is governed by all these processes, and when expressed in terms of dissolved organic carbon (DOC), values typically range from 5 to 50 mg L^{-1} in surface and litter horizons, and from 0.5 to 5 mg L^{-1} or less in B and C horizons. When expressed on as annual fluxes, Michalzik et al. (2001) presented values of 10–40 g DOC m^{-2} $year^{-1}$ for surface organic horizons of temperate forest soils with an attenuation to 1–10 g DOC m^{-2} $year^{-1}$ through soil C horizons. Hope et al. (1994) suggested that DOC fluxes from terrestrial ecosystems should range from 1 to 10 g DOC m^{-2} $year^{-1}$ based on river DOC fluxes.

Reductions in DOC flux on passage through mineral soils are typically attributed to the processes of biologically mediated mineralization and adsorption to soil minerals. In a review of laboratory incubation studies, Kalbitz and Kaiser (2008) indicated that between 5% and 93% of the DOM present in soil solutions can be biologically mineralized. Kalbitz et al. (2003) concluded that the susceptibility to biological mineralization of DOM derived from forest floors, peats, and A horizons decreased with increasing degree of decomposition of the materials from which it was derived. Aromatic compounds potentially derived from lignin appeared to be the most biologically stable form of DOM (Kalbitz et al., 2003). Other important chemical characteristics of DOM that influence its reactivity, sorption behavior, and susceptibility to decomposition include its molecular size (10^2 to >10^5 g mol^{-1} [Homann and Grigal, 1992]), acidity (6–15 mol C kg^{-1} [Herbert and Bertsch, 1995]), and degree of hydrophobicity.

Differential adsorption of DOC onto soil minerals in two adjacent catchments with similar vegetation but different geological origins was proposed to account for the large differences in DOC concentration in catchment drainage water (Nelson et al., 1993). DOC sorption capacity was shown to vary significantly with soil depth, mineralogy, and organic C content (Guggenberger and Kaiser, 2003). Guggenberger and Kaiser (2003) found that surface-mineral horizons rich in organic carbon had little capacity to adsorb additional DOC (1–2 g DOC m^{-2}) compared to iron- and aluminum-rich B horizons (often >150 g DOC m^{-2}). Such measurements suggest that attenuation of DOC flux with increasing depth is most likely dominated by mineralization in the organic-rich surface horizons and then by adsorption reactions in iron- and/or aluminum-rich B horizons.

11.2.3.2 Particulate- and Mineral-Associated Organic Materials

Other than the organic matter that enters as individual molecules (DOM and root exudates), organic matter enters soil as particles that can vary significantly in size and chemical composition.

Particulate organic matter (POM) is composed of a variety of molecular components (e.g., cellulose, lignin, lipids, proteins) that have different availabilities to processes of decomposition. As POM is decomposed and mixed into the mineral soil, particle size is reduced and the potential for interacting with soil minerals and entering the mineral-associated organic matter fraction increases. A variety of names have been applied to the mineral-associated organic matter. In this work, we will use the abbreviation of mSOM to denote mineral-associated SOM. Interactions between minerals and decomposing SOM will result in SOM/mineral particle associations that increase in density as the extent of association increases. The density will be greatest for individual organic molecules adsorbed onto the surfaces of soil minerals (mSOM). Density will be lowest for large POM particles with only a small mass of mineral particles bound to their surfaces. As a result, methods for extracting, quantifying, and characterizing the POM in soils are based on selective collection of materials with different particle sizes, densities, or a combination of both (Christensen, 1996a, 2001).

A prerequisite to quantifying the allocation of SOM to different particle-size or density fractions is the use of disruptive techniques that separate the soil into primary particles or aggregations of particles. It is essential that any chemically or biologically induced alteration of the SOM or redistribution of SOM in response to soil structural degradation and exposure of mineral surfaces previously inaccessible mineral surfaces be minimized. Thus, the use of strong acids, alkalis, or chemical oxidation pre-treatments and the exposure of disrupted suspensions to high temperatures for prolonged periods should be avoided.

Early approaches to SOM fractionation used a strong initial ultrasonic disruption to disperse soil in an attempt to destroy all aggregation and obtain a suspension of primary particles after which particle-size and density fractionation procedures were initiated (e.g., Baldock et al., 1992). However, Golchin et al. (1994a) and Amelung and Zech (1999) showed that recovery of coarse POM decreased with increasing sonification time or energy resulting in a decrease in the physical size of POM and a redistribution of carbon into finer particle-size classes. Development of a two-step disruption process was demonstrated to minimize these problems (Amelung and Zech, 1999). Free POM not associated with mineral materials was removed subsequent to an initial minimal disruption process in which the integrity of soil aggregates was maintained. A more vigorous second step achieving complete aggregate disruption and dispersion was then applied to release pieces of SOM that were occluded within soil aggregates and allow the separation of this material from mSOM. In its simplest form, this approach results in the isolation of three SOM fractions:

- fPOM—free pieces of organic residue found between soil particles
- oPOM—occluded pieces of organic residue found within aggregations of soil particles
- mSOM—organic matter strongly bound to mineral particle surfaces

Additional approaches to disrupt soils and perform particle-size and density fractionations have used a combination of sodium saturation and physical dispersion methods (e.g., Skjemstad et al., 2004) or have applied a more complex combination of particle-size and density fractionation steps with increasing amounts of disruption and/or increasing density fractionation steps. These latter approaches have been designed to progressively isolate SOM materials from free through occluded to various fractions associated to different extents with mineral components (e.g., Six et al., 2001; John et al., 2005; Sohi et al., 2005; Swanston et al., 2005).

Once isolated from bulk soil, a variety of methodologies can be applied to particle-size and density fractions to quantify rates of turnover and chemical composition. Acquiring $\Delta^{14}C$ measurements for the organic carbon found in 2000-63, 63-2, and <2 μm fractions of soils allowed Trumbore and Zheng (1996) to estimate turnover rates. After normalization of the $\Delta^{14}C$ values to those measured for the 2000-63 μm fraction, both depletions and enrichment of ^{14}C were found with decreases in particle size, indicating a respective increase or decrease in age of the carbon with decreasing particle size. Schöning et al. (2005) measured the percentage of modern carbon in the Ah horizons of luvisols, leptosols, and phaeozems under a European beech (*Fagus sylvatica* L.) forest and found a consistent trend of decreasing amounts of modern SOC with decreasing particle size.

The extent of decomposition and turnover times of organic carbon found in fine (<0.2 μm) and coarse (0.2–2.0 μm) mSOM fractions was assessed by Kahle et al. (2003) using $\delta^{13}C$ and $\Delta^{14}C$ measurements. Organic carbon found in the fine-clay fraction was more enriched with ^{13}C and ^{14}C, suggesting a greater extent of microbial processing but a shorter turnover time than coarse-clay organic carbon. Enrichments in ^{13}C and a decrease in C/N ratio with decreasing particle size were also observed by Amelung et al. (1999). The general lack of consistency with respect to changes in ^{13}C and ^{14}C enrichment with decreasing particle size suggests that different processes of SOC stabilization operate in different soils and that relatively young and potentially labile SOC may be stabilized against mineralization, particularly through the formation of mSOM via interactions with mineral surfaces.

The application of density fractionation, either independently or combined with particle-size fractionation methods, has also been used to isolate and characterize SOC fractions with different liabilities. Trumbore and Zheng (1996) found that SOC in dense soil fractions (>2.0 g cm⁻³) was more depleted in ^{14}C than that found in less-dense fractions (<2.0 g cm⁻³). John et al. (2005) determined the $\delta^{13}C$ of the following four density fractions isolated from a silty soil that had undergone a C3/C4 plant species transition:

- Free-POM <1.6 g cm⁻³ ($fPOM_{<1.6}$)
- Light-occluded POM <1.6 g cm⁻³ ($oPOM_{<1.6}$)
- Dense-occluded POM 1.6–2.0 g cm⁻³ ($mSOM_{1.6-2.0}$)
- Mineral-associated SOM >2 g cm⁻³ ($mSOM_{>2.0}$)

The $\delta^{13}C$ data were used to calculate a mean age of the C in each pool. Elemental analyses showed a decreasing trend in C:N ratio in progressing from the $fSOM_{<1.6}$ through to the $mSOM_{>2.0}$ fractions

that were suggested to indicate an increase in the degree of degradation and humification. The mean age of SOC in these fractions and the values obtained for percent modern carbon (Rethemeyer et al., 2005) did not follow the same trend, suggesting that the oldest carbon in a soil may not be the most decomposed. Such a situation can only arise when organic materials having a chemical composition that renders them biologically labile are stabilized against biological attack within the soil matrix.

A simpler density fractionation scheme using a density solution of 1.7 g cm^{-3} was used by Swanston et al. (2005) to isolate three fractions (fSOM$_{<1.7}$, oSOM$_{<1.7}$, and mSOM$_{>1.7}$) from ^{14}C-labeled and unlabeled forest soils. The fSOM$_{<1.7}$ was found to be the most active fraction based on the measurement of SOC content, C/N ratio, and Δ^{14}C. The oSOM$_{<1.7}$ fraction appeared to be less dynamic with a minimal entry of ^{14}C since the labeling event. Based on ^{13}C NMR analyses (Golchin et al., 1994a; Sohi et al., 2001, 2005; Poirier et al., 2005), oPOM is more degraded compared to fPOM; however, this is not consistent with the higher C/N ratios measured by Swanston et al. (2005). Such high C/N ratios would be consistent with the presence of a significant amount of charcoal C, which could mask the entry of new labeled ^{14}C into this pool based on Δ^{14}C measurements alone. ^{13}C NMR analyses of oPOM fractions shown in Figure 11.3 (Sohi et al., 2001; Poirier et al., 2005) do indeed suggest the presence of charcoal-like carbon given the significant resonances in the vicinity of 130 mg kg^{-1}. A significant movement of new labeled ^{14}C into the dense mineral-associated SOC fraction was also measured by Swanston et al. (2005). The depleted ^{14}C signature of this dense fraction at the near-background control site suggested that at the ^{14}C-labeled site, the dense fraction consisted of at least two different pools of SOM: a fast-cycling pool and an older more stable pool. The presence of a labile pool of carbon within the dense mineral-associated SOM fraction was

supported by the lack of a difference in rate of carbon respiration from free particulate and dense mineral-associated SOM over the first 120 days of an incubation study (Swanston et al., 2002).

Although it is challenging to apply all the methods of analysis described to a soil sample (i.e., elemental contents and ratios, isotopic measurements for establishing turnover rates, and NMR to define chemical composition), doing so allows a more complete picture of the nature and dynamics of SOM fractions.

11.2.4 Biological Composition

Research aimed at examining SOM in a biological context have attempted to differentiate organic matter on the basis of its susceptibility to decomposition and mineralization. Schemes developed to quantify the amount of biologically labile carbon have taken two general approaches: (1) quantification of the amount of carbon or nitrogen associated with the living soil microbial biomass or (2) quantification of a product of a biological process (e.g., CO_2–C mineralized over a defined time interval).

Measurements of soil microbial biomass are assumed to be representative of the total mass of living microorganisms present in a soil. The importance of microorganisms to soil functioning is well recognized (Dalal, 1998; Stockdale and Brookes, 2006). The soil microbial biomass regulates most transformations of SOM, responds rapidly to variations in climate or management, and has been defined as the eye of the needle through which all soil carbon passes (Jenkinson, 1977; Smith and Paul, 1990). Soil microbial biomass is routinely measured using a variety of direct and indirect techniques based on either carbon or nitrogen dynamics. Direct techniques attempt to quantify the amount of carbon released to the soil after a fumigation event designed to lyse living microbial cells. The carbon released is measured by an extraction or incubation technique (Jenkinson, 1976;
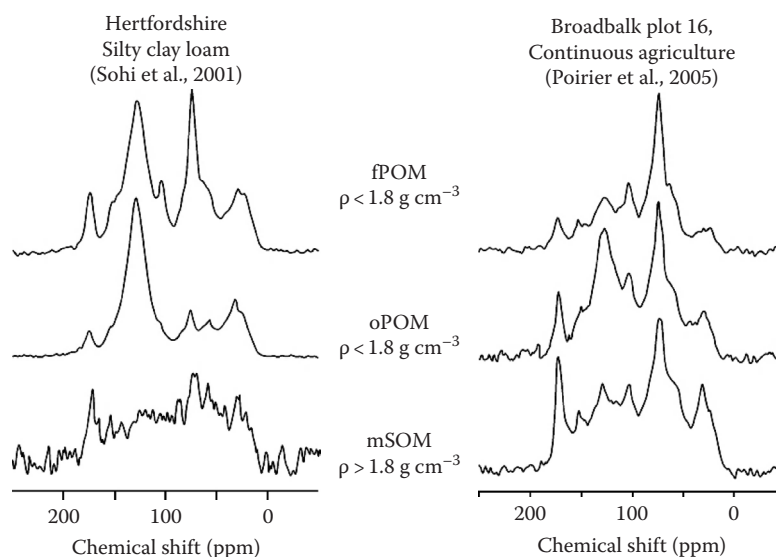
FIGURE 11.3 Solid-state ^{13}C NMR spectra obtained for fPOM, oPOM, and mSOM collected by Sohi et al. (2001) and Poirier et al. (2005) showing strong unsaturated carbon resonances, particularly in the oPOM fraction, consistent with the presence of charcoal and other forms of thermally altered organic matter.

Vance et al., 1987a, 1987b). Indirect techniques involve quantification of the respiration response on addition of a degradable substrate (Anderson and Domsch, 1978). Although the preceding comments have focused on the use of carbon to quantify soil microbial biomass, equivalent approaches exist based on the measurements of nitrogen (Brookes et al., 1985), sulfur (Saggar et al., 1981), and phosphorus (Brookes et al., 1982).

Once the measurements of extracted or mineralized carbon, nitrogen, phosphorus, or sulfur are obtained, conversion factors are applied to account for a lack of complete cell lysis and nonquantitative detection. For example, conversion factors of 0.45, 0.45, and 0.40 have been suggested to convert measurements of extracted carbon, nitrogen, and phosphorus, respectively, into their equivalent values within soil microbial biomass (Brookes et al., 1982, 1984, 1985; Jenkinson et al., 2004). Although these conversion factors have been shown to be adequate for microorganisms in cultures, they may vary across the diverse community of microorganisms found in soils and between different soils. Furthermore, the calculation of the size of the microbial biomass from CO_2 production requires a reliable estimate of the proportion of C that is assimilated and retained by the cell, that is, the microbial efficiency. This microbial efficiency, is not easily quantified, differs for different species within the soil microbial community (Anderson and Domsch, 1973) and varies on the basis of soil properties (Schimel, 1988) and the quality of the decomposing residue (Hart et al., 1994).

Carbon associated with the soil microbial biomass can account for 0.3%–7% of the total SOC (Wardle, 1992). Due to its short turnover time of <1 year (Jenkinson and Rayner, 1977; Jenkinson and Ladd, 1981; Ladd et al., 1981; Jenkinson and Parry, 1989; Wardle, 1992), the carbon associated with soil microbial biomass is considered to be a component of the active or labile pool of SOM. Measurements of soil microbial biomass are considered to provide a sensitive indicator of potential direction and relative magnitude of management-induced changes in SOM (Powlson and Jenkinson, 1976). However, the high spatial and temporal variability in the measurements of the soil microbial biomass and its dependence on soil water content, temperature, and substrate availability in the field will limit the usefulness of single point in time measures as a robust indictor of changes in SOC status. This was shown in an assessment of several Australian field trials, which indicated that very high numbers of field replicates (up to 93) were needed to significantly detect a 20% difference in soil microbial biomass carbon from control treatments (Broos et al., 2007).

There is no doubt that measures of soil microbial biomass provide an assessment of the potential rate of soil biological processes. However, such measures may or may not be related to actual biological process rates. Strong relationships will not exist where only a certain component of the entire microbial community involved in the process being quantified or where all individuals capable of contributing are doing so at different rates.

Quantification of process rates provides an alternative to define the functional capability a soil microbial population. An example of quantifying a process rate is the measurement of the amount of organic carbon mineralized to carbon dioxide over a given time period. Haynes (2005) has reviewed the use of potentially mineralizable carbon and nitrogen to assess the biological lability of SOM. In this approach, the amount of organic carbon or nutrient (typically nitrogen) that can be converted into an inorganic from over a defined time interval is quantified (e.g., Campbell et al., 1991; Franzluebbers et al., 1994). It is important to note that the mineralization of carbon and nitrogen are not analogous. Mineralization of organic carbon results in a gaseous product that, for the most part, cannot be used by soil microorganisms and can be quantitatively recovered in an adequate experimental apparatus. Alternatively, organic nitrogen that has been mineralized to ammonium can be reused (immobilized) by soil organisms, volatilized as ammonia under certain soil pH conditions, nitrified to nitrate, and subsequently denitrified or leached if open incubation systems are used. Estimates of gross nitrogen mineralization are required to obtain data analogous to carbon mineralization (Bengtsson et al., 2003; Murphy et al., 2003; Flavel and Murphy, 2006).

Mineralization of carbon and nutrients results from a complex set of biochemical processes conducted by a wide range of organisms and thus provides a measure of soil functional capacity. Organic C mineralization is often called "soil respiration," "basal respiration," or "microbial respiration." The amount or rate of C mineralization measured over periods from a few days to a few weeks is commonly used as an indicator of biological activity, whereas the total amount of CO_2–C released on a longer time frame (>3 months) is considered to provide information about the fraction of SOC that is readily available to decomposer organisms. Mineralized C can be expressed either per unit mass of soil (mg CO_2–C g^{-1} soil) or as a proportion of the original SOC present (mg CO_2–C g^{-1} SOC). When expressed per unit mass of soil, information regarding the size of the mineralizable C fraction is obtained; whereas, when expressed per unit mass of SOC, an indication of the degradability of the organic carbon present in a soil is obtained.

Although measurements of mineralizable nitrogen require careful interpretation, the measure of net nitrogen mineralization (change in inorganic nitrogen status through time) can provide a measure of the contribution that decomposition processes can make to the supply of plant-available N. Mineralizable soil organic N (SON) is composed of various organic substrates, including microbial biomass, residues of recent crops (mainly POM), and humus (HUM). The mineralizable SON can be measured as potentially mineralizable N using an aerobic incubation under optimum moisture and temperature conditions (Franzluebbers et al., 1994; Chan et al., 2002) or under field conditions (Dalal et al., 2005) using the method developed by Raison et al. (1987). Attempts to correlate mineralizable SON with measures of total SON or SOC have not been very successful. It is suspected that one of the major reasons for this lack of success is related to differences in the allocation of SON and SOC to different forms of SOM, the variability of C:N ratio of these materials and the impact this would have on the mineralization/immobilization balance.

11.2.5 Defining Biologically Relevant Soil Organic Matter Fractions

Variability in elemental, chemical, physical, and biological composition provides the different components of SOM with an ability to contribute to a range of soil properties and processes (see Section 11.4). This variability also leads to differences in rates of turnover as demonstrated using ¹⁴C measurements (e.g., Ladd et al., 1981; Anderson and Paul, 1984; Swanston et al., 2005). In the context of SOM turnover studies, one objective of fractionating SOM is to quantify the allocation of SOM to materials that are differentially available to decomposition. In this section, a conceptual model of SOM decomposition will be presented and used to identify biologically relevant SOM fractions that can be quantified through measurement.

Prior to the presentation of the model, decomposition and its component processes will be defined to ensure no ambiguity exists pertaining to the processes being referred to. Decomposition is used to define the alteration of the original form of an organic material. Mineralization is used to define the conversion of an organic form of an element into an inorganic from (e.g., organic carbon to CO_2 or organic N to NH_3). Mineralization of carbon and nutrients is thought to dominantly occur within cells in response to respiration and other metabolic processes. The extent to which mineralization, particularly nutrient mineralization, can occur external to cells remains unquantified. However, if mineralization external to the cell is a dominant process, then the concept of ecological stoichiometry would be flawed and constraining carbon and nutrient cycling simulation models with carbon to nutrient ratios would be unsuccessful. Assimilation is defined as the use of carbon and

other elements in an organic material to create new biochemicals required for maintenance and growth of the decomposer community. As some of the assimilated carbon and nutrients may be excreted in an organic form as metabolic waste products, net assimilation should be used to refer to the amount of carbon and nutrients taken up and retained within the decomposer community. Additionally, since organic materials in soils are acted on by extracellular enzymes, it is possible that substrate alteration can occur without subsequent assimilation or mineralization. Possible examples of this include the adsorption of products of extracellular enzyme activity on soil mineral surfaces or the leaching of these products through the soil profile before they can be used. In both instances, the original organic material is altered but not accompanied by either a mineralization or assimilation. In the context of these definitions, decomposition is taken to represent the total of the individual processes of mineralization, assimilation, and alteration.

In the decomposition model (Figure 11.4), all processes occur in the aqueous phase of the soil and the behavior of simple and complex substrates is differentiated. Simple substrates are defined as soluble molecules that can cross cell membranes without further alteration. Once inside a cell, simple substrates can be mineralized, assimilated, or transformed into products of metabolic activity that are excreted back into the soil solution. Excreted metabolic products may be further transformed in the soil solution and/or reused by individuals within the decomposer community. Complex substrates consist of polymeric or multicomponent mixtures that are not soluble and cannot pass through a cell membrane in their original state.

Additional processes are required for decomposer organisms to use complex substrates including excretion of extracellular
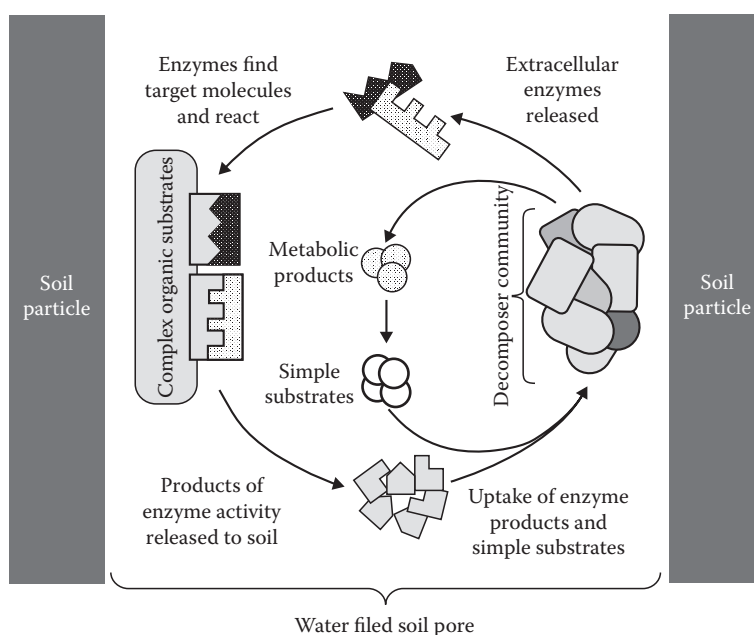
FIGURE 11.4 Schematic model of the decomposition of organic materials in soil. Substrates may exist as simple soluble molecules or in particulate form requiring enzymes to liberate molecules capable of passing cell membranes. Products of metabolic activity may or may not serve as subsequent sources of either energy or nutrients.

enzymes, diffusion of the enzymes to their target sites, release of the products of enzyme activity, diffusion of the products back to the decomposers, and uptake of the products. For both simple and complex substrates, diffusion of soluble materials is required, and thus, the relative position of the substrate and the organism is important. Minimizing the path lengths of diffusion of both enzymes and products will enhance the rates of substrate decomposition and assimilation. Although each of the various organic components identified in Figure 11.4 (extracellular enzymes, products of enzyme activity, simple substrates, and metabolic products) is grouped together in the figure, this is only to illustrate the components involved in the process of decomposition. In reality, all these components will diffuse throughout the soil solution.

Baldock (2007) indicated that two types of controls operate to determine the biological stability of organic materials in soils (Figure 11.5). Biological capability controls whether or not a particular form of organic matter can decompose, while biological capacity controls the rate at which decomposition will proceed. Controls over the biological capability were suggested to include three factors: chemical recalcitrance, decomposer capability, and protection by the soil matrix. Chemical recalcitrance describes the type and arrangement of atoms and bonds within an organic substrate. Some organic material in soil offers little chemical recalcitrance to decomposition (e.g., cellulose and proteins) while others contain components that are highly resistant at least on a timescale of decades (e.g., charcoal). Decomposer capability refers to the presence of the appropriate DNA sequences required to construct the enzymes needed to attack and decompose the molecular components within the organic materials present. The third factor is protection by the soil matrix, which describes whether a molecule is in a position in the soil where it is accessible to enzymes. Biological capacity

FIGURE 11.5 Factors important to defining the biological stability of organic materials in soil. Factors describing biological capability define whether a particular form of SOM can be decomposed or not. Factors describing biological capacity define the rate at which decomposition occurs rather than whether it will or will not occur. Rate modifiers include properties such as temperature and water content, which need to be combined with the duration of exposure to favorable conditions to define the biological capacity. (Modified from Baldock, J.A. 2007. Composition and cycling of organic carbon in soil, p. 1–35. *In* P. Marschner and Z. Rengel (eds.) Soil biology. Vol. 10. Nutrient cycling in terrestrial ecosystems. Springer-Verlag, Berlin, Germany.)

refers to processes that affect the rate of decomposition but not whether a given molecule is decomposable. Biological capacity is defined by a series of factors governing the rate of biological processes (e.g., soil temperature and water content) and the duration of exposure to favorable conditions.

Combining the model of decomposition processes (Figure 11.4) with the concepts of biological stability (Figure 11.5) provides a means to identify biologically relevant fractions of SOM that decompose at different rates. DOM contains a variety of molecular structures with different chemical recalcitrance that are free to diffuse throughout the soil solution. Given the requirement for all substrates to pass through an aqueous phase prior to uptake and utilization by the soil decomposer community, DOM is likely to be the best indicator of the instantaneous availability of organic material at any given point in time. However, quantitative estimates of DOM content are unlikely to provide a good indication of the longer-term biological availability of SOM.

One of the most important forms of SOM, from the point of view of provision of a longer-term supply of energy to soil organisms, is the POM fraction with its high content of more easily degradable material (e.g., cellulose). POM may or may not provide nutrients depending on its elemental ratios. POM with high C:nutrient ratios (e.g., C/N ratio >40) will tend to immobilize nutrients as its carbon is mineralized. This may reduce the availability of nutrients to plants, but also enhance nutrient retention within surface soil layers by limiting losses due to leaching. The mSOM fraction tends to have a chemical composition more indicative of decomposed materials [lower C:nutrient ratios and higher ratio of alkyl:*O*-alkyl carbon (Baldock et al., 1997a)]. When considered in conjunction with its high level of interaction with mineral surfaces, the mSOM fraction is more resistant to decomposition than the POM fraction. The low C:nutrient ratios of mSOM suggest that it can contribute significantly to the provision of nutrients to plants and decomposer organisms. An additional fraction of SOM that is important to consider and monitor is charcoal or black carbon. This material has been found to account for up to 60% of the carbon in a soil (Skjemstad et al., 1996, 1998, 1999, 2002; Schmidt et al., 1999; Skjemstad and Taylor, 1999). Although a fraction of newly created charcoal may be decomposable (Hamer et al., 2004; Marschner et al., 2008), a high recalcitrance of charcoal C to biological mineralization in laboratory-based incubation has also been demonstrated (Baldock and Smernik, 2002). Long-residence times measured for charcoal in soil also suggest that a significant proportion is resistant to decomposition (Pressenda et al., 1996; Skjemstad et al., 1998; Swift, 2001; Krull et al., 2003, 2006).

Given the more recalcitrant behavior of charcoal carbon than the other forms of soil carbon, it is important to be able to selectively quantify the amount of charcoal carbon. In Figure 11.6, the impact that the presence of increasing proportions of charcoal carbon can have on SOC dynamics is shown using a modeling exercise completed with a modified RothC soil carbon model. In this modified model, the original resistant plant material (RPM)

	Proportion of SOC allocated to charcoal			
	0.00	0.10	0.25	0.50
Total organic carbon	55.3	62.6	73.5	91.6
Non-charcoal organic carbon	55.3	49.8	41.5	27.7

FIGURE 11.6   Changes in total organic carbon (TOC), 2.0–0.053 mm POC, carbon associated with the <0.053 mm fraction (HUM), and charcoal carbon (Char-C) estimated from the RothC soil carbon model with different initial contents of charcoal carbon. For this modeling exercise, the total organic carbon at the start of the simulation was kept constant, the fraction of noncharcoal carbon allocated to the POC and HUM fractions was kept constant, and carbon inputs over the simulation period were set to zero. The DPM and two biomass pools in the RothC model were included in the modeling exercise but are omitted from the figure. The 1970–2005 average climate data for Beverly, WA, Australia were used. The table at the bottom of the figure gives the final values obtained after 50 years of simulation for total organic carbon and the noncharcoal organic carbon (POC + HUM) in units of Mg C ha^{-1} for the 0–30 cm soil layer.

and HUM fractions are equated to the POM and mSOM fractions described above. For all four modeling scenarios, the same starting level of SOC was used (127 Mg C ha^{-1}), and the allocation of carbon to each of the noncharcoal fractions was fixed at a given percentage of the total noncharcoal C components (27% for POC and 69% for mSOM). Over the 50-year period, inputs of carbon were set to 0. With increasing initial allocation of SOC to the charcoal fraction, the decline in TOC over the simulation period was reduced. However, when the dynamics of the POC and mSOM fractions are examined, despite there being more total carbon where charcoal accounted for 50% of the initial SOC, the amount of noncharcoal C remaining after 50 years was largest in the soil with no charcoal. If soil productivity was related more to the amount of mSOM carbon than charcoal carbon present, despite the greater loss of carbon from the soil with no charcoal, a higher level of productivity would be retained in this soil due to the higher amounts of mSOM maintained.

As a result of the preceding discussion, selective quantification of biologically relevant fractions of SOM requires an allocation of SOM across POM, mSOM, and charcoal fractions for longer-term studies (weeks to years) and an additional inclusion of DOM for short-term studies (hours to days). Although many approaches varying in number of fractions isolated and complexity of the fractionation process exist, it is suggested that the most simple fractionation system capable of allocating SOM to the POM, mSOM, charcoal fractions, and possibly DOM, depending on the application, would be most practical for the

potential broadscale application of quantifying the implications of land use and land-use change on SOM cycling. Baldock and Skjemstad (1999) proposed a three-component fractionation scheme to identify measurable SOM fractions for soil samples sieved to <2 mm: POC (organic carbon associated with 2000-53 μm particles), HUM (organic carbon associated with <53 μm particles), and ROC (resistant organic carbon associated with <53 μm particles after removal of nonresistant materials with UV photo-oxidation and correction with ^{13}C NMR analyses). Recently, modifications have been made to the Baldock and Skjemstad (1999) fractionation scheme to cover all forms of SOM at a given location and to account for recent analyses that have indicated the presence of significant quantities of charcoal in the POM fraction (Figure 11.7).

11.2.6 Consistency between Biologically Relevant Soil Organic Matter Fractions and Pools of Carbon in Simulation Models

Simulation models of SOC cycling (e.g., Rothamsted [Jenkinson et al., 1987], Century [Parton et al., 1987], and APSIM [McCown et al., 1996]) are often based on conceptual pools of carbon that are not measured directly. The construct of most SOC models is similar and includes fractions of SOC with a rapid turnover (annual), moderate turnover (decadal), and slow turnover (millennial) as well as a passive or inert component. It has been
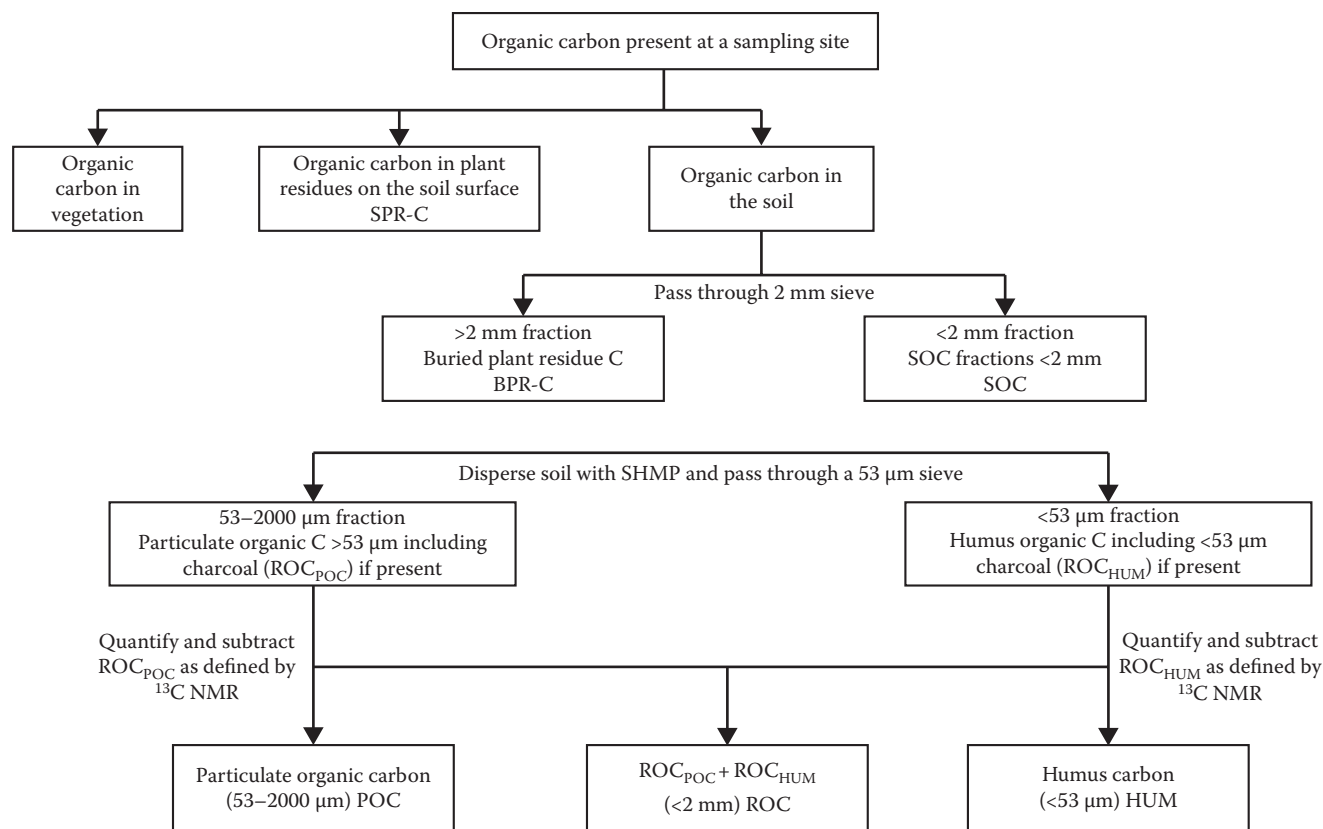
FIGURE 11.7 Schematic presentation of the proposed methodology for defining the allocation of total SOC to its ecologically significant component fractions. SPR-C is the organic carbon associated with plant residues on the soil surface collected on an area basis. BPR-C is the organic carbon in buried plant residues having a size >2 mm. SOC is the total amount of organic carbon found in the <2 mm fraction of the soil. POC is the organic carbon found in the noncharcoal component of particles 53–2000 μm in size. HUM is the organic carbon found in the noncharcoal component of particles <53 μm. ROC is the resistant organic carbon obtained by adding together the charcoal C found in particles <2 mm (the sum of charcoal C found in the 53–2000 μm and the <53 μm fractions). SHMP is sodium hexametaphosphate.

recognized that developing a capability to replace these conceptual pools of SOC with measurable pools would offer several advantages: (1) internal verification of appropriate allocations of SOC to pools, (2) greater mechanistic understanding of the implication of management and environment on the components of SOC most affected, and (3) improved confidence in simulation outcomes. A suitable fractionation procedure should be capable of isolating and quantifying the allocation of SOC to pools that differ significantly in their biological availability (Baldock, 2007).

Several methods have been proposed to link measurable fractions of SOC to the conceptual pools contained within models using density fractionation techniques (Christensen, 1996b; Poirier et al., 2005; Sohi et al., 2005). However, the biological availability of the carbon in each fraction of these studies was never measured, and no attempt was made to substitute the measurable pools of SOC into a working carbon simulation model to demonstrate the utility of this proposal. Skjemstad et al. (2004) showed that the pool structure of the RothC model could be approximated using a three-component fractionation scheme as described by Baldock and Skjemstad (1999). Skjemstad et al. (2004) demonstrated that the POC, HUM, and ROC fractions

could be used to replace the RPM, HUM, and inert organic matter (IOM) pools of the RothC model, respectively. This was an important step forward in simulating SOC dynamics and demonstrated the potential for "modeling the measurable" (Christensen, 1996b; Magid et al., 1996; Baldock, 2007).

It must be noted that the soil microbial biomass (BIOf and BIOs in the RothC model) and decomposable plant materials (DPM in the RothC model) were not included in the fractionation scheme described by Baldock and Skjemstad (1999) or used in the RothC calibration by Skjemstad et al. (2004). The main reasons for this were as follows:

1. Difficulties associated with quantitatively separating the microbial biomass from the other forms of SOC
2. The questionable link between measures of microbial biomass carbon and rates of mineralization of SOC
3. The contributions made by the DPM, BIOf, and BIOs fractions to the total SOC are small and within the errors associated with measurement of the other larger fractions
4. These fractions equilibrate quickly in SOC simulation models

11.3 Quantifying Soil Organic Matter Content and Allocation to Fractions

Soil organic matter content, as indicated in Section 11.2, is typically quantified by measuring the content of organic carbon. Initially, measurements of SOM or SOC were completed to investigate pedogenic processes and, given its diverse and important functions in soils (Section 11.1), to provide a measure of soil productivity. However, being able to accurately define the content of organic carbon in soil has become important because of the potential for soils to sequester atmospheric CO_2–C (Section 11.1). An additional requirement to move beyond the measurements of total organic carbon content and define the allocation of organic carbon to its ecologically significant fractions has emerged to support and improve predictions related to the influence of land use and management on soil carbon dynamics.

11.3.1 Direct Measurement of Soil Organic Content and Component Fractions

Methods to measure the content of SOC have been assembled recently by Skjemstad and Baldock (2008) and previously by Nelson and Sommers (1996) and include dry and wet combustion methods. Total C analyses involve the complete conversion of all C (organic and inorganic) in a soil to CO_2 and quantification of the evolved CO_2 by various means (e.g., infrared detection, increased mass of an ascarite trap, or others). Corrections for the presence of inorganic C must be performed when using total C methods to quantify SOC contents. Of the methodologies currently available, a dry combustion-automated analyzer that measures CO_2 evolution with an infrared detector is the preferred methodology for determining SOC, provided accurate estimates of inorganic C contents can be obtained where required or samples can be pretreated to remove all inorganic C (Baldock and Skjemstad, 1999).

The successful calibration of the RothC model using measurable fractions of SOC has led to the development of the protocol described in Figure 11.7. This methodology differs slightly from that presented by Baldock and Skjemstad (1999) in that the resistant fraction (charcoal) is quantified in both the 2000-53 µm particulate and <53 µm HUM fractions rather than just the HUM fraction. Once isolated and dried, the carbon content of the various fractions (SPR-C, BPR-C, SOC, 53–2000, and <53 µm) is quantified using the dry combustion technique. The amounts of carbon associated with the POC, ROC, and HUM fractions are then defined after using ${}^{13}C$ NMR to quantify the allocation of carbon to the ROC fractions in the 53–2000 and <53 µm materials (ROC$_{POC}$ and ROC$_{HUM}$, respectively).

11.3.2 Proximate Analyses

The SOM fractionation process described in Section 11.1 is time-consuming to complete (3–6 weeks to progress a sample from start to finish depending on the contents of carbon present in the soil and its fractions), requires the use of expensive analytical instrumentation (e.g., a ${}^{13}C$ NMR spectrometer and a UV photo-oxidizer system), and involves the use of hazardous chemicals (e.g., HF). It would therefore be unlikely that this SOM fractionation process would move beyond a research implementation and become a routine measure available to land managers interested in understanding the impact of management practices. Provision of a more cost-effective and rapid means of quantifying SOC content and its allocation to component fractions is required. The development of such a capability would also be very useful to support the carbon cycling work and scenario predictions by providing appropriate data for initializing soil carbon models.

Diffuse reflectance mid-infrared spectroscopy (MIR) offers a simple, rapid, and low-cost methodology that is sensitive to both mineral and organic materials present in soils. In MIR, the chemical bonds associated with a variety of organic functional groups (alkyl, carbohydrate, carboxyl, amide, amine, aryl) can be identified (Janik et al., 2007). The problem is that signals from these components are often hidden within soil MIR spectra by signals derived from soil mineral components. The development of partial least squares (PLS) regression approaches and their application to spectroscopic data has allowed the signals obtained throughout the entire MIR spectrum to be examined for correlation to a set of analytical values derived from traditional laboratory procedures. The ability of MIR/PLS to predict total organic carbon was demonstrated by Janik and Skjemstad (1995) and Janik et al. (1998). Using the PLS process detailed by Haaland and Thomas (1988), Janik et al. (2007) applied the MIR/PLS approach to the data collected for total organic carbon content and its allocation to the POC and ROC (charcoal C) fractions. Successful calibration was achieved and used to compare laboratory-derived analytical data with corresponding data derived from the MIR/PLS predictions (Figure 11.8). The predictive capability for total organic carbon was excellent and that for ROC (charcoal C) was good. Agreement between predicted and measured values for POC was not as good but remained acceptable. Given the stronger dependence of POC on the nature of the vegetation from which it was derived, it is perhaps not surprising that predictability of this fraction was lower. With time and the development of additional calibration data, it may be possible to define more specific calibrations for POC derived from soils under different vegetation types. The current approach defined by Janik et al. (2007) calculates the allocation of carbon to the HUM fraction by difference (HUM = total – particulate – resistant).

It should be recognized that the MIR/PLS process is completely empirical and may be based as much on positive correlation with MIR signals from organic materials as positive or negative signals derived from particular soil minerals. Subsequent unpublished MIR/PLS work completed in the same laboratory has indicated that although the calibrations presented by Janik et al. (2007) provided adequate estimates of soil carbon fractions, improved predictions could be obtained by performing sample set–specific calibrations. To perform these sample set–specific
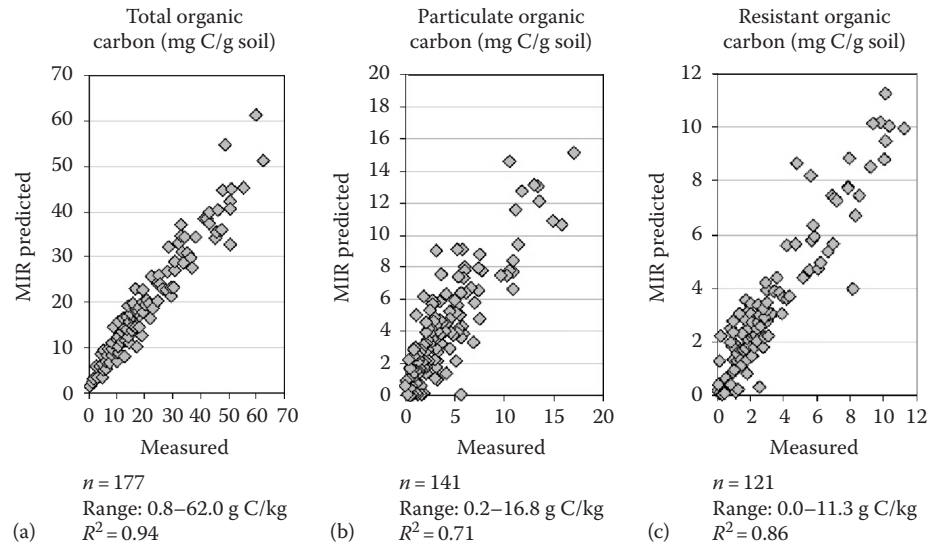
FIGURE 11.8 Relationship between measured and MIR/PLS predicted contents of (a) total organic carbon, (b) POC (53–2000 μm), and (c) ROC (dominated by charcoal). (Adapted from Janik, L.J., J.O. Skjemstad, K.D. Shepherd, and L.R. Spouncer. 2007. The prediction of soil carbon fractions using mid-infrared-partial least square analysis. Aust. J. Soil Res. 45:73–81.)

calibrations, a subset of 10%–20% of the samples were selected to span the range of predicted values, analyzed using the laboratory procedure and used to recalibrate the MIR/PLS estimation procedure.

The MIR/PLS procedure appears to offer a robust, rapid, and cost-effective way to predict the total amount of organic carbon in soils as well as its allocation to fractions. With further testing and development of calibration data sets, this methodology may be able to underpin soil carbon modeling activities and provide useful data to land managers.

11.4 Functions of Organic Matter in Soil

Despite its often minor contribution to the total mass of mineral soils, SOM can influence a variety of soil properties, ecosystem functioning, and the magnitude of various obligatory ecosystem processes (Table 11.3). The properties influenced by SOM have been classified into three groups: biological, chemical, and physical. It should be noted that strong interactions and interdependencies exist between these groups. For example, the ability of SOM to chelate multivalent cations can affect its potential to

TABLE 11.3 Properties and Functions of Organic Matter in Soil

Property	Function
Biological properties	
Reservoir of metabolic energy	Organic matter provides the metabolic energy, which drives soil biological processes
Source of macronutrients	The mineralization of SOM can significantly influence (positively or negatively) the size of the plant-available macronutrient (N, P, and S) pools
Ecosystem resilience	The buildup of significant pools of organic matter and associated nutrients can enhance the ability of an ecosystem to recover after imposed natural or anthropogenic perturbations
Physical properties	
Stabilization of soil structure	Through the formation of bonds with the reactive surfaces of soil mineral particles, organic matter is capable of binding individual particles and aggregations of soil particles into water-stable aggregates at scales ranging from <2 μm for organic molecules through to mm for plant roots and fungal hyphae
Water retention	Organic matter can directly affect water retention because of its ability to absorb up to 20 times its mass of water and indirectly through its impact on soil structure and pore geometry
Thermal properties	The dark color that SOM imparts on a soil may alter soil thermal properties
Chemical properties	
CEC	The high-charge characteristics of SOM enhance retention of cations (e.g., Al^{3+}, Fe^{3+}, Ca^{2+}, Mg^{2+}, NH_4^+, and transition metal micronutrients)
Buffering capacity and pH effects	In slightly acidic to alkaline soils, organic matter can act as a buffer and aids in the maintenance of acceptable soil pH conditions
Chelation of metals	Stable complexes formed with metals and trace elements enhance the dissolution of soil minerals, reduce losses of soil micronutrients, reduce the potential toxicity of metals, and enhance the availability of phosphorus
Interactions with xenobiotics	Organic matter can alter the biodegradability, activity, and persistence of pesticides in soils
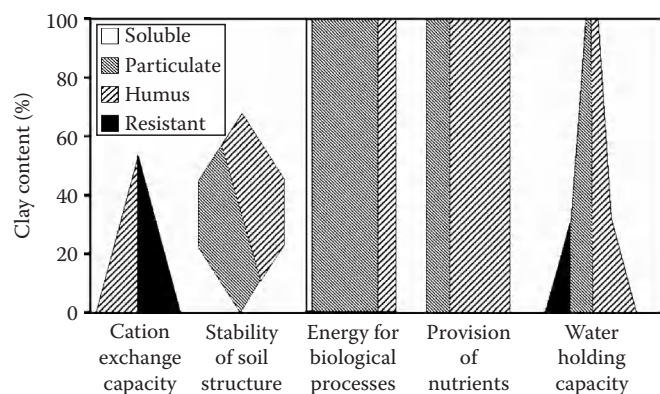
FIGURE 11.9 Conceptual representation of the contribution that SOM and its component fractions make to some of the soil properties to which organic matter makes a positive contribution. Changes in the width of the overall shape with changing clay content define the relative contribution made by organic matter to the function defined. Within each shape, the width of the various shadings defines the importance of each type of organic matter to the function defined. (Adapted from Krull, E.S., J.O. Skjemstad, and J.A. Baldock. 2005. Functions of soil organic matter and effects on soil properties, p. 107. Cooperative Research Centre for Greenhouse Accounting, Canberra, Australia.)

stabilize soil structure and also its biodegradability. In addition, the effects of SOM on soil properties often involve interactions with the soil mineral fraction. Thus, variations in SOM function across different soils may not be solely a consequence of qualitative or quantitative variations in the soil organic component but may also arise in response to changes in soil clay content. In Figure 11.9, an attempt has been made to conceptualize this idea using a series of shapes to illustrate the selective importance of SOM fractions in performing specific functions. In Figure 11.9, the width of the overall shape and that of the SOM components are meant to provide an indication of the relative importance to the function identified at the base of the figure. As the shapes widen, the relative importance increases and vice versa. Shapes have been used in this conceptual framework due to the absence of strong quantitative relationships linking SOM content and composition to the functions identified. The generation of quantitative relationships linking SOM content and composition to functions performed in soil should form the basis for future research projects.

11.4.1 Biochemical Functions

11.4.1.1 Reservoir of Metabolic Energy

The most fundamental function of SOM is the provision of metabolic energy, which drives soil biological processes and the direct and indirect effects, which this has on other soil properties and processes. Photosynthesis fixes CO_2 into glucose, which is then converted into a wide range of organic compounds (e.g., cellulose, hemicellulose, lignin, lipids, and proteins) by various enzymatic processes. The C fixed into such compounds is deposited in or on the soil during plant growth and as the plant or a portion of its tissues senesce, thereby providing C substrates for soil decomposer organisms. Oades (1989) presented an estimate of the C flow

through fertile grassland where primary production via photosynthesis was 10 Mg C ha^{-1} year^{-1}. Of the 3 Mg C ha^{-1} year^{-1} that was added to the SOC fraction, the soil fauna were estimated to utilize 0.3–0.45 Mg C ha^{-1} year^{-1}, while the soil microbial biomass was estimated to utilize 2.4 Mg C ha^{-1} year^{-1}. The majority of SOM processing is thought to be completed by the soil microbial biomass. However, other activities of the soil fauna enhance the ability of soil microbial decomposers to utilize organic residues added to soil. These include (1) fragmentation of plant debris, which enhances the surface area per unit weight of plant residue available to microbial attack and (2) distributing organic materials throughout the soil matrix, which provides an avenue for greater contact between decomposer microorganisms and substrates. Of the various SOM fractions, the POC fraction is suspected to play the biggest role in the provision of metabolic energy to decomposer populations.

11.4.1.2 Source of Macronutrients

A result of SOM decomposition is the conversion of macronutrients (N, P, and S) locked within organic chemical structures into inorganic forms, which are either immobilized and used in the synthesis of new tissues within soil organisms or mineralized and released into the soil mineral nutrient pool. With the exception of intensively managed soil receiving significant fertilizer inputs, organic matter provides the largest pool of macronutrients in the soil, with HUM being clearly the dominant fraction holding most macronutrients. McGill and Cole (1981) proposed that the mineralization of C, N, P, and S followed a dichotomous system involving both biological and biochemical mineralization. Biological mineralization is driven by the need of decomposer organisms for C as an energy source and accounts for the mineralization of N- and C-bonded S. Biochemical mineralization refers to the release of phosphate and sulfate from the P and S ester pool via enzymatic hydrolysis outside of the cell membrane. As a result and in contrast to organic N, organic P and S accumulation and mineralization in soils can occur independently of C and N dynamics.

11.4.1.2.1 Nitrogen

The soil N pool is dominated by N found in organic structures. In soils with significant contents of K$^+$-containing clay minerals (e.g., illite) capable of fixing NH$_4^+$, approximately 90% of the soil N is contained in organic structures, 8% exists as fixed NH$_4^+$, and 1%–3% can be found in the inorganic plant-available pool (NO$_3^-$ and NH$_4^+$). In soils with little capacity to fix NH$_4^+$ in clay minerals, the proportion of organic N is >97% and the inorganic fraction is 1%–3%. On a global scale, Söderlund and Svensson (1976) estimated that the organic N fraction of soils accounted for 95% of the total soil N pool, which is equivalent to the average value presented by Bremner (1968).

SON has been traditionally divided into the following five fractions based on a variety of acid hydrolysis procedures: (1) acid-insoluble N, (2) ammonia N recovered after hydrolysis, (3) amino acid N, (4) amino sugar N, and (5) hydrolyzable unidentified N. Data summarized by Stevenson (1994) for 11 studies where acid hydrolysis procedures were applied to different soil types showed that there was as much variation in the contents

of each form of N within similar soils as between different soil types. The proportions of each form of organic N were 7%–44% acid-insoluble N, 9%–37% ammonia N, 13%–50% amino acid N, 1%–14% amino sugar N, and 4%–40% hydrolyzable unidentified N. Although methodological differences may account for a portion of the large variations noted in the composition of SON, it is evident that approximately 50% of the total soil N cannot be identified by acid hydrolysis procedures (acid-insoluble N + hydrolyzable unidentified N).

Initial attempts to identify the chemical composition of unidentifiable organic N utilized gel filtration followed by acetylation and GC/MS (Schnitzer, 1985; Schnitzer and Spiteller, 1986). Schulten et al. (1995, 1997) used Curie-point—Py-GC/MS with N-selective detection of the pyrolysis products. These studies suggested that heterocyclic N compounds represented an important component of unidentified SON (see Schulten et al., 1997, for examples of the chemical structure of the heterocyclic N compounds). The formation of heterocyclic N compounds via non-biological fixation of $^{15}NH_3$ by humic substances (IHSS Suwannee River fulvic acid and peat and leonardite humic acids) and by reacting ^{15}N-labeled aniline with humic materials was noted by Thorn and Mikita (1992) and by Thorn et al. (1996), respectively. In contrast to these results, studies utilizing solid-state ^{15}N NMR spectroscopy have failed to observe substantial contributions from heterocyclic N, and spectra tend to be dominated by signals arising from amides and terminal amino groups (Clinton et al., 1995; Knicker and Lüdemann, 1995; Knicker et al., 1995). Further effort is required to address these inconsistencies and to quantitatively characterize the composition of the fraction of N, which cannot be identified by conventional acid hydrolysis procedures.

11.4.1.2.2 Phosphorus

The composition and cycling of soil organic P have been reviewed by Stevenson (1986, 1994) and Sanyal and Datta (1991). As a result of potential adsorption and inorganic precipitation reactions capable of reducing the availability of P in soils, mineralization of organic P is important to soil fertility (Tiessen et al., 1984; Beck and Sanchez, 1994). The relative importance of organic P as a nutrient source tends to be greater on highly weathered soils (Duxbury et al., 1989). The principal organic P-containing compounds in soils and their approximate proportions include inositol phosphates (2%–50%), phospholipids (1%–5%), nucleic acids (0.2%–2.5%), trace amounts of phosphoproteins, and metabolic phosphates (Stevenson, 1994). Soil organic P accounts for a variable proportion of the total soil P. Halstead and McKercher (1975) and Uriyo and Kesseba (1975) presented soil organic P values ranging from 4 to 1400 $\mu g\ g^{-1}$ soil, which accounted for 3%–90% of the total soil P. Uriyo and Kesseba (1975) derived the relationship between organic P and organic C given in Equation 11.8, which produces an organic C:P ratio of 115 and is consistent with the average value of 117 proposed by Stevenson (1994):

$$\text{Organic C } (mg\ g^{-1}\ \text{soil})$$
$$= 4.9 + 0.059 \text{ organic P } (mg\ g^{-1}\ \text{soil}) (R^2 = 0.49). \quad (11.8)$$

11.4.1.2.3 Sulfur

Reviews of the cycling and chemical composition of soil organic S include Stevenson (1986, 1994) and Nguyen and Goh (1994). Sulfur-containing organic compounds found in soils are generally grouped into two pools: compounds in which the S can be reduced to H_2S by hydroiodic acid (HI) and compounds in which the S is directly bound to C. The HI-reducible fraction consists mainly of ester sulfates (C–O–S bonds) and some ester sulfamates (C–N–S bonds). The C-bonded S fraction contains amino acid S (C–S bonds) or sulfonates (C–SO_3 bonds). The ester sulfates and sulfamates are typically associated with aliphatic side chains of soil organic compounds (Bettany et al., 1979), while the C-bonded S is incorporated along with C and N into the core of soil organic compounds and is generally less biologically accessible (McGill and Cole, 1981; Stewart and Cole, 1989). Organic S typically accounts for >90% of the total S found in nonsaline and nontidal soils (Nguyen and Goh, 1994; Stevenson, 1994).

11.4.1.3 Ecosystem Resilience

The resilience of an ecosystem can be defined as its capacity to return to its initial state after being subjected to some form of disturbance or stress (e.g., Webster et al., 1975; DeAngelis, 1980). The important role played by SOM in determining the resilience of an ecosystem can be exemplified by a comparison of the contents of chemical energy and nutrients stored within the soil organic fractions in several ecosystems. In temperate grasslands, high SOM contents result from large belowground additions of photosynthate, limited leaching, and slow decomposition rates. Storage of C in such ecosystems is greater in the soil than in vegetation (Szabolcs, 1994). The large store of chemical energy and nutrients contained in SOM offers resistance to the loss of soil fertility induced by natural or agricultural disturbance. Temperate grassland soils (e.g., mollisols) will remain agriculturally productive with limited inputs for many years, despite the mining of energy and nutrient reserves contained within SOM (Janzen, 1987; Tiessen et al., 1994). Such systems can be considered resilient, at least initially, but one must question how long such systems can be sustained. Tiessen et al. (1983) showed that rates of organic P mineralization in a grassland soil were in excess of crop requirements over the first 60 years of agricultural production. Subsequent to the first 60 years, only the less labile, low-energy-providing forms of organic matter remained, and organic P mineralization rates decreased below crop demand.

In temperate forests, SOM contents are less than that of temperate grasslands and more C and nutrients are stored in aboveground vegetation than in the readily available soil organic materials (Szabolcs, 1994). As a result, the impact of a natural disturbance such as fire can significantly deplete ecosystem stores of energy and nutrients, and ecosystem recoveries (resilience) are slow due to low residual contents of SOC and associated nutrients. Where temperate forests are cleared and agricultural production is initiated, SOM losses must be minimized; however, production systems that increase SOM and nutrient reserves (e.g., crop rotations including legume pastures) can lead to highly productive and sustainable agriculture.

In tropical forest ecosystems, the storage of energy and nutrients in vegetation dominates, and the rapid utilization of plant residues by decomposer organisms and cycling of nutrients maintain ecosystem stability. This, when coupled with the low stores of energy and nutrients in organic matter of tropical soils, indicates a reduced importance of SOM in ecosystem resilience (Anderson, 1995). A comparison of a temperate grassland mollisol with a tropical oxisol (Tiessen et al., 1994) demonstrated the important contribution of SOM to the resilience of the grassland soil and its reduced significance in the tropical soil.

11.4.1.4 Stimulation and Inhibition of Enzyme Activities and Plant and Microbial Growth

Research pertaining to the impacts of SOM on plants, microorganisms, and enzyme activities has typically used humic substances (e.g., humic and fulvic acids) as surrogates for SOM. The influence of humic and fulvic acids, tannins, and melanins on the activity of various enzymes was summarized by Ladd and Butler (1975), Müller-Wegener (1988), and Gianfreda and Bollag (1996). Based on earlier studies, Ladd and Butler (1975) concluded that the effect of humic acids on the activity of proteolytic enzymes varied and that the mechanism of humic acid–enzyme interaction involved primarily the carboxyl groups of humic acids. Inhibition of nonproteolytic enzyme activities by humic acids has also been demonstrated (Sarkar and Bollag, 1987). Müller-Wegener (1988) indicated that possible humic acid–enzyme interactions, which could impact on enzyme activity, included the following: (1) a direct interaction of the humic acid with the enzyme resulting in a modification of enzyme structure or changes in the functioning of active sites, (2) interference in the equilibrium of the enzyme reaction via the humic substances acting as analogue substrates, and/or (3) a reduction in the availability of cations, which often act as cofactors required for enzyme catalysis or structural stabilization of the protein molecule, by fixation on the humic acid molecule.

The influence of soil humic substances on plant growth and cellular activity have been reviewed (Chen and Aviad, 1990; Clapp et al., 2001; Varanini and Pinton, 2001; Nardi et al., 2002) and are generally attributed to direct (e.g., enhanced biochemical activity of plants) and indirect (e.g., increased efficiency of nutrient uptake) effects typically involve the absorption or adsorption of humic substances and the impact of these processes on biochemical properties at cell walls, cell membranes, and/or in the cytoplasm. Information on the impacts of humic materials in field studies is scarce and often confounded with other impacts of humic materials on soil properties (e.g., cation exchange capacity [CEC], nutrient status). Favorable effects on plants grown in defined media have included the following: (1) increased uptake of water and germination rate of seeds, (2) enhanced growth of shoots and roots as assessed by measurements of length and fresh and/or dry mass, and (3) increased root elongation, number of lateral roots, and root initiation (Nardi et al., 1996, 2002). Canellas et al. (2009) found that exposure of maize seedlings to solutions with 20 mg of humic acid carbon per liter for 7 days significantly altered metabolic activity and

enhanced root growth (length or density). Da Rosa et al. (2009) observed enhanced growth and K uptake by common beans when exposed to humic substances extracted from charcoal. Vaccaro et al. (2009) noted positive effects on growth and enzymatic activity of maize seedlings when exposed to the hydrophilic component or the entire soluble component extracted from a compost. Verlinden et al. (2009) noted an overall positive effect on dry matter yield and N uptake for a variety of plants (permanent grassland, maize, potato, and spinach) due to the application of humic substances originating from leonardite formations. Such positive effects on plant growth have been postulated to result from increased permeability of cell membranes, increased chlorophyll content, increased rates of photosynthesis and respiration, enhanced protein synthesis resulting from a stimulation of ribonucleic acid synthesis, and enhanced enzyme activity (Vaughan and Malcolm, 1985).

Addition of humic substances to soil can also influence the activity of soil microorganisms through the provision of a metabolizable source of carbon, increased nutrient supply, and enhanced permeability of cell membranes toward required solutes (Valdrighi et al., 1996). Addition of humic substances at concentrations ≤ 30 mg L^{-1} to a nutrient solution increased growth rates in microbial cultures (Visser, 1985). Humic acid addition was also found to stimulate in vitro growth and activity of aerobic nitrifying bacteria, but not actinomycetes or filamentous fungi (Vallini et al., 1993, 1997; Valdrighi et al., 1995, 1996). A greater promotion of microbial growth has been noted as the molecular weight of the added humic substances decreased (Garcia et al., 1991; Valdrighi et al., 1995). It has been suggested that interactions between added humic materials and microbial cell surfaces (Stehlickova et al., 2009) or humic materials and hydrophobic pollutants (Vacca et al., 2005) may be responsible for enhanced rates of decomposition and mineralization. Enhanced microbial activity due to the addition of humic substances is not always noted, particularly where the added humic substances form the sole source of available carbon (Filip and Tesarova, 2004). Whiteley and Pettit (1994) noted a decreased ability to decompose wheat straw in the presence of humic acid derived from lignite, and Yasmeen et al. (2009) observed that humic acids isolated from oil palm compost inhibited the mycelial growth indicating the presence of a fungicidal activity.

11.4.2 Physical Functions

11.4.2.1 Stabilization of Soil Aggregates

Organic matter is considered important to the maintenance of the structural stability of a wide range of soil types including mollisols, alfisols, ultisols, and inceptisols. Its importance tends to be less in oxisols and andisols, where hydrous oxides play an important stabilizing role, and in self-mulching soils (e.g., some vertisols), which contain clays with a high shrink/swell potential. In soils where organic matter is an important agent binding mineral particles together, a hierarchical arrangement of soil aggregates exists in which aggregates break down in a stepwise

manner as the magnitude of an applied disruptive force increases (Tisdall and Oades, 1982; Oades and Waters, 1991; Oades, 1993). Golchin et al. (1997a) and others have proposed the existence of three levels of aggregation: (1) the binding together of clay plates into packets <20 μm, (2) the binding of clay packets into stable microaggregates (20–250 μm), and (3) the binding of stable microaggregates into macroaggregates (>250 μm).

The importance and nature of the organic materials associated with each level of aggregation varies. At the scale of packets of clays, aggregation is primarily dictated by soil mineralogical and chemical properties important in controlling the extent of dispersion and is often a function of pedological processes. The binding together of clay packets to form microaggregates occurs via a range of mechanisms. The dominant mechanism is proposed to involve polysaccharide-based glues (mucilages or mucigels) produced by plant roots and soil microorganisms (Ladd et al., 1996). Emerson et al. (1986) presented transmission electron micrographs showing mucilage located between packets of clay plates. Small microaggregates (<53 μm) held together by humified organic matter and biologically processed materials are bound together around a particulate organic core (Oades, 1984; Elliott, 1986; Beare et al., 1994a; Golchin et al., 1994b) to produce larger microaggregates and small macroaggregates <2000 μm. Macroaggregates >2000 μm are stabilized by the presence of roots, fungal hyphae, and larger fragments of plant residues, which interconnect soil aggregates via bonding to aggregate surfaces, penetration into or through aggregates, and/or physical enmeshment (Tisdall and Oades, 1982; Churchman and Foster, 1994; Foster, 1994).

11.4.2.2 Water Retention

Organic materials can influence soil water retention directly and indirectly. SOM can absorb and hold substantial quantities of water, up to 20 times its mass (Stevenson, 1994). This direct effect, however, depends on the morphological structure of the organic materials and will not impart any beneficial effect to the soil unless it serves to enhance the ability of soil to hold water at potentials within the plant-available range. Organic matter in the form of surface residues can also influence water retention directly by reducing evaporation and increasing the infiltration of water.

The indirect effect of SOM on water retention arises from its impact on soil aggregation and pore-size distribution, and thus on plant-available water-holding capacity, AWHC, of the soil (the difference between volumetric water content at field capacity and permanent wilting point). This effect is best exemplified by the inclusion of SOC content as a significant parameter in pedotransfer functions, which predict pore-size distribution (e.g., Vereecken et al., 1989; da Silva and Kay, 1997; Kay et al., 1997). Equation 11.9 presents the pedotransfer function derived by da Silva and Kay (1997) to describe the relationship between volumetric water content, θ_v (m³ m⁻³), and matric potential, ψ (MPa), clay content, CL (%), organic C content, OC (%), and bulk density, BD (Mg m⁻³). Using this equation, Kay et al. (1997) calculated predicted changes in AWHC for soils ranging in clay content from 7% to 35% when organic C content was increased

by 0.01 kg kg⁻¹. Increases in AWHC of 0.039 and 0.020 (m³ m⁻³) were obtained for the soils with 7% and 35% clay, respectively, at a relative bulk density of 0.75. Application of the same equations to a data set acquired by Wegner et al. (1989) for 80 South Australian red brown earths (alfisols) showed that the increase in AWHC induced by increasing organic C content by 0.01 kg kg⁻¹ soil could be expressed by Equation 11.10. These results indicate that the presence of additional organic matter enhances AWHC of soils. Although the magnitude of the increase decreases with increasing clay content, building SOC content would be expected to be more difficult on a sand than on a clay soil:

$$\theta_v = a\psi_m^b, \qquad (11.9)$$

where

$$a = \exp(-4.15 + 0.68\ln CL + 0.42\ln OC + 0.27\ln BD),$$

$$b = -0.54 + 0.11\ln CL + 0.02\ln OC + 0.10\ln BD,$$

$$\text{Change in AWHC} = -0.0012(\%\text{clay}) + 0.055(R^2 = 0.82). \quad (11.10)$$

11.4.2.3 Soil Thermal Properties

The typical dark color of SOM contributes to the dark color of surface-mineral soils and can enhance soil warming and promote biological processes related to temperature in cooler climates (e.g., plant growth and mineralization of C and nutrients contained in SOM). However, the presence of litter layers or organic horizons can insulate a soil against fluctuations in air temperature and solar heating. On several Canadian forest soils subject to cold winters and cool springs, average soil temperatures and the growth of fertilized seedlings were greater where the litter layers were removed compared to where they were left intact (Burgess et al., 1995). Similar effects have been observed in a comparison of cropping systems, which leave different amounts of crop residue on the surface of the mineral soil (Fortin, 1993).

11.4.3 Chemical Functions

11.4.3.1 Cation Exchange Capacity

Organic matter contributes 25%–90% of CEC of the surface layers of mineral soils and practically all of the CEC of peats and forest litter and humus layers (Stevenson, 1994). The percent contribution is greatest for soils with low clay content or where the clay fraction is dominated by minerals with a low-charge density, such as kaolinite, and is lowest for soils with high contents of highly charged minerals, such as vermiculite or smectite. Organic matter will contribute most significantly to soil CEC in sandy soils.

The contribution of organic matter to soil CEC is pH dependent. At typical soil pH values (>5), the CEC of organic matter is derived principally from carboxyl functional groups, but phenol, enol, and imide groups may also contribute at higher pH values. Given that an increase in degree of oxidation is typically associated with decomposition of organic materials in soil, more

highly degraded organic materials would be expected to have a higher CEC than their less decomposed analogues. An increase in CEC was noted by Roig et al. (1988) during the degradation of manure over time. Beldin et al. (2007) measured the CEC of light and heavy soil fractions and found a positive correlation between CEC and %C for the light fractions but not for the heavy fractions, presumably due to the high content of mineral soil constituents that contribute to CEC in the heavy fractions (e.g., clay). The result from Beldin et al. (2007) and results obtained for black carbon by Liang et al. (2006) indicate that the different forms of SOM found in soil will contribute differently to CEC and that significant contributions can be made by the nonhumified components of SOM.

Beldin et al. (2007) also noted that predicted whole soil CEC values obtained by mathematically combining the CEC values measured for the light and heavy fractions on a mass basis were much greater than those measured for the whole unfractionated soils. The potential for carboxylic acid groups to be involved in organomineral interactions and the complexation of cations may reduce their ability to contribute to soil CEC. CEC measurements made on organic matter fractions isolated from soils must therefore be treated with caution, particularly where the potential exists to break organomineral associations and displace complexed cations during the fractionation process.

One approach used to assess the impact of SOM on soil CEC has involved performing CEC measurements before and after organic matter removal (Tan and Dowling, 1984; Thompson et al., 1989; Turnpault et al., 1996). However, this approach may result in an underestimation of the CEC of SOM since organic matter removal may expose inorganic CEC sites that were previously involved in organomineral interactions and not capable of contributing to whole soil CEC. Derivation of regression relationships between CEC and SOM/carbon has also been used to define the contribution of organic materials to soil CEC (Asadu et al., 1997; Oorts et al., 2003; Liang et al., 2006; Rashidi and Seilsepour, 2008; Seilsepour and Rashidi, 2008; Yimer et al., 2008). CEC values generated for SOM have ranged from 15 to >600 $cmol_c$ kg^{-1} C. As a general rule, each weight percentage of SOC contributes approximately 3 $cmol_c$ kg^{-1} soil (300 $cmol_c$ kg^{-1} SOC) to the CEC of neutral permanent charge soils (McBride, 1994) and approximately 1 $cmol_c$ kg^{-1} soil (100 $cmol_c$ kg^{-1} SOC) to the CEC of variable charge soils (Oades, 1989).

11.4.3.2 Buffering Capacity and Soil pH

The presence of weakly acidic chemical functional groups on soil organic molecules that can act as conjugate acid/base pairs makes SOM an effective buffer. The diversity in chemical composition of the functional groups (e.g., carboxylic, phenolic, acidic alcoholic, amine, amide, and others) provides organic matter with the ability to act as a buffer over a wide range of soil pH. James and Riha (1986) reported buffer capacities of 18–36 and 1.5–3.5 $cmol_c$ kg^{-1} (pH unit)$^{-1}$ for the organic and mineral horizons, respectively, of forest soils. Starr et al. (1996) obtained a good correlation between acid buffer capacity and organic matter content for 29 organic and 87 mineral soil horizons (E, B, and C horizons)

exhibiting buffering capacities of 9.8–40.8 and 0.1–5.2 $cmol_c$ kg^{-1} (pH unit)$^{-1}$, respectively. For 59 agricultural soil samples taken from the 0–15 cm layer of cultivated fields, Curtin et al. (1996) noted that titratable acidity could be described by Equation 11.11 in which the terms OC and clay represent the soil organic C and clay contents expressed in units of kg kg^{-1} soil and ΔpH is the reference pH (e.g., 8) minus the initial pH. Assuming the organic C content of SOM is 58%, Equation 11.11 indicates that the buffering capacity offered by organic matter was approximately 34 $cmol_c$ kg^{-1} (pH unit)$^{-1}$ and was an order of magnitude greater than that offered by clay (34 versus 3 $cmol_c$ kg^{-1} [pH unit]$^{-1}$). The average clay/organic C ratio for the soils studied by Curtin et al. (1996) was 7.9/1, indicating that even though most soils contained much more clay than organic C, organic C accounted for about two-thirds of the soil buffering capacity.

> Titratable acidity to pH 8
>
> $$= 0.02 + 59 OC \Delta pH + 3.0 \, clay \, \Delta pH \; (R^2 = 0.95). \quad (11.11)$$

Addition of organic matter to soil may result in increases or decreases in soil pH, depending on the influence that the addition has on the balance of the various processes that consume and release protons. A detailed presentation of these soil processes and their ability to release or consume protons is given by van Breemen et al. (1983). Factors that need to be considered include the chemical nature of the soil and that of the organic materials added as well as environmental properties including water content and extent of leaching. The net effect of adding organic matter to acidic soils is generally an increase in pH values (e.g., Yan et al., 1996; Pocknee and Sumner, 1997) with the main processes leading to the increase being (1) a decomplexation of metal cations, (2) mineralization of organic N, and (3) denitrification. Pocknee and Sumner (1997) found that on the acid Cecil soil, the extent of the increase in pH was controlled by the N content to basic cation content ratio. The decarboxylation of organic acids has also been shown to increase the pH of acid soils (Yan et al., 1996). Under alkaline soil conditions, however, these processes would be ineffective and would contribute to a reduction in soil pH as a result of their influence on soil CO_2 concentrations. The addition of organic matter to alkaline soils tends to acidify them especially under waterlogged and leaching conditions (Nelson and Oades, 1997). The main processes involved in the acidification of alkaline soils on addition of organic materials include (1) mineralization of organic S and P, (2) mineralization followed by nitrification of N, (3) leaching of the mineralized and nitrified organic N, (4) dissociation of organic ligands, and (5) dissociation of CO_2 during decomposition.

11.4.4 Complexation of Inorganic Cations

The presence of various functional groups on SOM provides the capacity for interaction with inorganic cations. Possible interactions can take the form of simple cation exchange reactions, such as that between negatively charged carboxyl groups

and monovalent cations, or more complex interactions where coordinate linkages with organic ligands are formed, such as occurs between amino acids and Cu^{2+} (Harter and Naidu, 1995; Baldock and Skjemstad, 2000). The influence that the complexation of inorganic cations by SOM has on soil properties and processes includes the following:

1. Altered solubility and degradability of associated organic materials (Skyllberg and Magnusson, 1995; Christl and Kretzschmar, 2007; Scheel et al., 2007)

2. Increased availability of insoluble mineral P through complexation of Fe^{3+} and Al^{3+} in acid soil and Ca^{2+} in calcareous soil, competition for P adsorption sites, and displacement of adsorbed P (Stevenson, 1994; Cajuste et al., 1996)

3. The release of plant nutrients through the weathering of rocks and soil parent materials by the removal of structural cations from silicate minerals (Robert and Berthelin, 1986; Tan, 1986)

4. Enhanced availability of trace elements in the upper portion of the soil profile as a result of upward translocation by plant roots and subsequent deposition on the soil surface and complexation during residue decomposition (Stevenson, 1994)

5. Facilitated adsorption of organic materials to soil minerals, which aids in the generation and/or stabilization of soil structure (Oades, 1984; Emerson et al., 1986)

6. Buffering of excessive concentrations of otherwise toxic levels of metal cations (e.g., Al^{3+}, Cd^{2+}, and Pb^{2+}; Anderson, 1995)

7. Pedogenic translocation of metal cations to deeper soil horizons (McKeague et al., 1986) and the formation of minerals (Huang and Violante, 1986)

11.5 Factors Determining the Content of Organic Matter in Soil

The amount of organic matter present in a soil is defined by the balance between the competing rates of input and loss of organic matter. Rates of input are typically defined by the amount of plant residue added to the soil. Any practice that enhances the amount of carbon captured by plants and the return of organic residues to the soil (above- and belowground) will increase inputs. For example, appropriate use of fertilizers to maximize productivity will also maximize returns of organic residues to the soil under any given management regime. Other factors such as the availability of water may place an upper limit on input rates by constraining potential plant productivity and thereby placing an upper limit on input rates. Where organic wastes are available (e.g., municipal green waste, residual materials derived from animal production, and biosolids), their application to soil can increase rates of input beyond that defined by environmental, soil, and management factors that normally limit plant production.

Losses of organic matter from soil result from decomposition and subsequent mineralization of organic carbon and its associated elements. Processes that accelerate decomposition increase the rate of loss. Additionally, losses of organic matter can occur through erosion or leaching. Erosion losses are rapid and event driven and may be catastrophic in localized areas. Losses by leaching are typically small compared to mineralization losses, but over time can lead to a significant removal or redistribution of organic matter within the soil profile.

During pedogenesis, organic matter accumulation in soil goes through a series of development phases. In the initial phase, a slow colonization by photosynthetic organisms occurs. A lack of available nutrients places a ceiling on the amount of CO_2–C that can be fixed. Low rates of carbon capture and low nutrient status limit both production and decomposition of SOM. Accumulation of SOM in this phase, expressed in units of g m^{-2}, proceeds slowly and can be aided by interactions with soil mineral components that are capable of biologically protecting SOM against decomposition. With continued soil development, SOM content and the activity of decomposer organisms increase to a point where a continued supply of nutrients in a plant-available form is reached. At this point, the rate of CO_2–C capture and organic matter deposition is greater than mineralization, and SOM accumulates at an exponential rate. With increasing SOM content, the ability of the soil to protect additional organic matter declines and an increasing proportion of added organic matter remains accessible to decomposition. As a result, the increase in organic matter content through time proceeds through an inflection point and then begins to decrease. Once the capacity for biological protection offered by soil mineral components is approached, the rate of mineralization of SOM tends toward the rate of deposition of fresh organic residues and SOM levels approach an equilibrium value. It is important to note that this biological protection rarely equates to a permanent and complete removal of organic C from the decomposing pool, but rather to a reduction in its rate of decomposition, when compared to similar materials existing in an unprotected state (Baldock and Skjemstad, 2000). As the older protected C is slowly mineralized, its position in the biologically protected pool is replaced with younger modern organic C.

The progression of SOM content with soil development and the magnitude of the equilibrium level of SOM will also depend on interactions, which occur between the factors of soil formation. Where cold and water-saturated soil conditions persist, decomposition is confined to slow anaerobic processes and organic matter contents expressed in units of gram per square meter may continue to increase leading to the formation of organic soils. In sandy soils, the extent of biological protection offered by the soil mineral component will be lower than that offered by clay-rich soils and large differences in SOM content can develop. Therefore, with the exceptions of peatland and wetland soils, which have been estimated to accumulate 0.1–0.3 Pg C $year^{-1}$ globally (Post et al., 1990), organic matter levels in soil do not increase indefinitely but rather tend to equilibrium values dictated by the soil-forming factors of climate, biota (vegetation and soil organisms), parent material, and topography (Baldock and Skjemstad, 1999; Six et al., 2002).

An additional factor that must be considered in an examination of factors influencing organic matter contents in soils used for agriculture and forestry is land management practice. Land management can induce rapid and drastic changes to equilibrium contents of SOM attained under natural undisturbed conditions and completely override the influence of soil-forming factors. For example, the conversion of native ecosystems to agriculture often, but not always (Skjemstad and Spouncer, 2003), results in a net loss of SOM (Mann, 1986; Davidson and Ackerman, 1993; Paustian et al., 1997b).

When measurements of SOM are conducted on perturbed systems, it must be acknowledged that they may still be in the process of attaining a new equilibrium content. When combined with the potential impacts that variations in climatic conditions can have, it can be difficult to detect the true direction of SOM change induced by alterations to land use at timescales <10 years. In fact, more than 50 years may be required to reestablish equilibrium conditions representative of a new land use (Baldock and Skjemstad, 1999). Combining this observation with early results presented by Jenny (1930) suggests that the relative importance of the soil-forming factors on SOM content can be viewed as management > climate > biota (vegetation and soil organisms) > topography = parent material (Baldock and Skjemstad, 1999).

An independent evaluation of the influence of any single factor on SOM contents is difficult because of the requirement that all other factors remain constant. Variations in the soil-forming factors experienced on a landscape scale and the interdependence of these factors contribute to the large variability noted for SOM contents, even within localized areas. When trying to assess changes of SOM content in a soil, it must be noted that rates of change in SOM (typically less than 0.5 Mg C ha^{-1} year^{-1}) are quite small compared to the large amounts of SOM often present (as high as 100 Mg C ha^{-1}, or more, in the top 60 cm soil layer; Ellert et al., 2008). Thus, changes in SOM can only be reliably measured over a period of years or even decades (Post et al., 2001). Since the distribution of SOM in space is inherently variable, temporal changes (e.g., attributable to management practices, environmental shifts, successional change) must be distinguishable from spatial ones (e.g., attributable to landform, long-term geomorphic processes, nonuniform management) to assess whether SOM is either increasing or decreasing (Ellert et al., 2008).

Computer simulation models of SOM dynamics, as defined through changes in SOC (Jenkinson et al., 1987; Parton et al., 1987; McCown et al., 1996), can be used to provide valuable information pertaining to the interaction of soil-forming factors on SOC levels and thus, ecosystem functioning, provided the models are structured appropriately. However, it is essential that field data are available to validate predictions (Burke et al., 1989).

11.5.1 Climate

Climate impacts SOM content primarily through the effects of temperature, moisture, and solar radiation on the array and growth rate of plant species and the rate of SOC mineralization. Post et al. (1982) found that the amount of SOC was positively correlated with precipitation and, at a given level of precipitation, negatively correlated with temperature. Similar results were observed by Guo et al. (2006) in an analysis of factors controlling SOC in the United States using maps generated from the State Soil Geographical database (STATSGO). Guo et al. (2006) found that SOC content increased as the mean annual precipitation (MAP) increased up to 700–850 m and then fluctuated as the MAP increased further. When other variables were highly restricted, there was a clear decline in SOC with increasing temperature. In the Great Plains of North America, precipitation controls NPP and temperature is considered to exert its strongest control over rates of SOC mineralization (Parton et al., 1987; Sala et al., 1988; Burke et al., 1989). Ladd et al. (1985) compared the mean loss of ^{14}C-labeled plant residues from four soils in South Australia with that obtained by Jenkinson and Ayanaba (1977) for soils in England and Nigeria and observed a doubling of the rate of substrate C mineralization for an 8°C –9°C increase in mean annual temperature. An influence of temperature on decomposition can also be inferred from ^{14}C content of SOC, which showed a latitudinal gradient in the mean residence time of SOC (Bird et al., 1996).

The observed trend of decreasing SOC content with increasing temperature implies that the relative temperature sensitivity of decomposition is greater than that of NPP. Because of the strong interactions between temperature, water availability, and substrate quantity, it is difficult to assess the temperature dependence of decomposition without confounding effects. In a compilation of data extracted from controlled incubation studies where water limitations were avoided and a common substrate was used at all temperatures, Kirschbaum (1995) showed that the Q_{10} value (rate of change of a process with a 10°C temperature increase) of C mineralization from soil was greater than that for NPP developed by Lieth (1973), especially at temperatures <15°C. Increases in temperature, particularly when starting from temperatures <15°C, will enhance decomposition more than NPP. The greater sensitivity of carbon mineralization to climatic variation was also observed at a regional (Goulden et al., 1998; Saleska et al., 2003) and global scale (Hicke et al., 2002). Hicke et al. (2002) showed that the variability in NPP was considerably less than the variability in growth rate of atmospheric CO_2 using inverse modeling, suggesting that the cause of the year-to-year variability in carbon fluxes is largely from varying rates of respiration rather than photosynthesis.

Climate has also been shown to affect the chemical structure of SOC. Using Py-GC to characterize the chemical structure of SOC in a climosequence of nine New Zealand soils, Bracewell et al. (1976) observed significant correlations between changes in the intensity of peaks in the chromatograms and MAP and temperature. By including both temperature and precipitation in a regression analysis, the resultant regression line explained 90% of the variation in chromatogram peak intensities. Amelung et al. (1997) used 10 grassland samples originating from different climatic zones of the North American Great Plains to investigate the impacts of mean annual temperature and precipitation on the chemical structure of SOC using a combination of chemical

methods and ^{13}C NMR. MAP was capable of accounting for only 10% of the variation in alkaline CuO oxidizable lignin. Higher precipitation tended to favor an accumulation of polysaccharide carbon; however, at a given MAP, polysaccharide carbon tended to decrease with increasing temperature. Amelung et al. (1997) suggested that the increased content of polysaccharide carbon in more humid conditions may have resulted from (1) a positive feedback mechanism in which increased plant production enhanced microbial activity and soil structural conditions, thereby offering the potential for protecting microbial polysaccharides within aggregates, and/or (2) an enhanced activity of earthworms that elevated polysaccharide content relative to the surrounding soils (Guggenberger et al., 1995) and offered organic carbon some physical protection against mineralization (Lavelle and Martin, 1992). Accompanying the decrease in polysaccharide carbon noted with increasing temperature, Amelung et al. (1997) further noted an increase in aliphatic carbon content. This accumulation of alkyl carbon at high temperature may be explained by (1) enhanced mineralization of carbohydrates and selective preservation of plant or microbially derived alkyl structures by adsorption onto clay particles (Baldock et al., 1989, 1992) and/or (2) higher inputs of plant-derived alkyl carbon in plant residues due to the presence of thicker cuticles on plants growing in warmer climates.

In a study using a soil sequence along an elevational gradient ranging from subtropical to subalpine climate zones in the Etna region (Sicily, southern Italy), Egli et al. (2007) examined changes in SOC and C:N ratios. Egli et al. (2007) showed that the concentration of SOC in the topsoil, the stocks of SOC in soil profile and the nature of the SOC were strongly related to elevation and, thus climate and vegetation. However, the C:N ratio in the topsoil was more defined by the vegetation type. A better protection of SOC at lower altitude was found and suggested to be an effect of the specific climate conditions with more pronounced change in periods of humidity alternating with periods of drought and resultant fire activity. Repeated bushfires had played a significant role in the soil formation as indicated by the presence of aromatic compounds and charcoal (Egli et al., 2007).

11.5.2 Soil Mineral Parent Materials and Products of Pedogenesis

The mineral phase of soils can exert a strong influence on SOM contents as a result of mechanisms capable of protecting organic materials against biological attack (Baldock and Skjemstad, 2000). Each soil has a given capacity to protect SOC dictated by the following:

1. The chemical nature of soil minerals
2. The presence of multivalent cations and their ability to form complexes with organic molecules in soils
3. The adsorptive capacity of soil minerals for organic materials as governed by particle size and surface area
4. Physical protection mechanisms, which restrict access of organic materials to biological attack, that is, the architecture of the soil matrix

The degree and amount of protection offered by each mechanism depends on the chemical and physical properties of the mineral matrix and the morphology and chemical structure of the organic matter. Furthermore, as with other aspects of SOC dynamics, strong interactions can exist between these characteristics (e.g., the presence and type of multivalent cations will undoubtedly be related to the chemical nature of the minerals present). Finally, each mineral matrix will have its unique and finite capacity to protect organic matter. In the extreme case where mineral protection mechanisms are not present, such as in well-aerated peat or forest litter layers, decomposability will be controlled by the recalcitrance offered by the chemical structure of the SOM itself.

11.5.2.1 Chemical Nature of the Soil Mineral Fraction

An analysis of different soil types indicates that soils with high contents of calcium carbonate ($CaCO_3$) and amorphous Al and Fe tend to have higher organic C contents compared to other soil types (Spain et al., 1983; Oades, 1988; Sombroek et al., 1993). In a study of the influence of soil properties on SOC genesis, Duchaufour (1976) suggested that the presence of calcium carbonate in a rendzina could protect both particulate organic and HUM carbon. Thin carbonate coatings visible under magnification and a precipitation of organic molecules induced by Ca^{2+} complexation were implicated in the protection of POM and HUM, respectively, and helped to explain the observed impedance of mineralization. Protection of SOC in high-base-status soils with less reactive or low contents of calcium carbonate results predominantly from the formation of Ca-organic linkages. In such soils, the initial decomposition of plant residues is rapid, but the subsequent utilization of initial decomposition products is slow leading to higher SOC contents, lower C:N ratios, and longer retention times. Soils with high base status typically have higher clay contents, are more fertile, and have greater annual vegetative inputs than similar low-base-status soils. Establishment of causative relationships between base status and SOC contents must therefore be examined carefully because of the potential confounding effects of increased vegetative inputs and protection mechanisms involving clay minerals.

Soils derived from volcanic ash (andisols) are typically characterized by large accumulations of SOC, high C:N ratios, and high allophane contents. The formation of Al-organic complexes is considered to be important to the biological protection of organic C in andisols. Boudot et al. (1986, 1988) obtained a significant correlation between the amount of native C mineralized from 10 French highland soils and the contents of amorphous Al and allophane, without observing significant correlations with clay content, exchangeable Al^{3+}, or crystalline iron oxides. Decreased organic C mineralization rates from ^{14}C-labeled organic substrates in allophanic soils and nonallophanic soils amended with allophane, relative to that noted in unamended nonallophanic soils, also demonstrated a protective effect of allophanic material on SOC (Zunino et al., 1982; Boudot et al., 1988, 1989). Zunino et al. (1982) demonstrated that the influence of allophane on mineralization of C from an organic substrate

varied with the chemical structure of the substrate. The presence of amorphous Fe compounds and Fe^{3+} cations has been shown to have a similar effect to that of allophane and Al^{3+} cations on the mineralization of C from organic materials; however, the magnitude of the protective effect was reduced (Boudot et al., 1989).

11.5.2.2 Impacts of Multivalent Cations

The presence of multivalent cations in soil has important implications on the behavior of clays and organic materials and the biological availability of organic C. When saturated with multivalent cations, clays remain flocculated, which reduces exposure of organic materials adsorbed onto their surfaces (Section 11.5.2.3) and macromolecular organic materials bearing functional groups become more condensed (altering their 3D structure) and thus less susceptible to the enzymatic attack.

The dominant multivalent cations present in soils include Ca^{2+} and Mg^{2+} in neutral and alkaline soils and hydroxypolycations of Fe^{3+} and Al^{3+} in acidic, ferrallitic, and andic soils. A protecting effect of Ca^{2+}, relative to Na^+, on organic C mineralization was effectively demonstrated by Sokoloff (1938), where the extent of mineralization and solubility of organic C in two soils was reduced by the addition of Ca^{2+} salts and enhanced by the addition of Na^+ salts. Other studies have also shown a decreased solubility of SOC in the presence of Ca^{2+} (Muneer and Oades, 1989c) and reduced mineralization of native organic materials and organic substrates on the addition of Ca^{2+} in incubation studies (Linhares, 1977; Muneer and Oades, 1989a, 1989b). In such studies, the question remained as to whether the effect of Ca^{2+} addition on mineralizable C resulted from an indirect effect on colloidal dispersibility or from a direct effect of Ca^{2+} complexation on the biodegradability of the organic molecule.

A direct effect of multivalent cation complexation on biodegradability in soil has been demonstrated by the following results:

1. A reduced oxygen absorption on incubation of humic acids saturated with Ca^{2+}, Al^{3+}, or Fe^{3+} in the same soil, relative to that noted for Na^+-saturated humic acids (Juste and Delas, 1970; Juste et al., 1975)
2. An increased protection of Al^{3+} and Fe^{3+} forms of plant and microbial polysaccharides (Martin et al., 1966, 1972)
3. A threefold increase in the amount of C mineralized from an organic soil after replacing Ca^{2+} cations with K^+ during a 25 week incubation (Gaiffe et al., 1984)

Indirect evidence for the involvement of cations in the accumulation of SOC can also be obtained through a comparison of the organic C contents of a variety of soil types. Using data derived from Spain et al. (1983) for the organic C contents of 29 Australian great soil groups, Oades (1988) showed that, excluding soils subject to waterlogging, there was a positive correlation between SOC contents and either high base status or the presence of substantial contents of Al and Fe oxides. Of interest was the comparison of siliceous and calcareous sands, which have little or no clay, but indicate an increased SOC content in the presence of Ca^{2+}-containing mineral fractions (<0.5% to 1.5% versus 1.5% to >4% organic C, respectively.)

11.5.2.3 Adsorption of Organic Materials onto Mineral Surfaces

Clay particles provide a reactive surface onto which organic materials can be adsorbed, and it is generally accepted that such adsorption reactions provide a mechanism of protecting SOC against microbial attack. The mineralogy, surface charge characteristics, and precipitation of amorphous Fe and Al oxides on clay mineral surfaces give clay minerals the capacity to adsorb organic matter and protect it from biological degradation. Correlations between SOC and clay contents have been observed (Schimel et al., 1985a, 1985b; Spain, 1990; Feller et al., 1991), and the various interactions between soil clays and organic materials have been summarized by Oades (1989). Such interactions are principally defined by the chemical nature of organic materials (functional group content, molecular size, etc.) and the type of clay mineral (kaolinite, illite, smectite, etc.). Numerous studies utilizing isotopically labeled organic substrates have shown a positive relationship between the contents of residual substrate C and soil clay content (Amato and Ladd, 1992). Clay particles have also been observed to encapsulate particles or patches of SOM (Baldock, 2002).

Not only the amount of clay, but also the specific surface area (SSA) of the clay and more generally the soil mineral particles is of importance. For example, Ransom et al. (1998) showed that adding even small amounts of high-SSA (100 m^2 g^{-1}) clay-size (<2 μm) material can have a significant effect on the total SSA of mineral particle mixtures. It must be noted that the relative magnitude of the effect of adding high-SSA clays to nonclay mineral particles decreases as the size of the nonclay mineral particles decreases (Baldock and Skjemstad, 2000).

In a field experiment where ^{14}C-labeled plant residues were added to four cultivated soils varying in clay content (5%–42%) but having similar clay mineralogy, climatic conditions, and no other organic inputs, the amounts of residual ^{14}C and total organic C in the topsoil (0–10 cm) remaining after 8 years of decomposition were nearly proportional to soil clay content (Ladd et al., 1985). Saggar et al. (1996) completed a similar study in which the decomposition of ^{14}C-labeled ryegrass was monitored over 6 years in four soils having variable clay content (16%–60%) and clay mineralogy. The mean residence time of the ^{14}C-labeled ryegrass was not related to clay content but rather to the SSA as measured by adsorption of *p*-nitrophenol. The increase in mean residence time with increasing SSA, suggested that the protective capacity of the soils toward transformed metabolites derived from plant residues was principally controlled by adsorption onto soil surfaces. Since the data presented by Ladd et al. (1985) were derived from soils with a similar clay mineralogy, SSAs would have been well correlated with clay content. The importance of available surface area was also suggested by the results of Sørensen (1972, 1975) where the addition of high-SSA montmorillonite to a soil/sand mixture protected microbial metabolites, but the addition of low-SSA kaolinite had little influence.

Hassink (1997) found a relationship between the silt- and clay-associated C and soil texture, whereas this relationship

did not exist for the C in the sand-sized fraction (i.e., POM C) suggesting that the capacity of soil to preserve C was linked to silt and clay particles. Using data from Hassink (1997), Six et al. (2002) were able to define significantly different relationships for 1:1 clays versus 2:1 clays, demonstrating the effect of clay type in C protection. Clearly, 1:1 and 2:1 clays have significant differences in CEC and SSA that may lead to a different capacity to adsorb organic materials.

SSA and mineralogy have been identified as key components in the preservation of organic matter in marine sediments (Keil et al., 1994; Mayer, 1994a, 1994b). Based on data generated in these studies, Ransom et al. (1998) showed that total organic carbon in the marine sediments was linearly related to SSA as well as to the content of high surface area minerals present. All these results suggest that the potential protective capacity of soil mineral particles, and more specifically that of the clay minerals, is more a function of the SSA available for adsorption of SOC than the absolute amount of clay.

The protective effect of mineral surfaces can also be shown experimentally by removing the minerals associated with the various fractions as defined by Skjemstad et al. (1999) with HF, since HF is known to dissolve soil minerals with little alteration to the composition of organic matter. In a recent experiment conducted by the authors, incubation studies were used to compare the mineralizability of HF- and non-HF-treated fractions of SOM (unpublished data, Figure 11.10). Mineralizability was defined as the cumulative amount of CO_2–C emitted per gram of organic C present in the sample being incubated. CO_2–C emissions were quantified by repeatedly sampling and then refreshing the headspace of an incubator system for up to 70 days using infrared detection of CO_2–C. Comparison of the mineralizability of SOC with and without HF treatment revealed an increased mineralizability with destruction of soil minerals consistent with the ability of minerals to stabilize a portion of the SOC against biological attack. The magnitude of the difference between the

non-HF- and HF-treated soils provides an assessment of the relative importance of mechanisms of mineral protection in defining the biological stability of the SOC present.

11.5.2.4 Physical Protection within Soil Matrix Offered by Soil Architecture

The architecture or structural condition of a soil can exert significant control over processes of biological decomposition through its effects on water and oxygen availability, dynamics of soil aggregation, and by limiting the accessibility of SOC to decomposer microorganisms and of microorganisms to their faunal predators. This limitation results from the ability of clays to encapsulate organic materials (Tisdall and Oades, 1982), the burial of SOC within aggregates (Golchin et al., 1997a, 1997b), and the entrapment of SOC within small pores (Elliott and Coleman, 1988; Strong et al., 2004). As outlined by van Veen and Kuikman (1990) and Hassink (1992), evidence of the importance of these processes in the protection of SOC in soils can be inferred from the following observations:

1. A faster turnover rate of organic substrates in liquid microbial cultures relative to that of similar substrates in mineral soils
2. An enhanced mineralization of C and N when soils are disrupted prior to incubation
3. A more rapid mineralization of organic C and plant residues in sandy soils than clay soils

A continuum of pore sizes exists in soils, starting with large macropores (>20 μm) and decreasing to micropores (<0.1 μm). Kilbertus (1980) suggested that bacteria can only enter pores >3 μm, which suggests that a significant proportion of the soil pore space may not be accessible to microbial decomposers. Organic materials adsorbed onto clay particles contained in pores <3 μm would only be decomposed as a result of diffusion of extracellular enzymes released by microorganisms followed
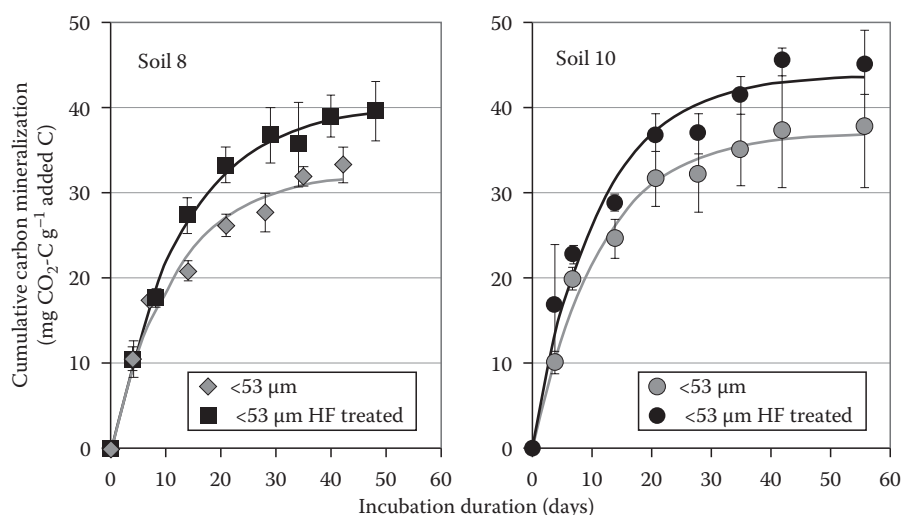
FIGURE 11.10 Change in mineralizability of <53 μm SOC fractionated from two different Australian soils induced by pretreatment with 2% HF according to Skjemstad (1994).

by a diffusion of the products of enzyme reactions back to the microorganisms (see Section 11.1). With increasing soil clay content, the proportion of the total soil pore space contained in micropores increases, and the potential for protection due to the exclusion of soil microorganisms increases. This concept of exclusion can be extended to the predation of microorganisms by soil fauna. van der Linden et al. (1989) suggested that protozoa and nematodes are excluded from pores <5 and <30 μm, respectively. Killham et al. (1993) showed that although placing glucose into pores <6 or <30 μm did not impact the rate of glucose decomposition, the turnover of glucose C incorporated into the microbial biomass was slower where glucose was only added to pores <6 μm. Strong et al. (2004) showed that decomposition occurs faster in soils with a large volume of pores with neck diameters of 15–60 μm. Their observations pointed to particularly rapid rates of decomposition near the air–water interface, most likely because of the ideal conditions for the organisms' mobility, nutrient or toxin diffusion, and oxygen supply. Furthermore, they suggested that on the one hand, SOC in large air-filled pores decomposes more slowly than in intermediate-sized pores (most likely due to decreased organism mobility, diffusion of solutes, and intimacy of contact between SOC and soil minerals) whereas on the other hand, the carbon trapped in the smallest pores was physically protected against decomposition.

The ability of clay particles to adsorb organic materials can also contribute to a biological protection of SOC through encapsulation and the formation of stable aggregates. Encapsulation of particulate organic residues in soils not only places a physical barrier between decomposer organisms indigenous to soils and potential substrates, but can also limit the movement of water and oxygen to sites of potentially active decomposition. A similar situation develops within soil aggregates. Relative to the larger pores between aggregates, the smaller pores within aggregates are more likely to remain filled with water during drying events, and therefore restrict oxygen movement into the aggregate. The presence of organic cores in aggregates (Beare et al., 1994a, 1994b; Golchin et al., 1994a, 1997a) will serve to increase this effect by enhancing oxygen consumption within the aggregate. It has been found that anaerobic conditions can exist in the core of moist aggregates even under well-aerated conditions (Sexstone et al., 1985). Smaller decomposition rates of SOC enclosed within soil aggregates compared to SOC located outside of soil aggregates have also been shown (Sollins et al., 1996; Angers et al., 1997). In native grasslands, Amelung and Zech (1996) demonstrated that the exterior 0.5 mm of >2 mm diameter peds contained less SOC and had a higher C:N ratio, less lignin, and more microbial-derived saccharides than ped interiors. The SOC associated with ped surfaces, therefore, appeared to turn over more rapidly and exhibited a greater degree of decomposition than that contained within peds.

Soil aggregation is a transient property and aggregates are constantly being formed and destroyed. Quantitative data on soil structural dynamics are, however, lacking. Recently, De Gryze et al. (2005) were able to define a macroaggregate turnover time of 40–60 days using data from a 3 week incubation experiment

and a model assuming an aggregate formation rate proportional to the respiration rate. There was no evidence that aggregate formation differed amongst the three soils examined, which all had a different structure. A subsequent study (De Gryze et al., 2006) confirmed this finding and showed that soil texture affected aggregate stabilization rather than aggregate formation.

11.5.3 Biota: Vegetation and Soil Organisms

11.5.3.1 Vegetative Inputs: Variations across and within Ecosystems

Vegetation influences SOC content as a result of the amount, placement, and biodegradability (chemical recalcitrance) of plant residues returned to the soil. Plant residues can be considered the dominant input and thus the primary source of organic carbon into or onto soils (Kögel-Knabner, 2002). Organic molecules created by the soil fauna and microorganisms are the secondary source of decomposable organic carbon in a soil. The greatest effects of vegetation on SOC contents are confined to the A horizon. Concentrations of organic C detected below the A horizon result from a combination of plant input through roots as well as pedogenic processes, which occur over longer timescales. Volkoff and Cerri (1988) showed that for Brazilian soil profiles, current vegetative cover was only in direct equilibrium with topsoil (A horizon) organic C, while that in subsoils was largely unaffected by the nature of vegetative cover. Once the SOC moves to depth (e.g., argillic or spodic horizons), it becomes less accessible to decomposer organisms, as exemplified by the increased depletion of ^{14}C with soil depth (Pressenda et al., 1996).

Scharpenseel et al. (1992) provided estimates of the amount of organic C contained in the vegetation, soil, and annual litterfall associated with various ecosystems (Figure 11.11). Across the tropical, temperate, and boreal forests, a continuous decrease in the amount of plant biomass and litter C is noted, with little change in the amount of organic C stored in soils. The decrease in the ratio of plant biomass C:SOC was associated with an increase in turnover time from 18 to 60 years. Presumably, most of this variation was related to the effect of temperature on litterfall decomposition; however, significant changes in litterfall quality and morphology are also evident. The amount of residue returned to the soil under similar types of vegetation appears to be a function of climatic factors, principally the amount of precipitation; however, this will depend on the nature of the factor most limiting plant growth. Where ample water is available, the amount of residues returned to the soil may be a function of some other factor such as nutrient supply. For example, it has been shown that P fertilization of Australian pasture soils can increase SOC by 150% or more relative to the native condition (Russel, 1960; Barrow, 1969; Ridley et al., 1990).

Where climatic and soil factors are constant, residue placement may become important. A comparison of the amounts of organic C contained in the plant biomass and soils of temperate grassland and forest ecosystems reveals that despite a much smaller amount of plant biomass in the grassland, annual litter C inputs, and SOC contents were approximately twice that of the
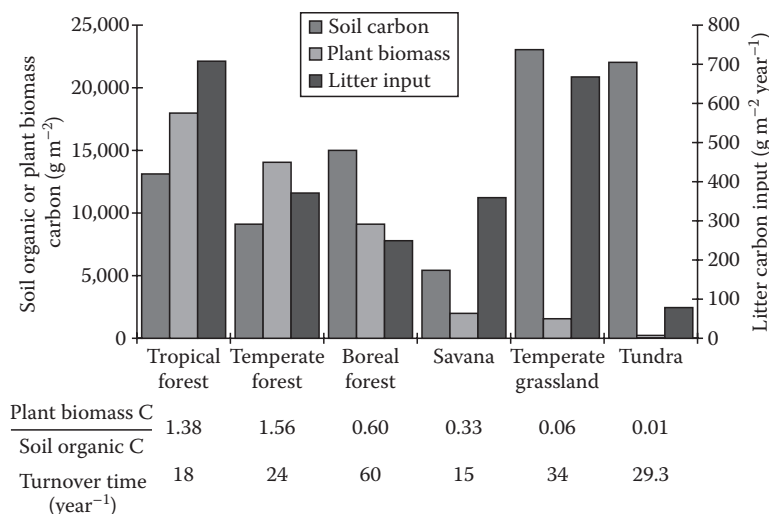
FIGURE 11.11   Variations in mean soil organic C contents, plant biomass C contents, and rate of litter deposition in various ecosystems. (From Scharpenseel, H.W., H.U. Neue, and S. Singer. 1992. Biotransformations in different climatic belts: Source sink relationships, p. 91–105. *In* J. Kubat (ed.) Humus, its structure and role in agriculture and environment. Elsevier Science Publishers, Amsterdam, the Netherlands.)

forests. The occurrence of deep organic-rich mineral horizons in temperate grassland soils (e.g., mollisols), in comparison with the concentration of organic materials in litter layers in boreal forest soil (e.g., spodosols), is an example of the influence that vegetation can have on SOC content and distribution within the soil profile. The apparent larger input of belowground residues in grassland soils compared to forest soils places organic C in close vicinity to the soil mineral components, thereby enhancing the potential for biological protection via the mechanisms discussed in Section 11.1.

The fate of surface-deposited residues depends on the activity of soil microorganisms and fauna and their ability to mix these residues into the surface-mineral horizons. In well-drained soils with high calcium status, the activity of earthworms and other soil fauna is high, leading to a mixing of organic residues through processes of particle-size diminution, ingestion and casting, and bioturbation. Under such conditions, a mull-type HUM layer is formed and litter layers do not develop. Plant residues and their decomposition products are intimately mixed with soil mineral particles, which facilitates potential biological protection through the various organomineral interactions, as discussed in Section 11.5.2. Soils low in calcium do not support as active soil faunal populations and plant residues tend to accumulate on the soil surface forming organic-rich, mor-type HUM layers. Within mor-type HUM, little potential exists for biological protection other than that due to the chemical recalcitrance of highly decomposed residues. The intermediate form of HUM is referred to as a moder.

11.5.3.2 Composition of Plant Materials: The Parent Material for Soil Organic C

Plant materials can be viewed as the parent material for SOC in much the same manner as we view primary minerals as the parent materials of soil mineral components. Plant materials are altered by soil fauna and microorganisms, predominantly after deposition in or on the soil, resulting in changes in the original chemical structure and in the synthesis of new compounds; just as some soil minerals dissolve and others precipitate during pedogenesis. An understanding of the chemical nature of plant materials is therefore important to studies of SOC genesis and composition. Kögel-Knabner (2002) provides a comprehensive review of the molecular composition of plant matter.

Plant materials consist of a range of different compounds varying in concentration across plant species, plant components (e.g., conducting, supporting, or photosynthetic tissues), growth stages, and space (distribution in the landscape). Plant cells can be divided up into three components, the cytoplasm, cell membranes, and cell walls. The cytoplasm contains the simple sugars, organic acids, amino acids, and enzymes essential to maintain metabolic activity. Cell membranes consist of globular proteins embedded within a lipid bilayer. Plant cell-wall components include hemicelluloses, celluloses, lignins, proteins, cuticular, and root waxes. Oades (1989) presented the following average contents for the major types of organic C in plant residues:

1. Extractable materials including water extractables (simple sugars, amino acids, and organic acids) and organic solvent extractables (free and bound alkyl molecules including fats, oils, and waxes)—200 g kg⁻¹
2. Hemicelluloses—200 g kg⁻¹
3. Celluloses—300 g kg⁻¹
4. Lignins—200 g kg⁻¹
5. Proteins—60 g kg⁻¹

The organic components of plant cell walls account for the majority of the mass of plant residues deposited in soils. Carbohydrate structures consist mainly of the polysaccharides cellulose and hemicellulose. Cellulose is the primary component of cell walls with a dominant structure of D-glucopyranose

Cellulose
D-glucopyranose residues linked by β-1,4-linkages

Noncellulosic polysaccharides (hemicellullose)
Pectin with α-1,4-linkages

Chitin
N-acetylglucosamine with β-1,4-linkages

Lignin precursors

Hydroxycinnamyl
alcohol   Coniferyl alcohol   Sinapyl alcohol

Examples of β-O-4-linked lignins

Guaiacyl   Syringyl   p-Hydroxyphenol

Cutin and suberin monomers

18-hydroxy-9-octadecenoic acid   9,10,18-trihydroxyoctadecanoic acid

$HOCH_2(CH_2)_7CH=CH(CH_2)_7COOH$

$HOCH_2(CH_2)_7CHCH(CH_2)_7COOH$

FIGURE 11.12   Representative chemical structures of the organic macromolecules found in plant and microbial residues entering the soil.

residues linked into a polymer via β-1,4-linkages (Figure 11.12). Cellulose can exist in either a crystalline or amorphous state as indicated by x-ray diffraction (Atalla and Vanderhart, 1984) and solid-state ^{13}C NMR (Vanderhart and Atalla, 1984). The crystalline state is more highly resistant to microbial and enzymatic degradation than the amorphous form (Ljungdahl and Eriksson, 1985). Hemicellulose is defined as the polysaccharide extractable in alkali solution. The hemicelluloses exist as linear and branched polymers of D-xylose, L-arabinose, D-mannose, D-glucose, D-galactose, and D-glucuronic acid monomers, which may be acetylated or methylated. Most hemicelluloses are composed of 2–6 of these monomers linked together primarily via a 1,4-β-linkage backbone as shown in Figure 11.12 for pectin, a glucuronic acid polymer.

Lignin represents the second most abundant organic compound in plant residues and accounts for approximately 5% of the mass of grasses and up to 30% of the mass of hardwood forest species (Haider, 1992). The basic building block of lignin, coumaryl alcohol, can be substituted with none, one, or two methoxyl groups at the C-3 and C-5 positions on the benzene ring to produce the p-hydroxyphenol, guaiacyl, and syringyl lignin monomeric units, respectively (Figure 11.12). The units are then linked together by more than 12 possible interunit linkages based on C–O or C–C bonds (McDougall et al., 1993). The major interunit linkage, accounting for about 60% of the linkages, is the β-*O*-4-linkage depicted in Figure 11.12 for the three lignin monomeric units. The nature of the lignin molecule changes with plant type: softwoods (gymnosperms) are dominated by

guaiacyl-based lignin, hardwoods (angiosperms) contain a mixture of guaiacyl- and syringyl-based lignin, and grasses are dominated by syringyl lignin. Results presented by Hedges et al. (1985) suggested that such changes in lignin composition can affect its biodegradability, with syringyl lignin being more susceptible to decomposition than guaiacyl lignin.

The protein and water-soluble components of plant residues, unless protected against biological attack, provide a readily decomposable substrate capable of supplying the chemical energy and nutrients required to drive soil biological processes. Enzymatic cleavage of the peptide linkages to form amino acids and mineralization of amino acid N to form NH_4^+ provide sources of N for soil biological processes, and the abiotic chemical processes, to be discussed subsequently.

Alkyl components of plant materials include free and bound lipids, polyesters, and nonsaponifiable alkyl C dominated biopolymers. Free and bound lipids represent a heterogeneous group of neutral and polar molecules, which are classified together based on their solubility in organic solvents (Tegelaar et al., 1989). The neutral component consists of triacylglycerols and waxes, which serve to protect external plant surfaces and to store energy. The polar component is dominated by the esterified fatty acids found in cell membranes. Insoluble polyesters, derived from hydroxy fatty acids, are found in cutin in plant cuticles and in suberin in roots. Cutin and suberin are composed of various long-chain (C_{16} and C_{18}), substituted fatty acids. The main substituent group is hydroxyl with lesser amounts of epoxy, ketone, and carboxyl groups also present (Holloway, 1982).

Two examples of cutin and suberin monomers are presented in Figure 11.12, and Tegelaar et al. (1989) have presented figures showing additional monomeric chemical structures and a proposed model of the structure of intact cutin and suberin. Plant cuticles and roots have also been shown to contain nonsaponifiable aliphatic biopolymers, which have been labeled cutan and suberan (Nip et al., 1986a, 1986b, 1987). Cutan and suberan are considered similar to cutin and suberin with the exception that they are highly cross-linked by nonester bonds.

In order to assess the influence of plant-residue composition on decomposition and mineralization, it is essential to remove other confounding effects such as climatic, soil, and biological parameters. In field studies, this can be accomplished by examining decomposition of all residues of interest at a single site. Vedrova (1997) assessed the impact of forest species on litter decomposition rates by placing litter collected from each species on small plots located within a single unforested site. Mean rates of mineralization measured for cedar, pine, larch, spruce, aspen, and birch litter over ≈2 years were 1.93, 1.57, 1.85, 2.20, 2.56, and 2.57 mg C g^{-1} litter C day^{-1}, respectively. A limitation of such studies is demonstrated, however, by the work of Elliott et al. (1993), in which the decomposition of four different forest litters (mixed hardwood, red pine, beech, and hemlock) was examined in each of the original four forest types. The rates of decomposition were principally a function of litter type, with mixed hardwood litter decomposing the fastest and hemlock litter the slowest. However, with the exception of the mixed hardwood litter, decomposition rates of the individual litter types were highest when they were placed in the forest type from which they were derived (i.e., decomposition of the hemlock litter was greatest in the hemlock forest). This interaction between litter type and forest type suggests that decomposition pathways in any given ecosystem may be tailored to the type of litter deposited. Thus, the results of decomposition studies where litters are removed from their ecosystem of origin, or where the community structure of the decomposer organisms is altered, may not accurately reflect the relative effects of residue composition on decomposability.

11.5.3.3 Relative Impacts of Soil Fauna and Microorganisms

The requirement of soil organisms for chemical energy and nutrients drives processes of heterotrophic decomposition in soils, which account for the major pathways through which SOC is mineralized. Abiotic chemical oxidation is unlikely to account for >20% of total C mineralization (Moorhead and Reynolds, 1989) and more often accounts for <5% (Lavelle et al., 1993). Microorganisms are the major contributors to soil respiration and are responsible for 80%–95% of the mineralization of C (Brady, 1990; Hassink et al., 1994). Hassink et al. (1994) calculated that the contribution of the fauna to C mineralization in two sandy and two loamy grassland soils to range from 5% to 13% of the total C mineralization. The pattern of C mineralization by the soil fauna through time differed from that of total C mineralization, suggesting that the activity of the soil fauna did not contribute substantially to the differences in total C

mineralization observed between the soils. Hassink et al. (1993) concluded that soil protozoa and nematodes did not significantly influence soil C mineralization despite a positive response of bacterivorous nematodes on the amount of N mineralized. Several other studies have shown that soil fauna enhanced nutrient mineralization had both positive and negative effects on SOC mineralization (Griffiths, 1994; Kajak, 1995; Alphei et al., 1996). In a study including protozoa, nematodes, and earthworms, Alphei et al. (1996) noted that none of the fauna studied significantly affected basal respiration. However, other studies showed that earthworm invasion can have a marked potential to alter (usually reduce) soil C storage on local and regional scales (Alban and Berry, 1994; Bohlen et al., 2004).

The role of soil fauna in decomposition processes should not be based only on their direct contribution to C mineralization. Soil fauna also act to reduce the particle size of litter, distribute it within the soil, transport otherwise immobile microorganisms to new sites within the soil matrix, and prime microorganism activities by the production of readily available substrates (e.g., earthworm intestinal mucus). In so doing, soil fauna generally enhance microbial activity and rates of decomposition. Soil conditions, which limit (e.g., water saturation and the development of anaerobic conditions) or enhance (e.g., tillage or installation of drains in imperfectly drained soils) the activity of soil microorganisms or fauna, will also impact significantly SOC mineralization rates and thus alter SOC levels.

11.5.3.4 Composition of the Microbial Community

The population of decomposer microorganisms in soil is extensive; densities up to 10^{10} bacteria and several kilometers of fungal hyphae per gram of soil have been measured in a wide range of soils (Lavelle et al., 1993). As a result of the diversity of decomposer organisms, the existence of interactions between specific types of organic residue and species of decomposer organisms can have pronounced effects on the chemical structure and biological availability of residual organic materials. The decomposition of woody materials provides an excellent example of how the species composition of the decomposer population can influence the chemical nature of decomposition products. Laboratory incubations of *Eucryphia cordifolia* wood with a brown-rot fungus (unidentified species) and a white-rot fungus (*Ganoderma australe*) showed a more selective utilization of carbohydrate C by the brown-rot fungus and a delignification by the white-rot fungus (Martínez et al., 1991). Using the same white-rot fungus in a solid-state fermentation procedure with beech wood, Martínez et al. (1991) noted little change in the chemical composition of the wood, despite a 36% mass loss. Barrasa et al. (1992) obtained similar results in an ultrastructural study. Selective delignification of *Laurelia philippiana* wood by the white-rot fungus *Phlebia chrysocrea* was noted, but decomposition of the same wood by *G. australe* resulted in increased lignin contents. The selective degradation of carbohydrates by brown-rot fungi appears to occur independently of the fungal or wood species involved. However, the presence of a selective or nonselective degradation process for white-rot fungi

appears to depend on interactions between the species of fungus and wood. Under anaerobic conditions, the activity of obligate aerobes such as wood-degrading fungi is limited and bacterial decomposition processes dominate. In examinations of buried woods, it has been found that decomposition processes invariably result in a preferential utilization of carbohydrates and a concentration of lignin (e.g., Bates and Hatcher, 1989; Bates et al., 1991). Such data indicate that changes in species composition of the decomposer community can significantly alter the decomposition processes and thus, rates of accumulation or loss of organic C from soils.

Earlier in this chapter, a conceptual framework to describe the controls over the decomposition process was presented (Figure 11.5) that was built around the concepts of biological capability and capacity. Various methods can provide information regarding the biological capability and capacity of decomposer communities and the way these properties can be affected by environmental parameters and management practices. For example, the potential degradative capabilities of decomposer communities can be measured using substrate utilization profiles based on Biolog plates (e.g., Bochner, 1989; Bucher and Lanyon, 2005) or microrespirometry like the MicroResp™ method (e.g., Campbell et al., 2003; Wakelin et al., 2008), which is a relatively recent method of community level physiological profiling (CLPP) and uses whole soil samples rather than soil extracts thereby eliminating extraction bias. Furthermore, the extraction and analysis of DNA and RNA (e.g., PCR-DGGE) from soils as well as phospholipid fatty acids (PLFA) can provide an indication of the genetic diversity and structure of soil microbial communities (Muyzer et al., 1993; Widmer et al., 2001; Crecchio et al., 2004; Wakelin et al., 2008). Several indices and multivariate statistical analyses can be used with these techniques to determine the influence of soil and environmental properties on the capability and capacity of the microbial community. Several factors have already been investigated like the influence of soil type (Schutter and Dick, 2000; Banu et al., 2004; Wakelin et al., 2008), crop rotation (Bending et al., 2004; Crecchio et al., 2004), application of various fertilizers (Bucher and Lanyon, 2005), vegetation (De Fede et al., 2001), and agricultural management practice (Wakelin et al., 2008). Although the methods have proven to be reproducible, different results between the different methods have been identified (Widmer et al., 2001). Therefore, it is recommended not to use these methods in isolation if a representative assessment of the composition of the microbial community is desired.

11.5.3.5 Relationship between Organic Residue Composition and Biochemical Recalcitrance

All organic C in soils can serve as a substrate. In addition to the potential mechanisms of biological protection of organic materials offered by the soil mineral fraction, the chemical structure of the organic residue itself can also impart a degree of biochemical recalcitrance. This biochemical recalcitrance of the potential substrate is defined by the strength of intra- and intermolecular bonds, the degree of polymerization and regularity of structural

units in polymers, and the content of aromatic and aliphatic functional groups (Baldock et al., 1997a; Gleixner et al., 2001). Rates of decomposition of known organic substances in soils were reviewed by Paul and van Veen (1978). Although variations in decomposition rates for any single substrate were evident as a result of differences in soils and incubation conditions, simple organic molecules and monomeric compounds decomposed most rapidly. Oades (1989) showed that the extent and rate of mineralization of C for a series of polysaccharides (glucose, dextran, cellulose, and a fungal polysaccharide) decreased with increasing molecular complexity and branching. Similar results were obtained by Martin and Haider (1975) for the mineralization of C from specifically ^{14}C-labeled benzoic and caffeic acid monomers and polymers. C mineralization was most extensive from carboxylic acid groups, less extensive from the aromatic ring C of the monomers, and least extensive from the polymeric aromatic ring C. Of the polymeric materials contained in plant residues, lignin and other polyphenolic C and aliphatic C appear to be the most recalcitrant, but, as discussed in the previous section, the stability of lignin C will be also related to the species composition of the decomposer community.

Many studies have demonstrated a relationship between decomposition and plant-residue characteristics thought to be indicative of residue quality (e.g., Edmonds and Thomas, 1995; Ågren and Bosatta, 1996; Cortez et al., 1996; Hobbie, 1996). Included in these residue characteristics are N concentration, C:N ratios, lignin and/or polyphenol concentration, lignin:nitrogen ratios, and acid-soluble carbohydrates (Heal et al., 1997). Ågren and Bosatta (1996) found that the proportions of extractable, acid-soluble, and acid-insoluble C obtained from a conventional chemical fractionation could be used to assess the quality of forest litter, particularly when the acid-insoluble fraction did not dominate.

During the decomposition of plant residues, significant changes in chemical composition of residual C are evident (Baldock et al., 1997a). In response to such changes, Berg and Staaf (1980) proposed a model of litter decay in which decomposition was controlled initially by N content but subsequently by lignin concentration. This was supported by the results of Edmonds and Thomas (1995), which showed that organic C mineralization rates from green needles of western hemlock and pacific silver fir were initially similar, but became more a function of litter chemistry (e.g., lignin:N ratio) as decomposition progressed.

11.5.4 Topography

Topography exerts its major control over SOC contents through a modification of climate and soil textural factors and through its impacts on the redistribution of water within a landscape. Soils in downslope positions are often wetter and have finer textures than soils in upslope positions or at the top of knolls. Topographic-induced changes in the soil microenvironment often lead to changes in plant communities (Sebastia, 2004), which in turn can influence the magnitude and quality of residue returns. Burke et al. (1995) examined the extent to which

SOC content varied at a landscape scale at two sites differing in soil texture but having similar climatic characteristics. Burke et al. (1995) noted increased organic C contents (and clay and silt contents) in downslope positions relative to the summits at both sites. Such a finding has been attributed to the downslope movement of organic C and organic-rich clay (Reiners, 1983). However, additional gradients in available water along slopes, especially in water-limited systems, influence plant production (Peterson et al., 1988), with greater biomass inputs and greater potential biological protection of organic C via higher clay contents at the base of slopes. Where excessive water exists, drainage of depressions in the landscape can be restricted, leading to the development of anaerobic conditions and a preservation of organic C relative to the better-drained higher landscape elements during wetter times of the year.

Guo et al. (2006) showed that the SOC decreases as elevation increases and that level topography had twice the SOC content of other slope classes. Particularly, in mountain areas that are by nature highly heterogeneous, taking topography into account is required to predict SOC stocks across the landscape. In a study investigating the SOC storage in mountain grasslands of the Pyrenees, Garcia-Pausas et al. (2007) found that the SOC stocks were particularly low at high altitudes probably as a result of an overall temperature limitation of NPP. Climatic and topographic variables were able to predict a significant part of the C storage variability in the mountain grasslands examined by Garcia-Pausas et al. (2007). The microclimate conditions related to topographic position (aspect and slope) were identified as important factors for predicting C storage in soils of the high-altitude grasslands and should be taken into account to achieve accurate estimations of C stocks in mountain ecosystems.

As a consequence of all the possible confounding factors, it is difficult to study the sole effects of topography. For example, the interactive effects of topography (depositional and erodible zones) and tillage (conventional versus minimum tillage [MT]) on the redistribution of C, N, and P within an agricultural landscape were examined by De Gryze et al. (2008). In general, organic matter content and moisture content were greater in lower, depositional areas compared to erodible areas, and the impact of topography on the stabilization and redistribution processes of nutrients was more pronounced in conventional tillage (CT) than MT. This last finding clearly demonstrates that interactions between topographic characteristics and land management practices should be considered in regional inventory assessments of SOC (De Gryze et al., 2008).

11.5.5 Land Management Practices

Paustian et al. (1997b) reviewed the influence of agricultural management practices on SOC levels. Hutchinson et al. (2007) summarized the relative rates of SOC change for a range of "carbon friendly" agricultural management treatments and found that most changes were <0.6 Mg C ha^{-1} year^{-1}. The influence of forestry management practices has also been reviewed (Johnson, 1992; Johnson and Curtis, 2001; Johnson et al., 2002).

The most dramatic influence of agricultural practices occurs when soils are first brought into production. Typically, SOC levels decrease for the first few decades after cultivation and then stabilize at a new equilibrium level, which is dictated principally by the ability of the soil to protect organic C and the amount, quality, and distribution of plant-residue inputs. For example, Haas et al. (1957) observed a loss of 28%–59% of the SOC following 30–43 years of cropping at 11 sites within the North American prairies. David et al. (2009) also measured reductions SOC (30%–50%) due to the conversion of prairies to annual cultivation and artificial drainage using archived samples and long-term resampling of soils.

The following characteristics of crop production systems, in comparison with those of native grasslands, help to explain the observed losses of SOC induced by cultivation:

1. 80% lower allocation of organic C to soils (Buyanovsky et al., 1987)
2. Reduced belowground allocation of photosynthate (Anderson and Coleman, 1985)
3. Enhanced aggregate disruption and exposure of physically protected organic C due to cultivation
4. Enhanced rates of decomposition of available organic C substrates due to more favorable abiotic conditions (e.g., aeration, temperature, and water content)

In practice, continuous alteration of land management and cropping practices (e.g., adoption of reduced tillage, the inclusion, or removal of pasture) will lead to a system where SOC levels are always in a state of flux, increasing with some practices and decreasing with others (Baldock and Skjemstad, 1999). Only if management practices are left in place long enough, it is possible to gain an indication of what the new equilibrium SOC value would be for any given land management system.

Field trials set up to examine the impact of fertilizer additions on SOC content, have revealed that the addition of N fertilizers typically enhances SOC contents. This is particularly true in temperate-zone ecosystems where N availability is thought to limit NPP (Peterson and Melillo, 1985; Schimel et al., 1996; Holland et al., 1997). Explanatory mechanisms suggest that N-fertilizer additions result in a greater return of plant residues to soils due to enhanced production, a reduction in decomposition rates due to enhanced soil drying (Andrén, 1987), a promotion of soil acidification (Thurston et al., 1976), a repression of lignolytic enzymes, and a formation of recalcitrant humic materials through the reaction of amino acids with humic precursors (Fog, 1988).

In tropical systems, the establishment of pastures after clearing of forests is widespread (Sombroek et al., 1993). Pasture establishment immediately after deforestation, using species with high proportions of belowground biomass, may increase SOC contents as demonstrated in the Brazilian Amazon (Serrâo et al., 1979), Latin America (Ligel, 1992), and East Africa (Boonman, 1993).

The intensity with which a soil is cultivated can impact both the total amount of SOC and its distribution with soil depth. No-till systems tend to concentrate residue inputs at the soil

surface and generally enhance soil organic C and N contents in soil surface layers (Angers and Eriksen-Hamel, 2008). To accurately evaluate the influence of tillage practices on SOC stocks, it is important to collect soil samples beyond the depth of tillage and to account for variations in soil bulk density. Paustian et al. (1997b) presented data from a number of long-term field trials indicating that SOC retention is typically enhanced under no-till relative to more intensive CT systems. However, in a review of Australian publications, Valzano et al. (2005) found that a positive effect of no-till or reduced till on SOC in the 0–30-cm soil layer could only be demonstrated at high annual rainfalls (>660 mm). Valzano et al. (2005) also noted that the separation of residue handling from the physical effect of passing an implement through the soil was difficult to separate given the wide range of residue handling practices associated with the different tillage systems.

11.6 Contribution of Soil Organic Matter to the Global Carbon Cycle

The organic carbon contained within SOM represents a significant reservoir of carbon within this global carbon cycle (Figure 11.13). SOC has been estimated to account for 1200–1550 Pg of C (1 Pg = 10^{15} g) to a depth of 1 m and for 2300–2450 Pg of C to a depth of 2 or 3 m (Eswaran et al., 1995; Jobbagy and Jackson, 2000; Lal, 2004a; Houghton, 2005). Comparative estimates of organic C contained in living biomass (550–560 Pg) and the atmosphere CO_2–C (760–780 Pg; Lal, 2004a; Houghton, 2005) indicate that variations in the size of the SOC store could significantly alter atmospheric CO_2–C concentrations. For example, a 5% shift in the amount of SOC stored in the 0–2 m soil profile has the potential to alter atmospheric CO_2–C by up to 16%.

Land-use changes can induce either a net emission or a net sequestration of organic carbon in soil depending on the resultant balance between losses and inputs. The net change in inputs and losses will be defined by the soil and environmental properties and land management practices discussed previously. Sequestration of organic carbon in soil is a slow process typically requiring decades, but is suggested to offer the most efficient natural strategy for offsetting increased atmospheric CO_2–C concentrations (Metting et al., 1999; Post et al., 1999; Lal, 2004b). Over the next century, improved land management strategies may have the capacity to sequester up to 150 Pg of CO_2–C (Lal et al., 1998; Lal, 2004a). Considerable uncertainty exists in such estimates because of an inability to accurately predict the potential sequestration of carbon that is possible in soil. Improving our understanding of SOC cycling processes and how these are affected by environment and land management practices will be vital to identify opportunities for building SOM and sequestering carbon in soils.

Estimates of the potential to sequester carbon in agricultural soils have been made for the United States (Lal et al., 1998), Canada (Bolinder et al., 2008), China (Han et al., 2005, 2006), the European Union (Freibauer et al., 2004; Janssens et al., 2005; Romanenkov et al., 2007), and South America (Cerri et al., 2004, 2006). Such estimates are typically based on the use of long-term field experiments or simulation modeling. Where estimates are based on field experiments, it is important to ensure that the potential for saturation of the soils capacity to protect carbon from decomposition is acknowledged and that measured carbon sequestration rates are not projected unimpeded into the future. A similar consideration must be applied to soil carbon simulation modeling activities. Most soil carbon models apply first-order kinetics to the decomposition of component carbon pools (Paustian, 1994). Therefore, predicted equilibrium carbon stocks are linearly proportional to carbon inputs (Paustian et al., 1997a, 1997b; Six et al., 2002), and modeling activities predict that unlimited increases in soil carbon stocks can occur provided inputs of carbon to the soil can continue to increase.
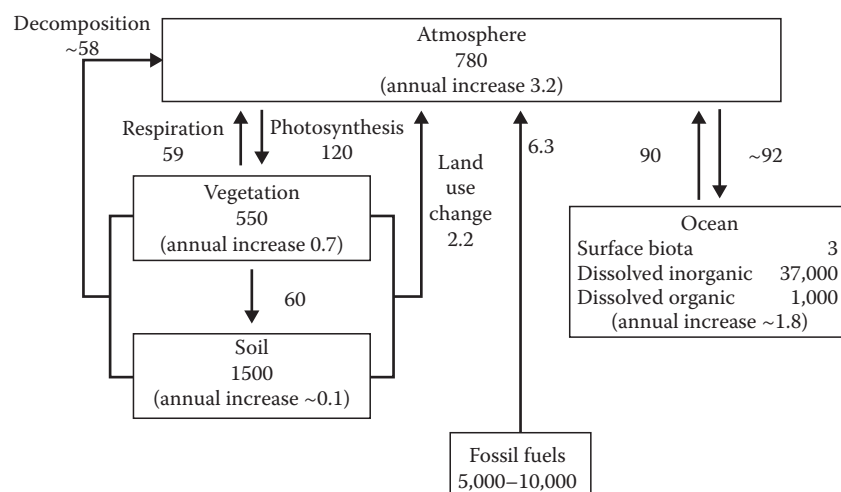
FIGURE 11.13 The global carbon cycle. All pool sizes are in units of Pg C and all fluxes are given in units of Pg C year^{-1}. (Adapted from Houghton, R.A. 2005. The contemporary carbon cycle, p. 473–513. *In* W.H. Schlesinger (ed.) Biogeochemistry. Elsevier Science, Amsterdam, the Netherlands.)

Since carbon inputs to soil will be limited by constraints placed on photosynthesis (e.g., availability of water, nutrients, heat, and ultimately light) and current soil carbon simulation models do not define a maximum soil carbon protective capacity, modeled estimates of potential carbon sequestration in soils need to be carefully scrutinized.

The concept of soil carbon saturation suggests that each soil has a unique carbon saturation level dictated by soil properties including texture, mineralogy, bulk density, and depth (Ingram and Fernandes, 2001; Six et al., 2002; Stewart et al., 2008a, 2008b). Soil carbon sequestration potential should therefore be estimated as the difference between the current and saturated soil carbon content. Defining soil carbon saturation values has proven difficult, although Hassink (1992) suggested that the amount of carbon obtained under long-term pastures may provide an indication. Furthermore, true soil carbon saturation values may be of limited importance if the inputs of organic carbon required to attain and maintain such levels are beyond the capture of carbon by photosynthesis and deposition within soil. Of more practical interest would be the behavior of soils as they approach their carbon saturation capacity as well as the influence of soil carbon saturation deficit on the efficiency of the SOC accumulation in saturated soils.

Although soil carbon does offer the possibility for sequestering atmospheric CO_2–C, the potential for soil carbon sequestration in soil is small compared to projected emission of CO_2 form current energy practices (about 150 Pg versus 600 Pg over the next 100 years). Soil carbon sequestration is also finite, an upper limit exists, and as soils move toward the upper limit, the rate of soil carbon sequestration will diminish. Enhancing soil carbon form atmospheric CO_2–C sequestration should therefore be viewed as a mechanism for "buying time" for the development and implementation of longer-lasting measures for reducing fossil fuel emissions (Watson et al., 2000; Houghton, 2005). However, it is also important to recognize that increasing sequestration of carbon in soils by building SOM may offer the added benefit of increased productivity through mechanisms discussed previously.

11.7 Summary

Soil organic matter is a complex mixture of a variety of materials derived initially from photosynthesis by plants but then altered through decomposition processes that enhance its diversity in composition. Such compositional diversity provides the SOM with the capability to contribute beneficially to the many functions that it serves in a soil. Key goals for future work on SOM should include the following:

1. An acknowledgment of the diversity in SOM composition and a development of fractionation systems capable of quantitatively dividing SOM up into biologically relevant components
2. Development of predictive relationships that quantify the role of each component to the various functions to which SOM contributes and an assessment of any soil type specificity that may exist
3. Parameterization of carbon cycling models build on measurable components and inclusion of the quantitative relationships between SOM composition and soil properties (functions) to allow the prediction of both SOM dynamics and the subsequent impacts on soil properties

With the potential for large-scale introduction of carbon trading schemes and a requirement to reduce the emission of greenhouse gases, a more complete understanding of SOM dynamics and how altering SOM will alter soil properties will be essential. Such understanding is required to develop viable land management practices that accommodate the potentially conflicting issues of maintaining food security for an increasing global population, ensuring profitability of individual farming enterprises, and enhancing the capture of atmospheric CO_2 as organic matter in soils while reducing emissions of other greenhouse gases (nitrous oxide and methane).

References

Achard, F.K. 1786. Chemische utersunchungen des torts. Crell's Chem. Ann. 2:391–403.

Ågren, G.I., and E. Bosatta. 1996. Quality: A bridge between theory and experiment in soil organic matter studies. Oikos 76:522–528.

Alban, D.H., and E.C. Berry. 1994. Effects of earthworm invasion on morphology, carbon, and nitrogen of a forest soil. Appl. Soil Ecol. 1:243–249.

Alphei, J., M. Bonkowski, and S. Scheu. 1996. Protozoa, nematoda and lumbricidae in the rhizosphere of *Hordelymus europaeus* (Poaceae): Faunal interactions, response of microorganisms and effects on plant growth. Oecologia 106:111–126.

Amato, M.A., and J.N. Ladd. 1992. Decomposition of ^{14}C-labelled glucose and legume material in soils: Properties influencing the accumulation of organic residue C and microbial biomass C. Soil Biol. Biochem. 24:455–464.

Amelung, W., R. Bol, and C. Friedrich. 1999. Natural ^{13}C abundance: A tool to trace the incorporation of dung-derived carbon into soil particle-size fractions. Rapid Commun. Mass Spectrom. 13:1291–1294.

Amelung, W., K.W. Flach, and W. Zech. 1997. Climatic effects on soil organic matter composition in the great plains. Soil Sci. Soc. Am. J. 61:115–123.

Amelung, W., and W. Zech. 1996. Organic species in ped surface and core fractions along a climosequence in the prairie, North America. Geoderma 74:193–206.

Amelung, W., and W. Zech. 1999. Minimisation of organic matter disruption during particle-size fractionation of grassland epipedons. Geoderma 92:73–85.

Anderson, D.W. 1995. The role of nonliving organic matter in soils, p. 81–92. *In* R.G. Zepp and C. Sonntag (eds.) Role of nonliving organic matter in the earth's carbon cycle. Vol. 16. John Wiley & Sons Ltd., Chichester, West Sussex, England.

Anderson, D.W., and D.C. Coleman. 1985. The dynamics of organic matter in grassland soils. J. Soil Water Conserv. 40:211–216.

Anderson, J.P., and K.H. Domsch. 1973. Quantification of bacterial and fungal contributions to soil respiration. Arch. Mikrobiol. 93:113–127.

Anderson, J.P.E., and K.H. Domsch. 1978. Physiological method for quantitative measurement of microbial biomass in soils. Soil Biol. Biochem. 10:215–221.

Anderson, D.W., and E.A. Paul. 1984. Organo-mineral complexes and their study by radiocarbon dating. Soil Sci. Soc. Am. J. 48:298–301.

Andrén, O. 1987. Decomposition of shoot and root litter of barley, lucerne and meadow fescue under field conditions. Swed. J. Agric. Res. 17:113–122.

Angers, D.A., and N.S. Eriksen-Hamel. 2008. Full-inversion tillage and organic carbon distribution in soil profiles: A meta-analysis. Soil Sci. Soc. Am. J. 72:1370–1374.

Angers, D.A., S. Recous, and C. Aita. 1997. Fate of carbon and nitrogen in water-stable aggregates during decomposition of ^{13}C ^{15}N labelled wheat straw. Eur. J. Soil Sci. 48:295–300.

Appuhn, A., R.G. Joergensen, M. Raubuch, E. Scheller, and B. Wilke. 2004. The automated determination of glucosamine, galactosamine, muramic acid, and mannosamine in soil and root hydrolysates by HPLC. J. Plant Nutr. Soil Sci. 167:17–21.

Asadu, C.L.A., J. Diels, and B. Vanlauwe. 1997. A comparison of the contributions of clay, silt, and organic matter to the effective CEC of soils of subSaharan Africa. Soil Sci. 162:785–794.

Atalla, R.H., and D.L. Vanderhart. 1984. Native cellulose: A composite of two distinct crystalline forms. Science 223:283–285.

Bachmann, J., G. Guggenberger, T. Baumgartl, R.H. Ellerbrock, E. Urbanek, M.O. Goebel, K. Kaiser, R. Horn, and W.R. Fischer. 2008. Physical carbon-sequestration mechanisms under special consideration of soil wettability. J. Plant Nutr. Soil Sci. 171:14–26.

Baldock, J.A. 2002. Interactions of organic materials and microorganisms with minerals in the stabilization of soil structure, p. 85–131. *In* P.M. Huang et al. (eds.) Interactions between soil particles and microorganisms and the impact on the terrestrial ecosystem. John Wiley & Sons, New York.

Baldock, J.A. 2007. Composition and cycling of organic carbon in soil, p. 1–35. *In* P. Marschner and Z. Rengel (eds.) Soil biology. Vol. 10. Nutrient cycling in terrestrial ecosystems. Springer-Verlag, Berlin, Germany.

Baldock, J.A., C.A. Masiello, Y. Gélinas, and J.I. Hedges. 2004. Cycling and composition of organic matter in terrestrial and marine ecosystems. Mar. Chem. 92:39–64.

Baldock, J.A., and P.N. Nelson. 2000. Soil organic matter, p. B25–B84. *In* M. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Baldock, J.A., J.M. Oades, P.N. Nelson, T.M. Skene, A. Golchin, and P. Clarke. 1997a. Assessing the extent of decomposition of natural organic materials using solid-state ^{13}C NMR spectroscopy. Aust. J. Soil Res. 35:1061–1084.

Baldock, J.A., J.M. Oades, A.M. Vassallo, and M.A. Wilson. 1989. Incorporation of uniformly labeled C-13 glucose carbon into the organic fraction of a soil. Carbon balance and CP MAS C-13 NMR measurements. Aust. J. Soil Res. 27:725–746.

Baldock, J.A., J.M. Oades, A.G. Waters, X. Peng, A.M. Vassallo, and M.A. Wilson. 1992. Aspects of the chemical structure of soil organic materials as revealed by solid-state ^{13}C NMR spectroscopy. Biogeochemistry 16:1–42.

Baldock, J.A., T. Sewell, and P.G. Hatcher. 1997b. Decomposition induced changes in the chemical structure of fallen red pine, white spruce and tamarack logs, p. 75–83. *In* G. Cadisch and K.E. Giller (eds.) Driven by nature: Plant litter quality and decomposition. CAB International, Wallingford, U.K.

Baldock, J.A., and J.O. Skjemstad. 1999. Soil organic carbon/soil organic matter, p. 159–170. *In* K.I. Peverill et al. (eds.) Soil analysis: An interpretation manual. CSIRO Publishing, Collingswood, Victoria, Australia.

Baldock, J.A., and J.O. Skjemstad. 2000. Role of the soil matrix and minerals in protecting natural organic materials against biological attack. Org. Geochem. 31:697–710.

Baldock, J.A., and R.J. Smernik. 2002. Chemical composition and bioavailability of thermally altered *Pinus resinosa* (Red pine) wood. Org. Geochem. 33:1093–1109.

Banu, N.A., B. Singh, and L. Copeland. 2004. Microbial biomass and microbial biodiversity in some soils from New South Wales, Australia. Aust. J. Soil Res. 42:777–782.

Barrasa, J.M., A.E. Gonzalez, and A.T. Martinez. 1992. Ultrastructural aspects of fungal delignification of Chilean woods by *Ganoderma australe* and *Phlebia chrysocrea*. A study of natural and *in vitro* degradation. Holzforschung 46:1–8.

Barrow, N.J. 1969. The accumulation of soil organic matter under pasture and its effect on soil properties. Aust. J. Exp. Agric. Anim. Husb. 9:437–444.

Bates, A.L., and P.G. Hatcher. 1989. Solid-state C-13 NMR studies of a large fossil gymnosperm from the Yallourn open cut, Latrobe Valley, Australia. Org. Geochem. 14:609–617.

Bates, A.L., P.G. Hatcher, H.E. Lerch, C.B. Cecil, S.G. Neuzil, and Supardi. 1991. Studies of a peatified angiosperm log cross section from Indonesia by nuclear magnetic resonance spectroscopy and analytical pyrolysis. Org. Geochem. 17:37–45.

Beare, M.H., M.L. Caberra, P.F. Hendrix, and D.C. Coleman. 1994a. Aggregate-protected and unprotected organic matter pools in conventional- and no-tillage soils. Soil Sci. Soc. Am. J. 58:787–795.

Beare, M.H., P.F. Hendrix, and D.C. Coleman. 1994b. Water-stable aggregates and organic matter fractions in conventional and no-tillage soils. Soil Sci. Soc. Am. J. 58:777–786.

Beck, M.A., and P.A. Sanchez. 1994. Soil phosphorus fraction dynamics during 18 years of cultivation on a Typic Paleudult. Soil Sci. Soc. Am. J. 58:1424–1431.

Beldin, S.I., B.A. Caldwell, P. Sollins, E.W. Sulzman, K. Lajtha, and S.E. Crow. 2007. Cation exchange capacity of density fractions from paired conifer/grassland soils. Biol. Fertil. Soils 43:837–841.

Bending, G.D., M.K. Turner, F. Rayns, M.-C. Marx, and M. Wood. 2004. Microbial and biochemical soil quality indicators and their potential for differentiating areas under contrasting agricultural management regimes. Soil Biol. Biochem. 36:1785–1792.

Bengtsson, G., P. Bengtson, and K.F. Månsson. 2003. Gross nitrogen mineralization-, immobilization-, and nitrification rates as a function of soil C/N ratio and microbial activity. Soil Biol. Biochem. 35:143–154.

Berg, B., and H. Staaf. 1980. Decomposition rate and chemical changes of Scots pine needle litter. II. Influence of chemical composition. Ecol. Bull. 32:373–390.

Bettany, J.R., J.W.B. Stewart, and S. Saggar. 1979. Nature and forms of sulfur in organic matter fractions of soils selected along an environmental gradient. Soil Sci. Soc. Am. J. 43:981–985.

Beyer, L. 1996. The chemical composition of soil organic matter in classical humic compound fractions and in bulk samples— A review. J. Plant Nutr. Soil Sci. 159:527–539.

Bird, M.I., A.R. Chivas, and J. Head. 1996. A latitudinal gradient in carbon turnover times in forest soils. Nature 381:143–146.

Blair, G.J., R.D.B. Lefroy, and L. Lisle. 1995. Soil carbon fractions based on their degree of oxidation, and the development of a carbon management index for agricultural systems. Aust. J. Agric. Res. 46:1459–1466.

Bochner, B.R. 1989. Sleuthing out bacterial identities. Nature 339:157–158.

Bohlen, P.J., P.M. Groffman, T.J. Fahey, M.C. Fisk, E. Suarez, D.M. Pelletier, and R.T. Fahey. 2004. Ecosystem consequences of exotic earthworm invasion of north temperate forests. Ecosystems 7:1–12.

Bolinder, M.A., O. Andren, T. Katterer, and L.E. Parent. 2008. Soil organic carbon sequestration potential for Canadian agricultural ecoregions calculated using the introductory carbon balance model. Can. J. Soil Sci. 88:451–460.

Boonman, J.G. 1993. East Africa's grasses and fodders: Their ecology and husbandry. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Boudot, J.P., A. Bel Hadi Brahim, and T. Chone. 1986. Carbon mineralisation in andisols and aluminium-rich highland soils. Soil Biol. Biochem. 18:457–461.

Boudot, J.P., A. Bel Hadi Brahim, and T. Chone. 1988. Dependence of carbon and nitrogen mineralisation rates upon amorphous metallic constituents and allophanes in highland soils. Geoderma 42:245–260.

Boudot, J.P., A. Bel Hadi Brahim, R. Steiman, and F. Seigle-Murandi. 1989. Biodegradation of synthetic organo-metallic complexes of iron and aluminium with selected metal to carbon ratios. Soil Biol. Biochem. 21:961–966.

Bracewell, J.M., G.W. Robertson, and K.R. Tate. 1976. Pyrolysis-gas chromatography studies on a climosequence of soils in tussock grasslands, New Zealand. Geoderma 15:209–215.

Brady, N.C. 1990. The nature and properties of soils. Macmillan Publishing Company, New York.

Bremner, J.M. 1968. The nitrogenous constituents of soil organic matter and their role in soil fertility. "Organic matter and soil fertility." Pontif. Acad. Sci. Scripta Varia. 32:143–180.

Brookes, P.C., A. Landman, G. Pruden, and D.S. Jenkinson. 1985. Chloroform fumigation and the release of soil nitrogen. A rapid extraction method to measure microbial biomass nitrogen in soil. Soil Biol. Biochem. 17:837–842.

Brookes, P.C., D.S. Powlson, and D.S. Jenkinson. 1982. Measurement of microbial biomass phosphorus in soil. Soil Biol. Biochem. 14:319–329.

Brookes, P.C., D.S. Powlson, and D.S. Jenkinson. 1984. Phosphorus in the soil microbial biomass. Soil Biol. Biochem. 16:169–175.

Broos, K., L.M. Macdonald, M.S.J. Warne, D.A. Heemsbergen, M.B. Barnes, M. Bell, and M.J. McLaughlin. 2007. Limitations of soil microbial biomass carbon as an indicator of soil pollution in the field. Soil Biol. Biochem. 39:2693–2695.

Bucher, A.E., and L.E. Lanyon. 2005. Evaluating soil management with microbial community-level physiological profiles. Appl. Soil Ecol. 29:59–71.

Burgess, D., J.A. Baldock, S. Wetzell, and D.G. Brand. 1995. Scarification, fertilization and herbicide treatment effects on planted conifers and soil fertility. Plant Soil 168:513–522.

Burke, I.C., E.T. Elliott, and C.V. Cole. 1995. Influence of macroclimate, landscape position, and management on soil organic matter in agroecosystems. Ecol. Appl. 5:124–131.

Burke, I.C., C.M. Yonker, W.J. Parton, C.V. Cole, K. Flach, and D.S. Schimel. 1989. Texture, climate and cultivation effects on soil organic matter in U.S. grassland soils. Soil Sci. Soc. Am. J. 53:800–805.

Buyanovsky, G.A., C.L. Kucera, and G.H. Wagner. 1987. Comparative analysis of carbon dynamics in native and cultivated ecosystems. Ecology 68:2023–2031.

Cadisch, G., and K.E. Giller. 1997. Driven by nature: Plant litter quality and decomposition. CAB International, Wallingford, U.K.

Cajuste, L.J., R.J. Laird, L. Cajuste, and B.G. Cuevas. 1996. Citrate and oxalate influence on phosphate, aluminum, and iron in tropical soils. Commun. Soil Sci. Plant Anal. 27:1377–1386.

Campbell, C.A., V.O. Biederbeck, R.P. Zentner, and G.P. Lafond. 1991. Effect of crop rotations and cultural practices on soil organic matter, microbial biomass and respiration in a thin black Chernozem. Can. J. Soil Sci. 71:363–376.

Campbell, C.D., S.J. Chapman, C.M. Cameron, M.S. Davidson, and J.M. Potts. 2003. A rapid microtiter plate method to measure carbon dioxide evolved from carbon substrate

amendments so as to determine the physiological profiles of soil microbial communities by using whole soil. Appl. Environ. Microbiol. 69:3593–3599.

Canellas, L.P., R. Spaccini, A. Piccolo, L.B. Dobbss, A.L. Okorokova-Facanha, G.D. Santos, F.L. Olivares, and A.R. Facanha. 2009. Relationships between chemical characteristics and root growth promotion of humic acids isolated from Brazilian oxisols. Soil Sci. 174:611–620.

Cerri, C.C., M. Bernoux, C.E.P. Cerri, and C. Feller. 2004. Carbon cycling and sequestration opportunities in South America: The case of Brazil. Soil Use Manage. 20:248–254.

Cerri, C.E.P., C.C. Cerri, M. Bernoux, B. Volkoff, and M.A. Rondon. 2006. Potential of soil carbon sequestration in the Amazonian tropical rainforests, p. 245–266. _In_ R. Lal et al. (eds.) Carbon sequestration in soils of Latin America. Food products press/ Harworthpress, Inc., New York.

Chan, K.Y., D.P. Heenan, and A. Oates. 2002. Soil carbon fractions and relationship to soil quality under different tillage and stubble management. Soil Tillage. Res. 63:133–139.

Chefetz, B., M.J. Salloum, A.P. Deshmukh, and P.G. Hatcher. 2002. Structural components of humic acids as determined by chemical modifications and carbon-13 NMR, pyrolysis, and thermochemolysis-gas chromatography/mass spectrometry. Soil Sci. Soc. Am. J. 66:1159–1171.

Chen, Y., and T. Aviad. 1990. Effects of humic substances on plant growth, p. 161–186. _In_ P. MacCarthy et al. (eds.) Humic substances in soil and crop sciences: Selected readings. American Society of Agronomy, Inc., Soil Science Society of America, Inc., Madison, WI.

Cheng, W.X., and Y. Kuzyakov. 2005. Root effects on soil organic matter decomposition, p. 119–143. _In_ Lobel, W., and S.F. Wright (eds.) Roots and soil management: Interactions between roots and the soil. American Society of Agronomy, Madison, WI.

Christensen, B.T. 1996a. Carbon in primary and secondary organomineral complexes, p. 97–165. _In_ M.R. Carter and B.A. Stewart (eds.) Advances in soil science: Structure and organic matter storage in agricultural soils. CRC Lewis Publishers, Boca Raton, FL.

Christensen, B.T. 1996b. Matching measurable soil organic matter fractions with conceptual pools in simulation models of carbon turnover: Revision of model structure, p. 143–159. _In_ D.S. Powlson et al. (eds.) Evaluation of soil organic matter models using existing long-term datasets. NATO ASI Series I: Global Environmental Change. Vol. 38. Springer-Verlag, Berlin, Germany.

Christensen, B.T. 2001. Physical fractionation of soil and structural and functional complexity in organic matter turnover. Eur. J. Soil Sci. 52:345–353.

Christl, I., and R. Kretzschmar. 2007. C-1s NEXAFS spectroscopy reveals chemical fractionation of humic acid by cation-induced coagulation. Environ. Sci. Technol. 41:1915–1920.

Churchman, G.J., and R.C. Foster. 1994. The role of clay minerals in the maintenance of soil structure, p. 17–34. 15th World Congr. Soil Sci., Acapulco, Mexico, July 10–16, 1994. Transactions, Vol. 8a: Commission VII Symposia, International Society of Soil Science, Acapulco, Mexico.

Clapp, C.E., Y. Chen, M.H.B. Hayes, and H.H. Cheng. 2001. Plant growth promoting activity of humic substances, p. 243–255. _In_ R.S. Swift and K.M. Sparks (eds.) Understanding and managing organic matter in soils, sediments, and waters. International Humic Science Society, Madison, WI.

Cleveland, C.C., and D. Liptzin. 2007. C:N:P stoichiometry in soil: Is there a "redfield ratio" for the microbial biomass? Biogeochemistry 85:235–252.

Clinton, P.W., R.H. Newman, and R.B. Allen. 1995. Immobilization of ^{15}N in forest litter studies by ^{15}N CPMAS NMR spectroscopy. Eur. J. Soil Sci. 46:551–556.

Cortez, J., J.M. Demard, P. Bottner, and L.J. Monrozier. 1996. Decomposition of Mediterranean leaf litters: A microcosm experiment investigating relationships between decomposition rates and litter quality. Soil Biol. Biochem. 28:443–452.

Crecchio, C., A. Gelsomino, R. Ambrosoli, J.L. Minati, and P. Ruggiero. 2004. Functional and molecular responses of soil microbial communities under differing soil management practices. Soil Biol. Biochem. 36:1873–1883.

Curtin, D., C.A. Campbell, and D. Messer. 1996. Prediction of titratable acidity and soil sensitivity to pH change. J. Environ. Qual. 25:1280–1284.

Dalal, R.C. 1998. Soil microbial biomass—What do the numbers really mean? Aust. J. Exp. Agric. 38:649–695.

Dalal, R.C., B.P. Harms, E. Krull, and W.J. Wang. 2005. Total soil organic matter and its labile pools following mulga (_Acacia aneura_) clearing for pasture development and cropping 1. Total and labile carbon. Aust. J. Soil Res. 43:13–20.

da Rosa, C.M., R.M.V. Castilhos, L.C. Vahl, D.D. Castilhos, L.F.S. Pinto, E.S. Oliveira, and O.D. Leal. 2009. Effect of humic-like substances on potassium uptake kinetics, plant growth and nutrient concentration in _Phaseolus vulgaris_ L. Rev. Bras. Cienc. Do Solo 33:959–967.

da Silva, A.P., and B.D. Kay. 1997. Estimating the least limiting water range of soils from properties and management. Soil Sci. Soc. Am. J. 61:877–883.

David, M.B., G.F. McIsaac, R.G. Darmody, and R.A. Omonode. 2009. Long-term changes in mollisol organic carbon and nitrogen. J. Environ. Qual. 38:200–211.

Davidson, E.A., and I.L. Ackerman. 1993. Changes in soil carbon inventories following cultivation of previously untilled soils. Biogeochemistry 20:161–193.

DeAngelis, D.L. 1980. Energy flow, nutrient cycling, and ecosystem resilience. Ecology 61:764–771.

De Fede, K.L., D.G. Panaccione, and A.J. Sexstone. 2001. Characterization of dilution enrichment cultures obtained from size-fractionated soil bacteria by BIOLOG® community-level physiological profiles and restriction analysis of 16S rRNA genes. Soil Biol. Biochem. 33:1555–1562.

De Gryze, S., J. Six, H. Bossuyt, K. Van Oost, and R. Merckx. 2008. The relationship between landform and the distribution of soil C, N and P under conventional and minimum tillage. Geoderma 144:180–188.

De Gryze, S., J. Six, C. Brits, and R. Merckx. 2005. A quantification of short-term macroaggregate dynamics: Influences of wheat residue input and texture. Soil Biol. Biochem. 37:55–66.

De Gryze, S., J. Six, and R. Merckx. 2006. Quantifying water-stable soil aggregate turnover and its implication for soil organic matter dynamics in a model study. Eur. J. Soil Sci. 57:693–707.

Duchaufour, P. 1976. Dynamics of organic matter in soils of temperate regions: Its action on pedogenesis. Geoderma 15:31–40.

Duncan, T.M., 1987. C-13 chemical shielding in solids. J. Chem. Phys. Ref. Data. 16:125–151.

Dunnivant, F.M., P.M. Jardine, D.L. Taylor, and J.F. McCarthy. 1992. Transport of naturally-occurring dissolved organic-carbon in laboratory columns containing aquifer material. Soil Sci. Soc. Am. J. 56:437–444.

Duxbury, J.M., M.S. Smith, and J.W. Doran. 1989. Soil organic matter as a source and sink of plant nutrients, p. 33–67. *In* D.C. Coleman et al. (eds.) Dynamics of soil organic matter in tropical ecosystems. NifTAL Project, University of Hawaii Press, Honolulu, HI.

Edmonds, R.L., and T.B. Thomas. 1995. Decomposition and nutrient release from green needles of western hemlock and Pacific silver fir in an old-growth temperate rainforest, Olympic National Park, Washington. Can. J. For. Res. 25:1049–1057.

Egli, M., L. Alioth, A. Mirabella, S. Raimondi, M. Nater, and R. Verel. 2007. Effect of climate and vegetation on soil organic carbon, humus fractions, allophanes, imogolite, kaolinite, and oxyhydroxides in volcanic soils of Etna (Sicily). Soil Sci. 172:673–691.

Ekschmitt, K., E. Kandeler, C. Poll, A. Brune, F. Buscot, M. Friedrich, G. Gleixner et al. 2008. Soil-carbon preservation through habitat constraints and biological limitations on decomposer activity. J. Plant Nutr. Soil Sci. 171:27–35.

Ellert, B.H., H.H. Janzen, A.J. VandenBygaart, and E. Bremer. 2008. Measuring change in soil organic carbon storage, p. 25–38. *In* M.R. Carter and E.G. Gregorich (eds.) Soil sampling and methods of analysis. 2nd edn. CRC Press, Taylor & Francis, Boca Raton, FL.

Elliott, E.T. 1986. Aggregate structure and carbon, nitrogen and phosphorus in native and cultivated soils. Soil Sci. Soc. Am. J. 50:627–633.

Elliott, E.T., and D.C. Coleman. 1988. Let the soil work for us. Ecol. Bull. 39:23–32.

Elliott, W.M., N.B. Elliott, and R.L. Wyman. 1993. Relative effect of litter and forest type on rate of decomposition. Am. Midl. Nat. 129:87–95.

Emerson, W.W., R.C. Foster, and J.M. Oades. 1986. Organo mineral complexes in relation to soil aggregation and structure, p. 521–548. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and microbes. Soil Science Society of America, Madison, WI.

Eswaran, H., E. Van den Berg, P. Reich, and J.M. Kimble. 1995. Global soil carbon resources, p. 27–43. *In* R. Lal et al. (eds.) Soils and global change. Lewis Publishers, Boca Raton, FL.

Feller, C., E. Fritsch, R. Poss, and C. Valentin. 1991. Effect of the texture on the storage and dynamics of organic matter in some low activity clay soils (West Africa, particularly). Cahiers ORSTOM Pedologie XXVI:25–36.

Filip, Z., and M. Tesarova. 2004. Microbial degradation and transformation of humic acids from permanent meadow and forest soils. Int. Biodeterior. Biodegrad. 54:225–231.

Flaig, W., H. Beutelspacher, and E. Rietz. 1975. Chemical composition and physical properties of humic substances, p. 1–211. *In* J.E. Gieseking (ed.) Soil components, Vol. I, Organic components. Springer-Verlag, New York.

Flavel, T.C., and D.V. Murphy. 2006. Carbon and nitrogen mineralization rates after application of organic amendments to soil. J. Environ. Qual. 35:183–193.

Flessa, H., W. Amelung, M. Helfrich, G.L.B. Wiesenberg, G. Gleixner, S. Brodowski, J. Rethemeyer, C. Kramer, and P.M. Grootes. 2008. Storage and stability of organic matter and fossil carbon in a luvisol and phaeozem with continuous maize cropping: A synthesis. J. Plant Nutr. Soil Sci. 171:36–51.

Fog, K. 1988. The effect of added nitrogen on the rate of decomposition of organic matter. Biol. Rev. Camb. Philos. Soc. 63:432–462.

Fortin, M.C. 1993. Soil temperature, soil water and no-till corn development following in-row residue removal. Agron. J. 85:571–576.

Foster, R.C. 1994. Microorganisms and soil aggregates, p. 144–155. *In* C.E. Pankhurst et al. (eds.) Soil biota: Management in sustainable farming systems. CSIRO Publishing, East Melbourne, Victoria, Australia.

Franzluebbers, A.J., F.M. Hons, and D.A. Zuberer. 1994. Seasonal changes in soil microbial biomass and mineralizable C and N in wheat management systems. Soil Biol. Biochem. 26:1469–1475.

Freibauer, A., M.D.A. Rounsevell, P. Smith, and J. Verhagen. 2004. Carbon sequestration in the agricultural soils of Europe. Geoderma 122:1–23.

Friedel, J.K., and E. Scheller. 2002. Composition of hydrolysable amino acids in soil organic matter and soil microbial biomass. Soil Biol. Biochem. 34:315–325.

Gaiffe, M., G. Duquet, H. Tavant, Y. Tavant, and S. Bruckert. 1984. Stabilité biologique et comportement physique d'un complet argilo-humique placé dans différentes condition de saturation en calcium ou en potassium. Plant Soil 77:271–284.

Garcia, C., T. Hemandez, F. Costa, and A. Polo. 1991. Humic substances in composted sewage sludge. Waste Manage. Res. 9:189–194.

Garcia-Pausas, J., P. Casals, L. Camarero, C. Huguet, M.T. Sebastià, R. Thompson, and J. Romanyà. 2007. Soil organic carbon storage in mountain grasslands of the Pyrenees: Effects of climate and topography. Biogeochemistry 82:279–289.

Gianfreda, L., and J.-M. Bollag. 1996. Influence of natural and anthropogenic factors on enzyme activity in soil, p. 123–193. *In* Stotzky, G., and J.-M. Boilag (eds.). Soil Biochemistry. Vol. 9 Marcel Dekker, New York.

Gillespie, A.W., F.L. Walley, R.E. Farrell, P. Leinweber, A. Schlichting, K.U. Eckhardt, T.Z. Regier, and R.I.R. Blyth. 2009. Profiling rhizosphere chemistry: Evidence from carbon and nitrogen K-edge XANES and pyrolysis-FIMS. Soil Sci. Soc. Am. J. 73:2002–2012.

Gleixner, G., C.J. Czimczik, C. Kramer, B. Luehker, and M.W.I. Schmidt. 2001. Plant compounds and their turnover and stabilization as soil organic matter, p. 201–215. *In* E.D. Schulze et al. (eds.) Global biogeochemical cycles in the climate system. Academic Press, San Diego, CA.

Golchin, A., J.A. Baldock, and J.M. Oades. 1997a. A model linking organic matter decomposition, chemistry and aggregate dynamics, p. 245–266. *In* R. Lal et al. (eds.) Soil processes and the carbon cycle. CRC Press, Boca Raton, FL.

Golchin, A., P. Clarke, J.A. Baldock, T. Higashi, J.O. Skjemstad, and J.M. Oades. 1997b. The effects of vegetation burning on the chemical composition of soil organic matter in a volcanic ash soil as shown by ^{13}C NMR spectroscopy. I. Whole soil and humic acid fraction. Geoderma 76:155–174.

Golchin, A., J.M. Oades, J.O. Skjemstad, and P. Clarke. 1994a. Study of free and occluded particulate organic matter in soils by solid state 13C Cp/MAS NMR spectroscopy and scanning electron microscopy. Aust. J. Soil Res. 32:285–309.

Golchin, A., J.M. Oades, J.O. Skjemstad, and P. Clarke. 1994b. Soil structure and carbon cycling. Aust. J. Soil Res. 32:1043–1068.

Goulden, M.L., S.C. Wofsy, J.W. Harden, S.E. Trumbore, P.M. Crill, S.T. Gower, T. Fries et al. 1998. Sensitivity of boreal forest carbon balance to soil thaw. Science 279:214–217.

Griffiths, B.S. 1994. Microbial feeding nematodes and protozoa in soil: Their effects on microbial activity and nitrogen mineralization in decomposition hotspots and the rhizosphere. Plant Soil 164:25–33.

Guggenberger, G., and K. Kaiser. 2003. Dissolved organic matter in soil: Challenging the paradigm of sorptive preservation. Geoderma 113:293–310.

Guggenberger, G., W. Zech, and R.J. Thomas. 1995. Lignin and carbohydrate alteration in particle-size separates of an oxisol under tropical pastures following native savanna. Soil Biol. Biochem. 27:1629–1638.

Guo, Y.Y., P. Gong, R. Amundson, and Q. Yu. 2006. Analysis of factors controlling soil carbon in the conterminous United States. Soil Sci. Soc. Am. J. 70:601–612.

Haaland, D.M., and E.V. Thomas. 1988. Partial least squares methods for spectral analyses 1: Relation to other quantitative calibration methods and the extraction of qualitative information. Anal. Chem. 60:1193–1202.

Haas, H.J., C.E. Evans, and E.F. Miles. 1957. Nitrogen and carbon changes in Great Plains soils as influenced by cropping and soil treatments. Technical Bulletin No. 1164. United States Department of Agriculture, Government Printing Office, Washington, DC. 111 pp.

Haider, K. 1992. Problems related to the humification processes in soils of temperate climates, p. 55–94. *In* G. Stotzky and J.M. Bollag (eds.) Soil biochemistry. Vol. 7. Marcel Dekker, New York.

Haider, K., and G. Guggenberger. 2005. Soil minerals and organic components: Impact on biological processes, human welfare, and nutrition, p. 3–16. Science Publishers, Inc., Enfield, NH.

Halstead, R.L., and R.B. McKercher. 1975. Biochemistry and cycling of phosphorus, p. 31–64. *In* E.A. Paul and A.D. McLaren (eds.) Soil biochemistry. Vol. 4. Marcel Dekker, New York.

Hamer, U., B. Marschner, S. Brodowski, and W. Amelung. 2004. Interactive priming of black carbon and glucose mineralisation. Org. Geochem. 35:823–830.

Han, B., X. Wang, and Z. Ouyang. 2005. Saturation levels and carbon sequestration potentials of soil carbon pools in farmland ecosystems of China. Rural Eco-Environ. 21:6–11.

Han, B., X.K. Wang, Z.Y. Ouyang, and F. Lu. 2006. Estimation of soil carbon saturation and carbon sequestration potential of an agro-ecosystem in China. Int. J. Sust. Dev. World Ecol. 13:459–468.

Hart, S.C., G.E. Nason, D.D. Myrold, and D.A. Perry. 1994. Dynamics of gross nitrogen transformations in an old-growth forest: The carbon connection. Ecology 75:880–891.

Harter, R.D., and R. Naidu. 1995. Role of metal-organic complexation in metal sorption by soils. Adv. Agron. 55:219–263.

Hassink, J. 1992. Effects of soil texture and structure on carbon and nitrogen mineralization in grassland soils. Biol. Fertil. Soils 14:126–134.

Hassink, J. 1997. The capacity of soils to preserve organic C and N by their association with clay and silt particles. Plant Soil 191:77–87.

Hassink, J., L.A. Bouwman, K.B. Zwart, and L. Brussard. 1993. Relationship between habitable pore space, soil biota and mineralization rates in grassland soils. Soil Biol. Biochem. 25:47–55.

Hassink, J., C. Chenu, J.W. Dalenberg, J. Bolem, and L.A. Bouwman, 1994. Interactions between soil biota, soil organic matter and soil structure, p. 57–58. 15th world congress of soil science, vol. 49, Acapulco, Mexico.

Haynes, R.J. 2005. Labile organic matter fractions as central components of the quality of agricultural soils: An overview, p. 221–268. *In* D.L. Sparks (ed.) Advances in agronomy. Vol. 85. Academic Press, San Diego, CA.

Heal, O.W., J.M. Anderson, and M.J. Swift. 1997. Plant litter quality and decomposition: An historical overview, p. 3–30. *In* G. Cadisch and K.E. Giller (eds.) Driven by nature: Plant litter quality and decomposition. CAB International, Wallingford, U.K.

Hecky, R.E., and P. Kilham. 1988. Nutrient limitation of phytoplankton in fresh-water and marine environments: A review of recent evidence on the effects of enrichment. Limnol. Oceanogr. 33:796–822.

Hedges, J.I., J.A. Baldock, Y. Gélinas, C. Lee, M.L. Peterson, and S.G. Wakeham. 2002. The biochemical and elemental compositions of marine plankton: A NMR perspective. Mar. Chem. 78:47–63.

Hedges, J.I., G.L. Cowie, J.R. Ertel, R.J. Barbour, and P.G. Hatcher. 1985. Degradation of carbohydrates and lignins in buried woods. Geochim. Cosmochim. Acta 49:701–711.

Hedges, J.I., and J.M. Oades. 1997. Comparative organic geochemistries of soils and marine sediments. Org. Geochem. 27:319–361.

Herbert, B.E., and P.M. Bertsch. 1995. Characterization of dissolved and colloidal organic matter in solutions: A review, p. 62–88. *In* W.W. McFee and J.M. Kelly (eds.) Carbon forms and functions in forest soils. Soil Science Society of America, Madison, WI.

Hicke, J.A., G.P. Asner, J.T. Randerson, C. Tucker, S. Los, R. Birdsey, J.C. Jenkins, C. Field, and E. Holland. 2002. Satellite-derived increases in net primary productivity across North America, 1982–1998. Geophys. Res. Lett. 29:69/1–69/4.

Hobbie, S.E. 1996. Temperature and plant species control over litter decomposition in Alaskan tundra. Ecol. Monogr. 66:503–522.

Holland, E.A., B.H. Braswell, J.F. Lamarque, A. Townsend, J. Sulzman, J.F. Muller, F. Dentener, G. Brasseur, H. Levy, J.E. Penner, and G.J. Roelofs. 1997. Variations in the predicted spatial distribution of atmospheric nitrogen deposition and their impact on carbon uptake by terrestrial ecosystems. J. Geophys. Res. Atmos. 102:15849–15866.

Holloway, P.J. 1982. The chemical constitution of plant cutin, p. 45–85. *In* D.F. Cutler et al. (eds.) The plant cuticle. Academic Press, London, U.K.

Homann, P.S., and D.F. Grigal. 1992. Molecular-weight distribution of soluble organics from laboratory-manipulated surface soils. Soil Sci. Soc. Am. J. 56:1305–1310.

Hope, D., M.F. Billett, and M.S. Cresser. 1994. A review of the export of carbon in river water: Fluxes and processes. Environ. Pollut. 84:301–324.

Houghton, R.A. 2005. The contemporary carbon cycle, p. 473–513. *In* W.H. Schlesinger (ed.) Biogeochemistry. Elsevier Science, Amsterdam, the Netherlands.

Huang, P.M. 2004. Soil mineral-organic matter-microorganism interactions: Fundamentals and impacts, p. 391–472. *In* D.L. Sparks (ed.) Advances in agronomy. Vol. 82. Academic Press, San Diego, CA.

Huang, P.M., and A. Violante. 1986. Influence of organic acids on crystallization and surface properties of precipitation products of aluminum, p. 160–221. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and microbes. Soil Science Society of America, Madison, WI.

Hutchinson, J.J., C.A. Campbell, and R.L. Desjardins. 2007. Some perspectives on carbon sequestration in agriculture. Agric. For. Meteorol. 142:288–302.

Ingram, J.S.I., and E.C.M. Fernandes. 2001. Managing carbon sequestration in soils: Concepts and terminology. Agric. Ecosyst. Environ. 87:111–117.

James, B.R., and S.J. Riha. 1986. pH buffering in forest soil organic horizons: Relevance to acid precipitation. J. Environ. Qual. 15:229–234.

Janik, L.J., R.H. Merry, and J.O. Skjemstad. 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? Aust. J. Exp. Agric. 38:681–696.

Janik, L.J., and J.O. Skjemstad. 1995. Characterization and analysis of soils using mid-infrared partial least-squares. 2. Correlations with some laboratory data. Aust. J. Soil Res. 33:637–650.

Janik, L.J., J.O. Skjemstad, K.D. Shepherd, and L.R. Spouncer. 2007. The prediction of soil carbon fractions using mid-infrared-partial least square analysis. Aust. J. Soil Res. 45:73–81.

Janssens, I.A., A. Freibauer, B. Schlamadinger, R. Ceulemans, P. Ciais, A.J. Dolman, M. Heimann, G.J. Nabuurs, P. Smith, R. Valentini, and E.D. Schulze. 2005. The carbon budget of terrestrial ecosystems at country-scale: A European case study. Biogeosciences 2:15–26.

Janzen, H.H. 1987. Soil organic matter characteristics after long term cropping to various spring wheat rotations. Can. J. Soil Sci. 67:845–856.

Jenkinson, D.S. 1976. Effects of biocidal treatments on metabolism in soil. 4. Decomposition of fumigated organisms in soil. Soil Biol. Biochem. 8:203–208.

Jenkinson, D.S. 1977. The soil microbial biomass. N. Z. Soil News 25:213–218.

Jenkinson, D.S., and A. Ayanaba. 1977. Decomposition of C-14 labeled plant material under tropical conditions. Soil Sci. Soc. Am. J. 41:912–915.

Jenkinson, D.S., P.C. Brookes, and D.S. Powlson. 2004. Measuring soil microbial biomass. Soil Biol. Biochem. 36:5–7.

Jenkinson, D.S., P.B.S. Hart, J.H. Rayner, and L.C. Parry. 1987. Modelling the turnover of organic matter in long-term experiments at Rothamsted. INTECOL Bull. 15:1–8.

Jenkinson, D.S., and J.N. Ladd. 1981. Microbial biomass in soil: Measurement and turnover. Soil Biochem. 5:415–471.

Jenkinson, D.S., and L.C. Parry. 1989. The nitrogen cycle in the Broadbalk wheat experiment: A model for the turnover of nitrogen through the soil microbial biomass. Soil Biol. Biochem. 21:535–541.

Jenkinson, D.S., and J.H. Rayner. 1977. Turnover of soil organic matter in some Rothamsted classical experiments. Soil Sci. 123:298–305.

Jenny, H. 1930. A study on the influence of climate upon the nitrogen and organic matter content of the soil. Missouri Agricultural Experimental station Bulletin 15, 66pp.

Jobbagy, E.G., and R.B. Jackson. 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol. Appl. 10:423–436.

John, B., T. Yamashita, B. Ludwig, and H. Flessa. 2005. Storage of organic carbon in aggregate and density fractions of silty soils under different types of land use. Geoderma 128:63–79.

Johnson, D.W. 1992. Effects of forest management on soil carbon storage. Water Air Soil Pollut. 64:83–120.

Johnson, D.W., and P.S. Curtis. 2001. Effects of forest management on soil C and N storage: Meta analysis. Forest Ecol. Manage. 140:227–238.

Johnson, D.W., J.D. Knoepp, W.T. Swank, J. Shan, L.A. Morris, D.H. Van Lear, and P.R. Kapeluck. 2002. Effects of forest management on soil carbon: Results of some long-term resampling studies. Environ. Pollut. 116:S201–S208.

Jokic, A., J.N. Cutler, D.W. Anderson, and F.L. Walley. 2004. Detection of heterocyclic N compounds in whole soils using N-XANES spectroscopy. Can. J. Soil Sci. 84:291–293.

Juste, C., and J. Delas. 1970. Comparison par une méthode repirométrique, des solubilités bioloiques d'un humate de calcium et d'un humate de sodium. C. R. Hebd. Seances Acad. Sci. D 270:1127–1129.

Juste, C., J. Delas, and M. Langon. 1975. Comparison de la stabilités biologique de différents humates metalliques. C. R. Hebd. Seances Acad. Sci. D 281:1685–1688.

Kahle, M., M. Kleber, M.S. Torn, and R. Jahn. 2003. Carbon storage in coarse and fine clay fractions of illitic soils. Soil Sci. Soc. Am. J. 67:1732–1739.

Kajak, A. 1995. The role of soil predators in decomposition processes. Eur. J. Entomol. 92:573–580.

Kalbitz, K., and K. Kaiser. 2008. Contribution of dissolved organic matter to carbon storage in forest mineral soils. J. Plant Nutr. Soil Sci. 171:52–60.

Kalbitz, K., J. Schmerwitz, D. Schwesig, and E. Matzner. 2003. Biodegradation of soil-derived dissolved organic matter as related to its properties. Geoderma 113:273–291.

Kalbitz, K., S. Solinger, J.H. Park, B. Michalzik, and E. Matzner. 2000. Controls on the dynamics of dissolved organic matter in soils: A review. Soil Sci. 165:277–304.

Kay, B.D., A.P. da Silva, and J.A. Baldock. 1997. Sensitivity of soil structure to changes in organic carbon content: Predictions using pedotransfer functions. Can. J. Soil Sci. 77:655–667.

Keil, R.G., E. Tsamakis, C.B. Fuh, C. Giddings, and J.I. Hedges. 1994. Mineralogical and textural controls on the organic composition of coastal marine sediments: Hydrodynamic separation using SPLITT-fractionation. Geochim. Cosmochim. Acta 58:879–893.

Keleman, S.R., M. Afeworki, M.L. Gorbaty, P.J. Kwiatek, M.S. Solum, J.Z. Hu, and R.J. Pugmire. 2002. XPS and ^{15}N NMR study of nitrogen forms in carbonaceous solids. Energy Fuels 16:1507–1515.

Kilbertus, G. 1980. Study of microhabitats in soil aggregates. Relation to bacterial biomass and size of procaryotes. Revue D Ecologie Et De Biologie Du Sol 17:543–557.

Killham, K., M. Amato, and J.N. Ladd. 1993. Effects of substrate location in soil and soil pore-water regime on carbon turnover. Soil Biol. Biochem. 25:57–62.

Kimble, J.M., C.W. Rice, D. Reed, S. Mooney, R.F. Follett, and R. Lal. 2007. Soil carbon management: Economic, environmental and societal benefits, CRC Press, Taylor & Francis Group, Boca Raton, FL, 280pp.

Kinyangi, J., D. Solomon, B.I. Liang, M. Lerotic, S. Wirick, and J. Lehmann. 2006. Nanoscale biogeocomplexity of the organomineral assemblage in soil: Application of STXM microscopy and C 1s-NEXAFS spectroscopy. Soil Sci. Soc. Am. J. 70:1708–1718.

Kirschbaum, M.U.F. 1995. The temperature dependence of soil organic matter decomposition and the effect of global warming on soil organic C storage. Soil Biol. Biochem. 27:753–760.

Knicker, H., G. Almendros, F.J. González-vila, H.-D. Lüdemann, and F. Martin. 1995. ^{13}C and ^{15}N NMR analysis of some fungal melanins in comparison with soil organic matter. Org. Geochem. 23:1023–1028.

Knicker, H., R. Fründ, and H.-D. Lüdeman. 1993. The chemical nature of nitrogen in native soil organic matter. Naturwissenschaften 80:219–221.

Knicker, H., and H.-D. Lüdemann. 1995. N-15 and C-13 CPMAS and solution NMR studies of N-15 enriched plant material during 600 days of microbial degradation. Org. Geochem. 23:329–341.

Knicker, H., and J.O. Skjemstad. 2000. Nature of organic carbon and nitrogen in physically protected organic matter of some Australian soils as revealed by solid-state ^{13}C and ^{15}N NMR spectroscopy. Aust. J. Soil Res. 38:113–127.

Kögel-Knabner, I. 2002. The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. Soil Biol. Biochem. 34:139–162.

Kögel-Knabner, I., K. Ekschmitt, H. Flessa, G. Guggenberger, E. Matzner, B. Marschner, and M. von Luetzow. 2008a. An integrative approach of organic matter stabilization in temperate soils: Linking chemistry, physics, and biology. J. Plant Nutr. Soil Sci. 171:5–13.

Kögel-Knabner, I., G. Guggenberger, M. Kleber, E. Kandeler, K. Kalbitz, S. Scheu, K. Eusterhues, and P. Leinweber. 2008b. Organo-mineral associations in temperate soils: Integrating biology, mineralogy, and organic matter chemistry. J. Plant Nutr. Soil Sci. 171:61–82.

Kögel-Knabner, I., M.V. Lützow, G. Guggenberger, H. Flessa, B. Marschner, E. Matzner, and K. Ekschmitt. 2005. Mechanisms and regulation of organic matter stabilisation in soils. Geoderma 128:1–2.

Krull, E.S., J.A. Baldock, and J.O. Skjemstad. 2003. Importance of mechanisms and processes of the stabilisation of soil organic matter for modelling carbon turnover. Funct. Plant Biol. 30:207–222.

Krull, E.S., J.O. Skjemstad, and J.A. Baldock. 2005. Functions of soil organic matter and effects on soil properties, p. 107. Cooperative Research Centre for Greenhouse Accounting, Canberra, Australia.

Krull, E.S., C.W. Swanston, J.O. Skjemstad, and J.A. McGowan. 2006. Importance of charcoal in determining the age and chemistry of organic carbon in surface soils. J. Geophys. Res. Biogeosci. 111, 6.04001, doi: 10.1029/2006J6.0001

Kruse, J., and P. Leinweber. 2008. Phosphorus in sequentially extracted fen peat soils: A K-edge X-ray absorption near-edge structure (XANES) spectroscopy study. J. Plant Nutr. Soil Sci. 171:613–620.

Kruse, J., P. Leinweber, K.U. Eckhardt, F. Godlinski, Y.F. Hu, and L. Zuin. 2009. Phosphorus L-2,L-3-edge XANES: Overview of reference compounds. J. Synchrotron Radiat. 16:247–259.

Ladd, J.N., M. Amato, and J.M. Oades. 1985. Decomposition of plant materials in Australian soils. III Residual organic and microbial biomass C and N from isotope-labelled legume materials and soil organic matter decomposing under field conditions. Aust. J. Soil Res. 23:603–611.

Ladd, J.N., and J.H.A. Butler. 1975. Humus-enzyme systems and synthetic, organic polymer-enzyme analogs, p. 143–194. *In* E.A. Paul and A.D. McLaren (eds.) Soil biochemistry. Vol. 4. Marcel Dekker, Inc., New York.

Ladd, J.N., R.C. Foster, P. Nannipieri, and J.M. Oades. 1996. Soil structure and biological activity, p. 23–78. *In* G. Stotzky and J.M. Bollag (eds.) Soil biochemistry. Vol. 9. Marcel Dekker, Inc., New York.

Ladd, J.N., J.M. Oades, and M. Amato. 1981. Microbial biomass formed from ^{14}C, ^{15}N-labelled plant material decomposing in soils in the field. Soil Biol. Biochem. 13:119–126.

Lal, R. 2004a. Soil carbon sequestration to mitigate climate change. Geoderma 123:1–22.

Lal, R. 2004b. Agricultural activities and the global carbon cycle. Nutr. Cycl. Agroecosyst. 70:103–116.

Lal, R., K. Kimble, R. Follet, and C. Cole (eds.). 1998. The potential of US cropland to sequester carbon and mitigate the greenhouse effect. Ann Arbor Press, Chelsea, MI.

Lavelle, P., E. Blanchart, A. Martin, S. Martin, A. Spain, F. Toutain, I. Barois, and R. Schaefer. 1993. A hierarchical model for decomposition in terrestrial ecosystems: Application to soils of the humid tropics. Biotropica 25:130–150.

Lavelle, P., and A. Martin. 1992. Small-scale and large-scale effects of endogenic earthworms on soil organic matter dynamics in soils of the humid tropics. Soil Biol. Biochem. 24:1491–1498.

Lawes, J. 1861. On the application of different manures to different crops and their proper distribution on the farm. Private publication cited Dyke (1993) (as cited by Heal et al. 1997).

Leavitt, S.W., R.F. Follett, and E.A. Paul. 1996. Estimation of slow- and fast-cycling soil organic carbon pools from 6 N HCl hydrolysis. Radiocarbon 38:231–239.

Lefroy, R.D.B., G.J. Blair, and E.T. Craswell. 1995. Soil organic matter management for sustainable agriculture: A workshop held in Ubon, Thailand, August 24–26, 1994.

Lefroy, R.D.B., G.J. Blair, and W.M. Strong. 1993. Changes in soil organic matter with cropping as measured by organic carbon fractions and ^{13}C natural isotope abundance. Plant Soil 155–156:399–402.

Lehmann, J., B.Q. Liang, D. Solomon, M. Lerotic, F. Luizao, J. Kinyangi, T. Schafer, S. Wirick, and C. Jacobsen. 2005. Near-edge X-ray absorption fine structure (NEXAFS) spectroscopy for mapping nano-scale distribution of organic carbon forms in soil: Application to black carbon particles. Global Biogeochem. Cy. 19:GB1013.

Lehmann, J., D. Solomon, J. Kinyangi, L. Dathe, S. Wirick, and C. Jacobsen. 2008. Spatial complexity of soil organic matter forms at nanometre scales. Nat. Geosci. 1:238–242.

Leifeld, J., and I. Kögel-Knabner. 2005. Soil organic matter fractions as early indicators for carbon stock changes under different land-use? Geoderma 124:143–155.

Leinweber, P., J. Kruse, F.L. Walley, A. Gillespie, K.U. Eckhardt, R.I.R. Blyth, and T. Regier. 2007. Nitrogen K-edge XANES—An overview of reference compounds used to identify 'unknown' organic nitrogen in environmental samples. J. Synchrotron Radiat. 14:500–511.

Liang, B., J. Lehmann, D. Solomon, J. Kinyangi, J. Grossman, B. O'Neill, J.O. Skjemstad, J. Thies, F.J. Luizao, J. Petersen, and E.G. Neves. 2006. Black carbon increases cation exchange capacity in soils. Soil Sci. Soc. Am. J. 70:1719–1730.

Liang, B., J. Lehmann, D. Solomon, S. Sohi, J.E. Thies, J.O. Skjemstad, F.J. Luizao, M.H. Engelhard, E.G. Neves, and S. Wirick. 2008. Stability of biomass-derived black carbon in soils. Geochim. Cosmochim. Acta 72:6069–6078.

Lieth, H. 1973. Primary production: Terrestrial ecosystems. Hum. Ecol. 1:303–332.

Ligel, L.H. 1992. An overview of carbon sequestration in soils of Latin America. *In* F.H. Beinroth (ed.) Organic carbon sequestration in the soils of Puerto Rico. Agronomy and Soils. University of Puerto Rico, Mayaguez, Puerto Rico.

Linhares, M. 1977. Contribution de l'ion calcium à la stabilisation biologique de la matière organique des sols. Thise Doc. Spéc., Université de Bordeaux III.

Ljungdahl, L.G., and K.-E. Eriksson. 1985. Ecology of microbial cellulose degradation, p. 237–299. *In* K.C. Marshall (ed.) Advances in microbial ecology. Vol. 8. Plenum press, New York.

Lorenz, K., and R. Lal. 2005. The depth distribution of soil organic carbon in relation to land use and management and the potential of carbon sequestration in subsoil horizons, p. 35–66. *In* D.L. Sparks (ed.) Advances in agronomy. Vol. 88. Academic Press, San Diego, CA.

Ludwig, B., K. Kuka, U. Franko, and M. von Luetzow. 2008. Comparison of two quantitative soil organic carbon models with a conceptual model using data from an agricultural long-term experiment. J. Plant Nutr. Soil Sci. 171:83–90.

MacCarthy, P., R.L. Malcolm, C.E. Clapp, and P.R. Bloom. 1990. An introduction to soil humic substances, p. 1–12. *In* P. MacCarthy et al. (eds.) Humic substances in crop and soil sciences: Selected readings. Soil Science Society of America, Madison, WI.

Magid, J., T. Mueller, L.S. Jansen, and N.E. Nielsen. 1996. Modelling the measurable: Interpretation of field-scale CO_2 and N-mineralisation, soil microbial biomass and light fractions as indicators of oilseed rape, maize and barley straw decomposition, p. 349–362. *In* G. Cadisch and K.E. Giller (eds.) Driven by nature: Plant litter quality and decomposition. CAB International, Wallingford, U.K.

Mahieu, N., D.C. Olk, and E.W. Randall. 2000. Accumulation of heterocyclic nitrogen in humified organic matter: A ^{15}N-NMR study of lowland rice soils. Eur. J. Soil Sci. 51:379–389.

Mainari, S., and F. Caporali. 2008. Soil carbon sequestration under organic farming in the Mediterranean environment. Transworld. Research Network, Kerala, India. 178 pp.

Mann, L.K. 1986. Changes in soil carbon storage after cultivation. Soil Sci. 142:289–288.

Marschner, B., S. Brodowski, A. Dreves, G. Gleixner, A. Gude, P.M. Grootes, U. Hamer et al. 2008. How relevant is recalcitrance for the stabilization of organic matter in soils? J. Plant Nutr. Soil Sci. 171:91–110.

Martens, D.A., and K.L. Loeffelmann. 2002. Improved accounting of carbohydrate carbon from plants and soils. Soil Biol. Biochem. 34:1393–1399.

Martens, D.A., and K.L. Loeffelmann. 2003. Soil amino acid composition quantified by acid hydrolysis and anion chromatography-pulsed amperometry. J. Agric. Food Chem. 51:6521–6529.

Martin, J.P., J.O. Ervin, and R.A. Shepherd. 1966. Decomposition of the iron, aluminium, zinc, and copper salts or complexes of some microbial and plant polysaccharides in soil. Proc. Soil Sci. Soc. Am. 30:196–200.

Martin, J.P., and K. Haider. 1975. Decomposition of specifically labelled benzoic acid and cinnamic acid derivatives in soil. Proc. Soil Sci. Soc. Am. 39:657–662.

Martin, J.P., S.J. Richards, and J.O. Ervin. 1972. Decomposition and binding action in soil of some mannose containing microbial polysaccharides and their Fe, Al, Zn and Cu complexes. Soil Sci. 113:322–327.

Martínez, A.T., A.E. González-Vila, M. Valmaseda, B.E. Dale, M.J. Lambregts, and J.F. Haw. 1991. Solid-state NMR studies of lignin and plant polysaccharide degradation by fungi. Holzforschung 45:49–54.

Masiello, C.A., M.E. Gallagher, J.T. Randerson, R.M. Deco, and O.A. Chadwick. 2008. Evaluating two experimental approaches for measuring ecosystem carbon oxidation state and oxidative ratio. J. Geophys. Res. Biogeosci. 113:9.

Mayer, L.M. 1994a. Surface area control of organic carbon accumulation in continental shelf sediments. Geochim. Cosmochim. Acta 58:1271–1284.

Mayer, L.M. 1994b. Relationships between mineral surfaces and organic carbon concentrations in soils and sediments. Chem. Geol. 114:347–363.

McBride, M.B. 1994. Environmental chemistry of soils. Oxford University Press, New York.

McCown, R.L., G.L. Hammer, J.N.G. Hargreaves, D.P. Holzworth, and D.M. Freebairn. 1996. APSIM: A novel software system for model development, model testing and simulation in agricultural systems research. Agric. Syst. 50:255–271.

McDougall, G.J., I.M. Morrison, D. Stewart, J.D.B. Weyers, and J.R. Hillman. 1993. Plant fibers: Botany, chemistry and processing for industrial use. J. Sci. Food Agric. 62:1–20.

McFee, W.W., and J.M. Kelly. 1995. Carbon forms and functions in forest soils. Soil Science Society of America, Madison, WI. 594 pp.

McGill, W.B., and C.V. Cole. 1981. Comparative aspects of cycling of organic C, N, S and P through soil organic matter. Geoderma 26:267–286.

McKeague, J.A., M.V. Cheshire, F. Andreux, and J. Berthelin. 1986. Organo-mineral complexes in relation to pedogenesis, p. 549–592. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and microbes. Soil Science Society of America, Madison, WI.

Mendham, D.S., A.M. O'Connell, and T.S. Grove. 2002. Organic matter characteristics under native forest, long-term pasture, and recent conversion to eucalyptus plantations in Western Australia: Microbial biomass, soil respiration, and permanganate oxidation. Aust. J. Soil Res. 40:859–872.

Metting, F., J. Smith, and J. Amthor. 1999. Science needs and new technology for carbon sequestration, p. 1–34. *In* N. Rosenberg et al. (eds.) Carbon sequestration in soils: Science, monitoring and beyond. Battelle Press, Columbus, OH.

Michalzik, B., K. Kalbitz, J.-H. Park, S. Solinger, and E. Matzner. 2001. Fluxes and concentrations of dissolved organic carbon and nitrogen: A synthesis for temperate forests. Biogeochemistry 52:173–205.

Moorhead, D.L., and J.F. Reynolds. 1989. The contribution of abiotic processes to buried litter decomposition in the northern Chihuahuan desert. Oecologia 79:133–135.

Muller, P.E. (ed.) 1887. Studien über die natürlichen Humusformen und deren Einwirkunger auf Vegetation und Boden. Springer, Berlin, Germany.

Müller-Wegener, U. 1988. Interactions of humic substances with biota, p. 179–192. *In* F.H. Frimmel and R.F. Christman (eds.) Humic substances and their role in the environment. John Wiley & Sons, New York.

Muneer, M., and J.M. Oades. 1989a. The role of Ca-organic interactions in soil aggregate stability. I. Laboratory studies with ^{14}C-glucose, $CaCO_3$ and $CaSO_4 \cdot 2H_2O$. Aust. J. Soil Res. 27:389–399.

Muneer, M., and J.M. Oades. 1989b. The role of Ca-organic interactions in soil aggregate stability. II. Field studies with ^{14}C-labelled straw, $CaCO_3$ and $CaSO_4 \cdot 2H_2O$. Aust. J. Soil Res. 27:401–409.

Muneer, M., and J.M. Oades. 1989c. The role of Ca-organic interactions in soil aggregate stability. III. Mechanisms and models. Aust. J. Soil Res. 27:411–423.

Murphy, D.V., S. Recous, E.A. Stockdale, I.R.P. Fillery, L.S. Jensen, D.J. Hatch, and K.W.T. Goulding. 2003. Gross nitrogen fluxes in soil: Theory, measurement and application of 15N pool dilution techniques, p. 69–118. *In* D.L. Sparks (ed.) Advances in agronomy. Vol. 79. Academic Press, San Diego, CA.

Muyzer, G., E.C. Dewaal, and A.G. Uitterlinden. 1993. Profiling of complex microbial populations by denaturing gradient gel-electrophoresis analysis of polymerase chain reaction amplified genes coding for 16s ribosomal RNA. Appl. Environ. Microbiol. 59:695–700.

Nardi, S., G. Concheri, and G. Dell'Agnola. 1996. Biological activity of humic substances, p. 361–406. *In* A. Piccolo (ed.) Humic substances in terrestrial ecosystems. Elsevier, Amsterdam, the Netherlands.

Nardi, S., D. Pizzeghello, A. Muscolo, and A. Vianello. 2002. Physiological effects of humic substances on higher plants. Soil Biol. Biochem. 34:1527–1536.

Neff, J.C., and G.P. Asner. 2001. Dissolved organic carbon in terrestrial ecosystems: Synthesis and a model. Ecosystems 4:0029–0048.

Nelson, P.N., and J.A. Baldock. 2005. Estimating the molecular composition of a diverse range of natural organic materials from solid-state C-13 NMR and elemental analyses. Biogeochemistry 72:1–34.

Nelson, P.N., J.A. Baldock, and J.M. Oades. 1993. Concentration and composition of dissolved organic carbon in streams in relation to catchment soil properties. Biogeochemistry 19:27–50.

Nelson, P.N., and J.M. Oades. 1997. Organic matter, sodicity and soil structure, p. 67. *In* M.E. Sumner and R. Naidu (eds.) Sodic soils: Distribution, processes, management and environmental consequences. Oxford University Press, New York.

Nelson, D.W., and L.E. Sommers. 1996. Total carbon, organic carbon, and organic matter, p. 961–1010. *In* D.L. Sparks et al. (eds.) Methods of soil analysis. Part 3. Chemical methods. ASA-CSSA-SSSA, Madison, WI.

Nguyen, M.L., and K.M. Goh. 1994. Sulfur cycling and its implications on sulfur fertiliser requirements of grazed grassland ecosystems. Agric. Ecosyst. Environ. 49:173–206.

Nip, M., J.W. Deleeuw, P.J. Holloway, J.P.T. Jensen, J.C.M. Sprenkels, M. Depooter, and J.J.M. Sleeckx. 1987. Comparison of flash pyrolysis, differential scanning calorimetry, C-13 NMR and IR spectroscopy in the analysis of a highly aliphatic biopolymer from plant cuticles. J. Anal. Appl. Pyrol. 11:287–295.

Nip, M., E.W. Tegelaar, H. Brinkhuis, J.W. Deleeuw, P.A. Schenck, and P.J. Holloway. 1986a. Analysis of modern and fossil plant cuticles by curie-point PY-GC and curie-point PY-GC-MS. Recognition of a new, highly aliphatic and resistant biopolymer. Org. Geochem. 10:769–778.

Nip, M., E.W. Tegelaar, J.W. Deleeuw, P.A. Schenck, and P.J. Holloway. 1986b. A new non-saponifiable highly aliphatic and resistant biopolymer in plant cuticles: Evidence from pyrolysis and C-13-NMR analysis of present day and fossil plants. Naturwissenschaften 73:579–585.

Oades, J.M. 1984. Soil organic matter and structural stability: Mechanisms and implications for management. Plant Soil 76:319–337.

Oades, J.M. 1988. The retention of organic matter in soil. Biogeochemistry 5:35–70.

Oades, J.M. 1989. An introduction to organic matter in mineral soils, p. 89–159. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. Soil Science Society of America, Madison, WI.

Oades, J.M. 1993. The role of biology in the formation, stabilization and degradation of soil structure. Geoderma 56:377–400.

Oades, J.M. 1995. Krasnozems: Organic matter. Aust. J. Soil Res. 33:43–57.

Oades, J.M., and A.G. Waters. 1991. Aggregate hierarchy in soils. Aust. J. Soil Res. 29:815–828.

Oorts, K., B. Vanlauwe, and R. Merckx. 2003. Cation exchange capacities of soil organic matter fractions in a ferric lixisol with different organic matter inputs. Agric. Ecosyst. Environ. 100:161–171.

Orlov, D.S. 1995. Humic substances of soils and general theory of humification. Brookfield, Rotterdam, the Netherlands. 323 pp.

Parton, W.J., D.C. Schimel, C.V. Cole, and D.S. Ojima. 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. Soil Sci. Soc. Am. J. 51:1173–1179.

Parton, W.J., J.W.B. Stewart, and C.V. Cole. 1988. Dynamics of C, N, P and S in grassland soils: A model. Biogeochemistry 5:109–131.

Paul, E.A., H.P. Collins, and S.W. Leavitt. 2001. Dynamics of resistant soil carbon of midwestern agricultural soils measured by naturally occurring ^{14}C abundance. Geoderma 104:239–256.

Paul, E.A., and H. van Veen. 1978. The use of tracers to determine the dynamic nature of organic matter, p. 61–102. Trans. 11th Int. Congr. Soil Sci. Vol. 3. International Society of Soil Science, Edmonton, Alberta, Canada.

Paustian, K. 1994. Modelling soil biology and biochemical processes for sustainable agriculture research, p. 182–193. *In* C.E. Pankhurst et al. (eds.) Soil biota: Management in sustainable farming systems. CSIRO Publishing, Melbourne, Australia.

Paustian, K., O. Andren, H.H. Janzen, R. Lal, P. Smith, G. Tian, H. Tiessen, M. Van Noordwijk, and P.L. Woomer. 1997a. Agricultural soils as a sink to mitigate CO_2 emissions. Soil Use Manage. 13:230–244.

Paustian, K., H.P. Collins, and E.A. Paul. 1997b. Management controls on soil carbon, p. 1549. *In* E.A. Paul et al. (eds.) Soil organic matter in temperate agroecosystems: Longterm experiments in North America. CRC Press, Boca Raton, FL.

Peterson, B.J., and J.M. Melillo. 1985. The potential storage of carbon caused by eutrophication of the biosphere. Tellus Ser.B Chem. Phys. Meteorol. 37:117–127.

Peterson, G.A., D.G. Westfall, C.W. Wood, and S. Ross. 1988. Crop and soil management in dryland agroecosystems. Colorado State University Technical Bulletin LTB886.

Piccolo, A. 2002. The supramolecular structure of humic substances: A novel understanding of humus chemistry and implications in soil science, p. 57–134. *In* D.L. Sparks (ed.) Advances in agronomy. Vol. 75. Academic Press, San Diego, CA.

Pocknee, S., and M.E. Sumner. 1997. Cation and nitrogen contents of organic matter determine its soil liming potential. Soil Sci. Soc. Am. J. 61:86–92.

Poirier, N., S.P. Sohi, J.L. Gaunt, N. Mahieu, E.W. Randall, D.S. Powlson, and R.P. Evershed. 2005. The chemical composition of measurable soil organic matter pools. Org. Geochem. 36:1174–1189.

Post, W.M., W.R. Emmanuel, P.J. Zinke, and A.G. Stangenberger. 1982. Soil carbon pools and world life zones. Nature 298:156–159.

Post, W., R. Izaurralde, L. Mann, and N. Bliss. 1999. Monitoring and verifying soil organic carbon sequestration, p. 41–66. *In* N. Rosenberg et al. (eds.) Carbon sequestration in soils: Science, monitoring and beyond. Battelle Press, Columbus, OH.

Post, W.M., R.C. Izaurralde, L.K. Mann, and N. Bliss. 2001. Monitoring and verifying changes of organic carbon in soil. Clim. Change 51:73–99.

Post, W.M., T.-H. Peng, W.R. Emanuel, A.W. King, V.H. Dale, and D.L. DeAngelis. 1990. The global carbon cycle. Am. Sci. 78:310–326.

Poulenard, J., J.C. Michel, F. Bartoli, J.M. Portal, and P. Podwojewski. 2004. Water repellency of volcanic ash soils from Ecuadorian páramo: Effect of water content and characteristics of hydrophobic organic matter. Eur. J. Soil Sci. 55:487–496.

Powlson, D.S., and D.S. Jenkinson. 1976. Effects of biocidal treatments on metabolism in soil. 2. Gamma-irradiation, autoclaving, air-drying and fumigation. Soil Biol. Biochem. 8:179–188.

Pressenda, L.C.R., R. Aravena, A.J. Melfi, E.C.C. Telles, R. Boulet, E.P.E. Vanencia, and M. Tomazello. 1996. The use of carbon isotopes (C13, C14) in soil to evaluate vegetation changes during the Holocene in central Brazil. Radiocarbon 38:191–201.

Preston, C.M. 1996. Applications of NMR to soil organic matter analysis: History and prospects. Soil Sci. 161:144–166.

Preston, C.M., J.A. Trofymow, B.G. Sayer, and J. Niu. 1997. ^{13}C nuclear magnetic resonance spectroscopy with cross-polarization and magic-angle spinning investigation of the proximate-analysis fractions used to assess litter quality in decomposition studies. Can. J. Bot. 75:1601–1613.

Prietzel, J., J. Thieme, U. Neuhäusler, J. Susini, and I. Kögel-Knabner. 2003. Speciation of sulphur in soils and soil particles by X-ray spectromicroscopy. Eur. J. Soil Sci. 54:423–433.

Prietzel, J., J. Thieme, M. Salome, and H. Knicker. 2007. Sulfur K-edge XANES spectroscopy reveals differences in sulfur speciation of bulk soils, humic acid, fulvic acid, and particle size separates. Soil Biol. Biochem. 39:877–890.

Raison, R.J., M.J. Connell, and P.K. Khanna. 1987. Methodology for studying fluxes of soil mineral N *in situ*. Soil Biol. Biochem. 19:521–530.

Randerson, J.T., C.A. Masiello, C.J. Still, T. Rahn, H. Poorter, and C.B. Field. 2006. Is carbon within the global terrestrial biosphere becoming more oxidized? Implications for trends in atmospheric O_2. Global Change Biol. 12:260–271.

Ransom, B., D. Kim, M. Kastner, and S. Wainwright. 1998. Organic matter preservation on continental slopes: Importance of mineralogy and surface area. Geochim. Cosmochim. Acta 62:1329–1345.

Rashidi, M., and M. Seilsepour. 2008. Modeling of soil cation exchange capacity based on some soil physical and chemical properties. J. Agric. Biol. Sci. 3:6–13.

Redfield, A. 1958. The biological control of chemical factors in the environment. Am. Sci. 46:205–221.

Reichstein, M. 2007. Impacts of climate change on forest soil carbon: Principles, factors, models, uncertainties, p. 127–135. Tree-Smith, P.H. et al. (eds.) CAB International, Wallingford, U.K.

Reiners, W.A. 1983. Transport processes in the biogeochemical cycles of carbon, nitrogen, phosphorus, and sulfur, p. 143–176. *In* B. Bolin and R.B. Cook (eds.) The major biogeochemical cycles and their interactions. Scope 21. John Wiley & Sons, New York.

Rethemeyer, J., C. Kramer, G. Gleixner, B. John, T. Yamashita, H. Flessa, N. Andersen, M.-J. Nadeau, and P.M. Grootes. 2005. Transformation of organic matter in agricultural soils: Radiocarbon concentration versus soil depth. Geoderma 128:94–105.

Ridley, A.M., W.J. Slattery, K.R. Helyar, and A. Cowling. 1990. The importance of the carbon cycle to acidification of a grazed annual pasture. Aust. J. Exp. Agric. 30:529–537.

Robert, M., and J. Berthelin. 1986. Role of biological and biochemical factors in soil mineral weathering, p. 455–495. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and microbes. Soil Science Society of America, Madison, WI.

Roig, A., A. Lax, J. Cegarra, F. Costa, and M.T. Hernandez. 1988. Cation exchange capacity as a parameter for measuring the humification degree of manures. Soil Sci. 146:311–316.

Romanenkov, V.A., J.U. Smith, P. Smith, O.D. Sirotenko, D.I. Rukhovitch, and I.A. Romanenko. 2007. Soil organic carbon dynamics of croplands in European Russia: Estimates from the "model of humus balance". Reg. Environ. Change 7:93–104.

Rovira, P., and V.R. Vallejo. 2000. Examination of thermal and acid hydrolysis procedures in characterization of soil organic matter. Commun. Soil Sci. Plant Anal. 31:81–100.

Rumpel, C., A. Seraphin, M.-O. Goebel, G. Wiesenberg, F. Gonzales-Vila, J. Bachmann, L. Schwark, W. Michaelis, A. Mariotti, and I. Kögel-Knabner. 2004. Alkyl C and hydrophobicity in B and C horizons of an acid forest soil. J. Plant Nutr. Soil Sci. 167:685–692.

Russel, J.S. 1960. Soil fertility changes in the long-term experimental plots at Kybybolite, South Australia. I. Changes in pH, total nitrogen, organic carbon and bulk density. Aust. J. Agric. Sci. 11:902–926.

Saggar, S., J.R. Bettany, and J.W.B. Stewart. 1981. Measurement of microbial sulfur in soil. Soil Biol. Biochem. 13:493–498.

Saggar, S., A. Parshotam, G.P. Sparling, C.W. Feltham, and P.B.S. Hart. 1996. ^{14}C-labelled ryegrass turnover and residence times in soils varying in clay content and mineralogy. Soil Biol. Biochem. 28:1677–1686.

Sala, O.E., W.J. Parton, L.A. Joyce, and W.K. Lauenroth. 1988. Primary production of the central grassland region of the United States. Ecology 69:40–45.

Saleska, S.R., S.D. Miller, D.M. Matross, M.L. Goulden, S.C. Wofsy, H.R. da Rocha, P.B. de Camargo et al. 2003. Carbon in Amazon forests: Unexpected seasonal fluxes and disturbance-induced losses. Science 302:1554–1557.

Sanderman, J., J.A. Baldock, and R. Amundson. 2008. Dissolved organic carbon chemistry and dynamics in contrasting forest and grassland soils. Biogeochemistry 89:181–198.

Sanyal, S.K., and S.K.D. Datta. 1991. Chemistry of phosphorus transformations in soil. Adv. Soil Sci. 16:1–20.

Sarkar, J.M., and J.M. Bollag. 1987. Inhibitory effect of humic and fulvic acids on oxidoreductases as measured by the coupling of 2,4-diclorophenol to humic substances. Sci. Total Environ. 62:367–377.

Scharpenseel, H.W., H.U. Neue, and S. Singer. 1992. Biotransformations in different climatic belts: Source sink relationships, p. 91–105. *In* J. Kubat (ed.) Humus, its structure and role in agriculture and environment. Elsevier Science Publishers, Amsterdam, the Netherlands.

Scheel, T., C. Dörfler, and K. Kalbitz. 2007. Precipitation of dissolved organic matter by aluminum stabilizes carbon in acidic forest soils. Soil Sci. Soc. Am. J. 71:64–74.

Schimel, D.S. 1988. Calculation of microbial-growth efficiency from N-15 immobilization. Biogeochemistry 6:239–243.

Schimel, D.S., D. Alves, I. Enting, M. Heimann, F. Joos, D. Raynaud, and T. Wigley. 1996. CO_2 and the carbon cycle, p. 76–86. *In* J.T. Houghton, L.G. Meira Filho, B.A. Callendar, N. Harris, A. Kattenberg, and K. Maskel (eds.) Climate change 1995. Cambridge University Press, Cambridge, U.K.

Schimel, D.S., D.C. Coleman, and K.A. Horton. 1985a. Soil organic matter dynamics in paired rangeland and cropland toposequences in North Dakota. Geoderma 36:201–214.

Schimel, D.S., M.A. Stillwell, and R.G. Woodmansee. 1985b. Biogeochemistry of C, N, and P in a soil catena of the short grass steppe. Ecology 66:276–282.

Schmidt, M.W.I., and G. Gleixner. 2005. Carbon and nitrogen isotope composition of bulk soils, particle-size fractions and organic material after treatment with hydrofluoric acid. Eur. J. Soil Sci. 56:407–416.

Schmidt, M.W.I., J.O. Skjemstad, E. Gehrt, and I. Kögel-Knaber. 1999. Charred organic carbon in German chernozemic soils. Eur. J. Soil Sci. 50:351–365.

Schnitzer, M. 1985. Nature of nitrogen in humic substances, p. 303–325. *In* G.R. Aiken et al. (eds.) Humic substances in soil, sediment and water: Geochemistry, isolation and characterization. John Wiley & Sons, New York.

Schnitzer, M. 2000. A lifetime perspective on the chemistry of soil organic matter. Adv. Agron. 68:1–58.

Schnitzer, M., and H.R. Schulten. 1995. Analysis of organic matter in soil extracts and whole soils by pyrolysis-mass spectrometry. Adv. Agron. 55:167–198, C1–C2, 199–217.

Schnitzer, M., and M. Spiteller. 1986. The chemistry of the "unknown" soil nitrogen. Trans. 13th Int. Congr. Soil Sci. 3:473–474.

Schöning, I., G. Morgenroth, and I. Kögel-Knabner. 2005. O/N-alkyl and alkyl C are stabilised in fine particle size fractions of forest soils. Biogeochemistry 73:475–497.

Schulten, H.R. 1987. Pytolysis and soft ionization mass-spectroscopy of aquatic-terresterial humic substances and soils. J. Anal. Pyrol. 12:149–186.

Schulten, H.-R., and P. Leinweber. 1996. Characterization of humic and soil particles by analytical pyrolysis and computer modeling. J. Anal. Appl. Pyrol. 38:1–53.

Schulten, H.R., P. Leinweber, and M. Schnitzer. 1998. Analytical pyrolysis and computer modelling of humic and soil particles, p. 281–324. *In* P.M. Huang, N. Senesi, and J. Buffle (eds.) Structure and surface reactions of soil particles. Wiley, New York.

Schulten, H.R., C. Sorge, and M. Schnitzer. 1995. Structural studies on soil nitrogen by curie-point pyrolysis gas chromatography mass spectrometry with nitrogen selective detection. Biol. Fertil. Soils 20:174–184.

Schulten, H.R., C. SorgeLewin, and M. Schnitzer. 1997. Structure of "unknown" soil nitrogen investigated by analytical pyrolysis. Biol. Fertil. Soils 24:249–254.

Schulze, D.G., and P.M. Bertsch. 1995. Synchrotron X-ray techniques in soil, plant, and environmental research. Adv. Agron. 55:1–66.

Schutter, M.E., and R.P. Dick. 2000. Comparison of fatty acid methyl ester (FAME) methods for characterizing microbial communities. Soil Sci. Soc. Am. J. 64:1659–1668.

Sebastia, M.T. 2004. Role of topography and soils in grassland structuring at the landscape and community scales. Basic Appl. Ecol. 5:331–346.

Seilsepour, M., and M. Rashidi. 2008. Modeling of soil cation exchange capacity based on soil colloidal matrix. American-Eurasian J. Agric. Environ. Sci. 3:365–369.

Serrâo, E.A.S., I.C. Falesi, J.B. da Vega, and J.R. Teizeira Neto. 1979. Productivity of cultivated pastures on low fertility soils of the Amazon region of Brazil, p. 195–225. *In* P.A. Sanchez and L.W. Tergas (eds.) Pasture production on acid soils of the tropics. CIAT, Cali, Columbia.

Sexstone, A.J., N.P. Revsbech, T.B. Parkin, and J.M. Tiedje. 1985. Direct measurement of oxygen profiles and denitrification rates in soil aggregates. Soil Sci. Soc. Am. J. 49:645–651.

Shevchenko, S.M., and G.W. Bailey. 1996. Life after death: Lignin-humic relationships reexamined. Crit. Rev. Environ. Sci. Technol. 26:95–153.

Six, J., R.T. Conant, E.A. Paul, and K. Paustian. 2002. Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. Plant Soil 241:155–176.

Six, J., G. Guggenberger, K. Paustian, L. Haumaier, E.T. Elliott, and W. Zech. 2001. Sources and composition of soil organic matter fractions between and within soil aggregates. Eur. J. Soil Sci. 52:607–618.

Skjemstad, J., and J.A. Baldock. 2008. Total and organic carbon, p. 225–238. *In* M.R. Carter and E.G. Gregorich (eds.) Soil sampling and methods of analysis. 2nd edn. CRC Press, Taylor & Francis Group, Boca Raton, FL.

Skjemstad, J.O., P. Clarke, J.A. Taylor, J.M. Oades, and S.G. McClure. 1996. The chemistry and nature of protected carbon in soil. Aust. J. Soil Res. 34:251–271.

Skjemstad, J.O., P. Clarke, J.A. Taylor, J.M. Oades, and R.H. Newman. 1994. The removal of magnetic materials from surface soils: A solid state 13C CP/MAS NMR study. Aust. J. Soil Res. 32:1215–1229.

Skjemstad, J.O., L.J. Janik, and J.A. Taylor. 1998. Non-living soil organic matter: What do we know about it? Aust. J. Exp. Agric. 38:667–680.

Skjemstad, J.O., D.C. Reicosky, A.R. Wilts, and J.A. McGowan. 2002. Charcoal carbon in US agricultural soils. Soil Sci. Soc. Am. J. 66:1249–1255.

Skjemstad, J.O., and L. Spouncer. 2003. Integrated soils modelling for the national carbon accounting system. Estimating changes in soil carbon resulting from changes in land use. National Carbon Accounting System Technical Report No. 36.

Skjemstad, J.O., L.R. Spouncer, B. Cowie, and R.S. Swift. 2004. Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools. Aust. J. Soil Res. 42:79–88.

Skjemstad, J.O., and J.A. Taylor. 1999. Does the Walkley-Black method determine soil charcoal? Commun. Soil Sci. Plant Anal. 30:2299–2310.

Skjemstad, J.O., J.A. Taylor, and R.J. Smernik. 1999. Estimation of charcoal (char) in soils. Commun. Soil Sci. Plant Anal. 30:2283–2298.

Skyllberg, U., and T. Magnusson. 1995. Cations adsorbed to soil organic matter—A regulatory factor for the release of organic carbon and hydrogen ions from soils to waters. Water Air Soil Pollut. 85:1095–1100.

Smernik, R.J., and J.A. Baldock. 2005. Does solid-state 15N NMR spectroscopy detect all soil organic nitrogen? Biogeochemistry 75:507–528.

Smernik, R.J., and J.M. Oades. 2000a. The use of spin counting for determining quantitation in solid state ^{13}C NMR spectra of natural organic matter. 1. Model systems and the effects of paramagnetic impurities. Geoderma 96:101–129.

Smernik, R.J., and J.M. Oades. 2000b. The use of spin counting for determining quantitation in solid state ^{13}C NMR spectra of natural organic matter. 2. HF-treated soil fractions. Geoderma 96:159–171.

Smith, P., and P. Ineson. 2007. The soil carbon dioxide sink, p. 50–57. *In* D. Reay, C.N. Hewitt, K.A. Smith, and J. Grace (eds.) Greenhouse gas sinks. CAB International, Oxfordshire, U.K.

Smith, J.L., and E.A. Paul. 1990. The significance of soil microbial biomass estimations. Soil Biochem. 6:357–396.

Snowdon, P., P. Ryan, and J. Raison. 2005. Review of C:N ratios in vegetation, litter and soil under Australian native forests and plantations. National Carbon Accounting System-Technical Report 45:72 pp.

Söderlund, R., and B.H. Svensson. 1976. The global nitrogen cycle. *In* B.H. Svensson and R. Söderlund (eds.) Nitrogen, phosphorous and sulphur-global cycles, SCOPE Report 7, Ecol. Bull. (Stockholm), 22:23–73.

Sørensen, L.H. 1972. Stabilization of newly formed amino acid metabolites in soil by clay minerals. Soil Sci. 114:5–11.

Sørensen, L.H. 1975. Influence of clay on rate of decay of amino acid metabolites synthesized in soils during decomposition of cellulose. Soil Biol. Biochem. 7:171–177.

Sohi, S.P., N. Mahieu, J.R.M. Arah, D.S. Powlson, B. Madari, and J.L. Gaunt. 2001. A procedure for isolating soil organic matter fractions suitable for modeling. Soil Sci. Soc. Am. J. 65:1121–1128.

Sohi, S.P., N. Mahieu, D.S. Powlson, B. Madari, R.H. Smittenberg, and J.L. Gaunt. 2005. Investigating the chemical characteristics of soil organic matter fractions suitable for modeling. Soil Sci. Soc. Am. J. 69:1248–1255.

Sokoloff, V.P. 1938. Effect of neutral salts of sodium and calcium on carbon and nitrogen in soils. J. Agric. Res. 57:201–216.

Sollins, P., P. Homann, and B.A. Caldwell. 1996. Stabilization and destabilization of soil organic matter: Mechanisms and controls. Geoderma 74:65–105.

Solomon, D., J. Lehmann, J. Kinyangi, B.Q. Liang, K. Heymann, L. Dathe, K. Hanley, S. Wirick, and C. Jacobsen. 2009. Carbon (1s) NEXAFS spectroscopy of biogeochemically relevant reference organic compounds. Soil Sci. Soc. Am. J. 73:1817–1830.

Solomon, D., J. Lehmann, I. Lobe, C.E. Martinez, S. Tveitnes, C.C. Du Preez, and W. Amelung. 2005. Sulphur speciation and biogeochemical cycling in long-term arable cropping of subtropical soils: Evidence from wet-chemical reduction and S K-edge XANES spectroscopy. Eur. J. Soil Sci. 56:621–634.

Sombroek, W.G., F.O. Nachtergaele, and A. Hebel. 1993. Amounts, dynamics and sequestering of carbon in tropical and subtropical soils. Ambio 22:417–426.

Spain, A.V. 1990. Influence of environmental conditions and some soil chemical properties on the carbon and nitrogen contents of some Australian rainforest soils. Aust. J. Soil Res. 28:825–839.

Spain, A.V., R.F. Isbell, and M.E. Probert. 1983. Soil organic matter, p. 551–563. Soils: An Australian viewpoint. CSIRO, Melbourne, Australia.

Starr, M., C.J. Westman, and J. AlaReini. 1996. The acid buffer capacity of some Finnish forest soils: Results of acid addition laboratory experiments. Water Air Soil Pollut. 89:147–157.

Stehlickova, L., M. Svab, L. Wimmerova, and J. Kozler. 2009. Intensification of phenol biodegration by humic substances. Int. Biodeterior. Biodegrad. 63:923–927.

Stevenson, F.J. 1986. Cycles of soil: Carbon, nitrogen, phosphorus, sulfur, micronutrients. John Wiley & Sons, New York.

Stevenson, F.J. (ed.). 1994. Humus chemistry: Genesis, composition and reactions. John Wiley & Sons, New York.

Stewart, J.W.B., and C.V. Cole. 1989. Influences of elemental interactions and pedogenic processes in organic matter dynamics. Plant Soil 115:199–209.

Stewart, C.E., K. Paustian, R.T. Conant, A.F. Plante, and J. Six. 2008a. Soil carbon saturation: Evaluation and corroboration by long-term incubations. Soil Biol. Biochem. 40:1741–1750.

Stewart, C.E., A.F. Plante, K. Paustian, R.T. Conant, and J. Six. 2008b. Soil carbon saturation: Linking concept and measurable carbon pools. Soil Sci. Soc. Am. J. 72:379–392.

Stockdale, E.A., and P.C. Brookes. 2006. Detection and quantification of the soil microbial biomass—Impacts on the management of agricultural soils. J. Agric. Sci. 144:285–302.

Strong, D.T., H. De Wever, R. Merckx, and S. Recous. 2004. Spatial location of carbon decomposition in the soil pore system. Eur. J. Soil Sci. 55:739–750.

Swanston, C.W., B.A. Caldwell, P.S. Homann, L. Ganio, and P. Sollins. 2002. Carbon dynamics during a long-term incubation of separate and recombined density fractions from seven forest soils. Soil Biol. Biochem. 34:1121–1130.

Swanston, C.W., M.S. Torn, P.J. Hanson, J.R. Southon, C.T. Garten, E.M. Hanlon, and L. Ganio. 2005. Initial characterization of processes of soil carbon stabilization using forest stand-level radiocarbon enrichment. Geoderma 128:52–62.

Swift, R.S. 2001. Sequestration of carbon by soil. Soil Sci. 166:858–871.

Szabolcs, I. 1994. The concept of soil resilience, p. 33–39. *In* D. Greenland and I. Szabolcs (eds.) Soil resilience and sustainable land use. CAB International, Wallingford, Oxon, England.

Tan, K.H. 1986. Degradation of soil minerals by organic acids, p. 1–27. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and microbes. Soil Science Society of America, Inc., Madison, WI.

Tan, K.H., and P.S. Dowling. 1984. Effect of organic-matter on CEC due to permanent and variable changes in selected temperate region soils. Geoderma 32:89–101.

Tegelaar, E.W., J.W. Deleeuw, and C. Saizjimenez. 1989. Possible origin of aliphatic moieties in humic substances. Sci. Total Environ. 81–82:1–17.

Thompson, M.L., H. Zhang, M. Kazemi, and J.A. Sandor. 1989. Contribution of organic matter to cation exchange capacity and specific surface area of fractionated soil materials. Soil Sci. 148:250–257.

Thorn, K.A., and M.A. Mikita. 1992. Ammonia fixation by humic substances: A N-15 and C-13 NMR study. Sci. Total Environ. 113:67–87.

Thorn, K.A., P.J. Pettigrew, and W.S. Goldenberg. 1996. Covalent binding of aniline to humic substances. 2. N-15 NMR studies of nucleophilic addition reactions. Environ. Sci. Technol. 30:2764–2775.

Thurston, J.M., E.D. Williams, and A.E. Johnston. 1976. Modern developments in an experiment on permanent grassland started in 1856: Effects of fertilizers and lime on botanical composition and crop and soil analyses. Ann. Agron. 27:1043–1082.

Tiessen, H., J.W.B. Stewart, and D.W. Anderson. 1994. Determinants of resilience in soil nutrient dynamics, p. 157–170. *In* D. Greenland and I. Szabolcs (eds.) Soil resilience and sustainable land use. CAB International, Wallingford, Oxon, England.

Tiessen, H., J.W.B. Stewart, and C.V. Cole. 1984. Pathways of phosphorus transformations in soils of differing pedogenesis. Soil Sci. Soc. Am. J. 48:853–858.

Tiessen, H., J.W.B. Stewart, and J.O. Moir. 1983. Changes in organic and inorganic phosphorus composition of two grassland soils and their particle size fractions during 60–90 years of cultivation. J. Soil Sci. 34:815–823.

Tisdall, J.M., and J.M. Oades. 1982. Organic matter and water-stable aggregates in soils. J. Soil Sci. 33:141–163.

Toor, G.S., S. Hunger, J.D. Peak, J.T. Sims, and D.L. Sparks. 2006. Advances in the characterization of phosphorus in organic wastes: Environmental and agronomic applications, p. 1–72. *In* D.L. Sparks (ed.) Advances in agronomy. Vol. 89. Academic Press, San Diego, CA.

Trumbore, S.E., and S.H. Zheng. 1996. Comparison of fractionation methods for soil organic matter ¹⁴C analysis. Radiocarbon 38:219–229.

Tuomela, M., M. Vikman, A. Hatakka, and M. Itävaara. 2000. Biodegradation of lignin in a compost environment: A review. Bioresour. Technol. 72:169–183.

Turner, R.E., N.N. Nancy, D. Justic, and Q. Dortch. 2003. Future aquatic nutrient limitations. Mar. Pollut. Bull. 46:1032–1034.

Turnpault, M.P., P. Bonnaud, J. Fichter, J. Ranger, and E. Dambrine. 1996. Distribution of cation exchange capacity between organic matter and mineral fractions in acid forest soils (Vosges mountains, France). Eur. J. Soil Sci. 47:545–556.

Uriyo, A.P., and A. Kesseba. 1975. Amounts and distribution of organic phosphorus in some profiles in Tanzania. Geoderma 13:201–210.

Vacca, D.J., W.F. Bleam, and W.J. Hickey. 2005. Isolation of soil bacteria adapted to degrade humic acid-sorbed phenanthrene. Appl. Environ. Microbiol. 71:3797–3805.

Vaccaro, S., A. Muscolo, D. Pizzeghello, R. Spaccini, A. Piccolo, and S. Nardi. 2009. Effect of a compost and its water-soluble fractions on key enzymes of nitrogen metabolism in maize seedlings. J. Agric. Food Chem. 57: 11267–11276.

Valdrighi, M.M., A. Pera, M. Agnolucci, S. Frassinetti, D. Lunardi, and G. Vallini. 1996. Effects of compost-derived humic acids on vegetable biomass production and microbial growth within a plant (*Cichorium intybus*)-soil system: A comparative study. Agric. Ecosyst. Environ. 58:133–144.

Valdrighi, M.M., A. Pera, S. Scatena, M. Agnolucci, and G. Vallini. 1995. Effects of humic acids extracted from mined lignite or composted vegetable residues on plant growth and soil microbial population. Compost. Sci. Util. 3:30–38.

Vallini, G., A. Pera, M. Agnolucci, and M.M. Valdrighi. 1997. Humic acids stimulate growth and activity of in vitro tested axenic cultures of soil autotrophic nitrifying bacteria. Biol. Fertil. Soils 24:243–248.

Vallini, G., A. Pera, L. Avio, M.M. Valdrighi, and M. Giovannetti. 1993. Influence of humic acids on laurel growth, associated rhizospheric microorganisms, and mycorrhizal fungi. Biol. Fertil. Soils 16:1–4.

Valzano, F., B. Murphy, and T. Loen. 2005. The impact of tillage on changes in soil carbon with special emphasis on Australian conditions. National Carbon Accounting System-Technical Report 43:164 pp.

van Breemen, N., J. Mulder, and C.T. Driscoll. 1983. Acidification and alkalinization of soils. Plant Soil 75:283–308.

Vance, E.D., P.C. Brookes, and D.S. Jenkinson. 1987a. Microbial biomass measurements in forest soils: The use of the chloroform fumigation incubation method in strongly acid soils. Soil Biol. Biochem. 19:697–702.

Vance, E.D., P.C. Brookes, and D.S. Jenkinson. 1987b. An extraction method for measuring soil microbial biomass-C. Soil Biol. Biochem. 19:703–707.

van der Linden, A.M.A., L.J.J. Jeurisson, J.A. Van Veen, and B. Schippers. 1989. Turnover of soil microbial biomass as influenced by soil compaction, p. 25–36. *In* J. Attansen and K. Henriksen (eds.) Nitrogen in organic wastes applied to soil. Academic Press, London, U.K.

van Krevelen, D.W. 1950. Graphical-statistical method for the study of structure and reaction processes of coal. Fuel 29:269–284.

van Veen, J.A., and P.J. Kuikman. 1990. Soil structural aspects of decomposition of organic matter by microorganisms. Biogeochemistry 11:213–233.

Vanderhart, D.L., and R.H. Atalla. 1984. Studies of microstructure in native celluloses using solid-state C-13 NMR. Macromolecules 17:1465–1472.

Varanini, Z., and R. Pinton. 2001. Direct versus indirect effects of soil humic substances on plant growth and nutrition, p. 141–158. *In* R. Pinton et al. (eds.) The rhizosphere. Marcel Dekker, Basel, Switzerland.

Vaughan, D., and R.E. Malcolm. 1985. Soil organic matter and biological activity. Martinus Nijhoff/Dr. W. Junk Publishers, Dordrecht, the Netherlands.

Vedrova, E.F. 1997. Organic matter decomposition in forest litters. Eurasian Soil Sci. 30:181–188.

Vereecken, H., J. Maes, J. Feyen, and P. Darius. 1989. Estimating the soil moisture retention characteristic from texture, bulk density and carbon content. Soil Sci. 148:389–403.

Verlinden, G., B. Pycke, J. Mertens, F. Debersaques, K. Verheyen, G. Baert, J. Bries, and G. Haesaert. 2009. Application of humic substances results in consistent increases in crop yield and nutrient uptake. J. Plant Nutr. 32:1407–1426.

Visser, S.A. 1985. Physiological action of humic substances on microbial cells. Soil Biol. Biochem. 17:457–462.

Volkoff, B., and C.C. Cerri. 1988. Humus of Brazilian soils. Nature and relationship with the environment. Cahiers ORSTOM 24:83–95.

von Liebig, J. 1840. Die Chemie in ihrer Anwendung auf Agrikultur and Physiologie Braunschweig. Vieweg. As cited by Hedges and Oades (1997).

von Lützow, M., I. Kogel-Knabner, B. Ludwig, E. Matzner, H. Flessa, K. Ekschmitt, G. Guggenberger, B. Marschner, and K. Kalbitz. 2008. Stabilization mechanisms of organic matter in four temperate soils: Development and application of a conceptual model. J. Plant Nutr. Soil Sci. 171:111–124.

von Lützow, M., J. Leifeld, M. Kainz, I. Kögel-Knabner, and J.C. Munch. 2002. Indications for soil organic matter quality in soils under different management. Geoderma 105:243–258.

Wakelin, S.A., L.M. Macdonald, S.L. Rogers, A.L. Gregg, T.P. Bolger, and J.A. Baldock. 2008. Habitat selective factors influencing the structural composition and functional capacity of microbial communities in agricultural soils. Soil Biol. Biochem. 40:803–813.

Wardle, D.A. 1992. A comparative assessment of factors which influence microbial biomass carbon and nitrogen levels in soil. Biol. Rev. Camb. Philos. Soc. 67:321–358.

Watson, C.J., G. Travers, D.J. Kilpatrick, A.S. Laidlaw, and E. O'Riordan. 2000. Overestimation of gross N transformation rates in grassland soils due to non-uniform exploitation of applied and native pools. Soil Biol. Biochem. 32:2019–2030.

Webster, J.R., J.B. Waide, and B.C. Patten. 1975. Nutrient recycling and stability of ecosystems, p. 1–27. *In* F.G. Howell et al. (eds.) Mineral cycling in Southeastern ecosystems. National Technical Information Service, Springfield, VA.

Wegner, P.F., C.J. McDowall, and A.B. Frensham. 1989. Monitoring farming systems. Effects of cropping practices. Limitations to wheat yields. Technical Paper 23. Department of Agriculture of South Australia.

Weihermuller, L., J. Siemens, M. Deurer, S. Knoblauch, H. Rupp, A. Gottlein, and I. Putz. 2007. In situ soil water extraction: A review. J. Environ. Qual. 36:1735–1748.

Whiteley, G.M., and C. Pettit. 1994. Effect of lignite humic acid treatment on the rate of decomposition of wheat straw. Biol. Fertil. Soils 17:18–20.

Widmer, F., A. Fließbach, E. Laczkó, J. Schulze-Aurich, and J. Zeyer. 2001. Assessing soil biological characteristics: A comparison of bulk soil community DNA-, PLFA-, and Biolog-analyses. Soil Biol. Biochem. 33:1029–1036.

Yan, F., S. Schubert, and K. Mengel. 1996. Soil pH increase due to biological decarboxylation of organic anions. Soil Biol. Biochem. 28:617–624.

Yasmeen, S., M. Sariah, I. Razi, R. Mawardi, and A. Asgar. 2009. In vitro fungicidal activity of humic acid fraction from oil palm compost. Int. J. Agric. Biol. 11:448–452.

Yimer, F., S. Ledin, and A. Abdelkadir. 2008. Concentrations of exchangeable bases and cation exchange capacity in soils of cropland, grazing and forest in the Bale Mountains, Ethiopia. For. Ecol. Manage. 256:1298–1302.

Zepp, R.G., and C. Sonntag. 1995. The role of nonliving organic matter in the earth's carbon cycle. Wiley, New York, 342 pp.

Zunino, H., F. Borie, S. Aguilera, J.P. Martin, and K. Haider. 1982. Decomposition of ^{14}C-labeled glucose, plant and microbial products and phenols in volcanic ash-derived soils of Chile. Soil Biol. Biochem. 14:37–43.

12
Soil Solution

Paul Schwab
Purdue University

The soil solution is the hub of chemical and biological activity in the soil. Most reactions are slow in the absence of water: Soil organisms become dormant or die, mineral transformations become imperceptibly slow, and soil chemical weathering and formation processes become greatly impeded. Addition of moisture to previously dry systems reinitiates these reactions.

12.1 Basic Concepts

12.1.1 Definitions

The simplest definition of the soil solution is "aqueous liquid phase of the soil and its solutes" (SSSA, 2009). A broader definition describes the soil solution as a natural system that is both open and dynamic. Its composition is the result of the concurrent reactions of the labile soil minerals, organic materials, and biological metabolism. Any discussion of the soil solution, however, must take into account that from physical, chemical, and biological perspectives, a significant fraction of the soil solution

seemingly does not participate in some important phenomena. Many examples exist in different disciplines of soil science that illustrate this point. In soil physics, the concepts of "mobile" and "immobile" water have been introduced to explain why water is often observed to move more rapidly through soil than would be predicted assuming that water participates in the transport process. Higher plants and microorganisms undergo serious moisture deficit even when measurable water still exists in the soil. The composition and kinetics of reaction of water very near solid surfaces in the soil are significantly different than in the bulk solution. Therefore, a more functional definition of soil solution goes beyond simply the "aqueous liquid phase" and reflects that the soil solution may be dependent upon the application and, in some instances, the methodology used to obtain this solution from the soil.

From the viewpoint of a soil chemist, the soil solution may be defined as "the aqueous liquid phase in soil with a composition that is influenced by exchanges of matter and energy with soil air, soil solid phases, the biota, and the gravitational field of the

earth" (Sposito, 2008). By defining it as a phase, Sposito implicitly requires that the soil solution has uniform bulk properties (such as composition and temperature) and can be isolated from the soil. The requirement of uniformity can be met only on small scales of time and space because of the dynamic and spatially variable nature of soils.

The soil solution viewed on a molecular level would reveal that it is not a distinct entity but a part of the

> continuum of phases exhibiting indistinct interfaces at the molecular level. Solutes in the aqueous phase may be associated with bound water at the surfaces of soil colloids, free water percolating through soil macropores, water in the free space of plant roots, or immobile water in soil micropores.
>
> **(Wolt, 1994)**

12.1.2 Composition

Because soil solutions are highly variable over space and time, their composition can be discussed only in general terms. Trends in relative concentrations of soluble inorganic and organic constituents (see Chapter 11) are similar for most soils, but natural and anthropogenic factors influence some or all components.

12.1.2.1 Inorganic Constituents

The most common inorganic cations in the soil solution (Table 12.1) usually are Ca^{2+}, Mg^{2+}, and K^+ (in that order) with a large number of minor cations, including various forms of Na^+, Fe^{2+}, Cu^{2+}, and Zn^{2+}. The most prevalent anions are HCO_3^-, Cl^-, and SO_4^{2-}. Soils that become flooded are strongly influenced by reducing conditions and microbial activities, and Fe^{2+}, HS^-, and SO_4^{2-} can take on greater importance. In contaminated environments, the entire composition of the soil solution will change to reflect the most soluble components of the contaminants.

The sum of soluble cations and anions usually is less than 10^{-2} mol L^{-1}. Among the cations, Ca^{2+} will comprise at least half of this total; for the anions, the division among HCO_3^-, Cl^-, and SO_4^{2-} will be dependent upon soil pH and composition of the soil solids. Naturally, saline soils (see Chapter 17 of *Handbook of Soil Sciences: Resource Management and Environmental Impacts*) will

contain elevated concentrations of Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, and sometimes Na^+, depending upon pH and the source of the salts.

Concentrations of inorganic constituents in the soil solution are controlled by many factors, including pH, redox potential, and solid-phase composition. These considerations will be addressed later in this chapter and elsewhere in Chapters 15 through 18.

12.1.2.2 Organic Constituents

The composition of the soluble organic compounds in the soil solution is not as complex as that in the organic solid phase (see Chapter 11) but is a reflection of the organic components of the solid phase. Soluble organics also originate from the organisms in the soil including exudates from plant roots and soil microbes. Because these compounds are generated and degraded by microbial activity, the concentrations of these compounds in the soil solution can be more transient and more variable than soluble inorganic constituents.

As with soluble inorganic species, the organic constituents can be positively charged, negatively charged, or neutral. The chemical reactions of these compounds will be influenced by their electronic charge, but hydrophobic and hydrophilic tendencies are critical as well. For example, positively charged compounds that are water soluble can be adsorbed by cation exchange sites, negatively charged organics may form strong complexes with iron or aluminum oxides, and uncharged compounds may adsorb onto hydrophobic sites of soil minerals and organic matter.

The low-molecular-weight carboxylic acids typify soluble organics (Table 12.2). They are continually exuded by plant roots and soil microorganisms and are readily degraded with a half-life of hours. Carboxylic acids have the COOH functional group. The proton usually is dissociated in soil solutions, and the resulting negatively charged ligand often can act as a complexing agent for metals. Formic and tartaric acids are exuded by grass roots, acetic acid is generated under anaerobic conditions, oxalic acid is often associated with ectomycorrhizae, and citric acid is excreted by fungi and plant roots (Sposito, 2008). Typical concentrations for these acids range from 10^{-5} to 10^{-3} mol L^{-1}. Stevenson (1967) published an extensive review of these and other organic constituents in the soil solution.

TABLE 12.1 Inorganic Components Found in Soil Solutions

Category	Major Components (10^{-4}–10^{-2} mol L^{-1})	Minor Components (10^{-6}–10^{-4} mol L^{-1})	Others[a]
Cations	Ca^{2+}, Mg^{2+}, Na^+, K^+	Fe^{2+}, Mn^{2+}, Zn^{2+}, Cu^{2+}, NH_4^+, Al^{3+}	Cr^{3+}, Ni^{2+}, Cd^{2+}, Pb^{2+}
Anions	HCO_3^-, Cl^-, SO_4^{2-}	$H_2PO_4^-$, F^-, HS^-	CrO_4^{2-}, $HMoO_4^-$
Neutral	$Si(OH)_4^o$	$B(OH)_3^o$	

Concentration ranges are estimates and could change depending upon specific environments. The components listed are not necessarily the dominant solution species.

[a] Components normally found in concentrations <10^{-6} mol L^{-1} unless in a contaminated environment.

TABLE 12.2 Organic Components Found in Typical Soil Solutions

Category	Major Components (10^{-4}–10^{-2} mol L^{-1})	Minor Components (10^{-6}–10^{-4} mol L^{-1})
Naturally occurring	Formate, acetate, oxalate, amino acids, simple sugars	Citrate, phenolics, siderophores, proteins, alcohols, sulfhydryls
Anthropogenic	[a]	Herbicides, fungicides, insecticides, PCBs, petroleum hydrocarbons, surfactants, solvents, antibiotics, hormones

Concentration ranges are estimates and are subject to variability depending upon specific environmental conditions.

[a] Organic contaminants are present in low concentrations except in the case of spills, leaks, and accidental releases.

Organic pollutants include those compounds that have been introduced artificially into the soil, whether by design or accident. Most organic pesticides are applied at very low concentrations (0.1–10 kg ha^{-1}). Organics that have been spilled or accidentally leaked onto the soil can be present at many orders of magnitude higher. The resulting concentration in the soil solution will be dependent upon the total concentration in the soil, water solubility, strength of adsorption, volatility, and degradability of the compound in question. For example, the triazine herbicide, atrazine, normally is applied at a rate of approximately 2 kg ha^{-1}, has a water solubility of 35 mg L^{-1}, is weakly adsorbed by the soil, has a half-life of approximately 30 days, and has limited volatility. In contrast, glyphosate is applied at about the same rate as atrazine but is 1000 times more soluble, very strongly adsorbed, and not volatile. Immediately after application, atrazine concentrations in the soil solution can exceed 5 mg L^{-1} and represent an environmental threat to surface and groundwater. In contrast, aqueous glyphosate concentrations are nearly always <1 μg L^{-1}, and glyphosate is seldom detected in drinking water. (McBride, 1994, provides an excellent overview of the chemical behavior of pesticides in soil.) Organic solvents and petroleum hydrocarbons tend to have very low water solubilities (<1 μg L^{-1}), but their potential toxicity is high, and even trace concentrations in the soil solution may be environmentally significant.

12.2 Sampling the Soil Solution

The definition of the soil solution given in Section 12.1.1 is idealized and serves as a point from which the soil solution may be conceptualized. At the experimental level, the "soil solution" is defined by the method used to separate the aqueous phase from the rest of the soil. As discussed later, sampling methodology has a profound influence on the composition of the soluble constituents. Thus, consistent with the Heisenberg uncertainty principle for the study of subatomic particles, one cannot sample and examine the soil solution without altering it.

For both laboratory and field studies of the soil solution, one of the major problems associated with obtaining an unaltered soil solution is that the moisture content of field soils can rapidly change from air dry to saturated, and the moisture content influences the chemical and microbiological dynamics in the aqueous phase. Most techniques for obtaining samples of the soil solution function poorly when the moisture content is below saturation, and very few function at all moisture tensions of ≤33 kPa (1/3 bar). The choice of method and moisture content for obtaining a sample of the soil solution must be made to minimize the impact on solution composition while realizing that the act of sampling necessarily changes it.

An example of the influence of moisture content on soil chemical properties is the often-observed change in pH at different soil:solution ratios (Figure 12.1). Soil pH was measured for four soils with moisture contents ranging from 1:1 to a ratio of 10:1 water:soil (on a volume/mass basis). As the amount of
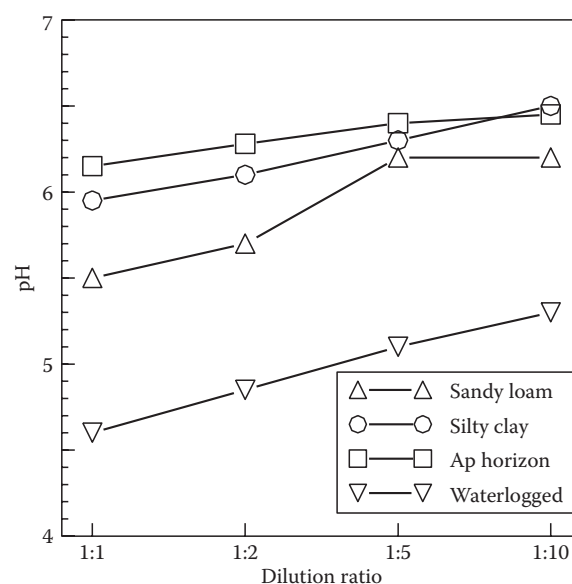
FIGURE 12.1 The impact of soil:solution ratio on the measured pH of two soils. Data for sandy loam and silty clay soils from Schwab (1992); surface horizon and waterlogged soil from Elberling and Matthiesen (2007). (Data from Schwab, A.P. 1992. Chemical and physical characterization of soils. *In* L.E. Erickson, S.C. Grant, and J.P. McDonald (eds.) Conf. Proc. Hazard. Subst. Res. University of Colorado, Boulder, Co., p. 326–344; Elberling, B., and H. Matthiesen. 2007. Methodologically controlled variations in laboratory and field pH measurements in waterlogged soils. Eur. J. Soil Sci. 58:207–214.)

water increased, the pH increased significantly. These observations lead to the following inescapable conclusion:

> Consideration of soil at field moisture contents is necessitated by the inability to predict consistently the effects of variation in soil to water ratios across broad ranges of soil solution composition; neither variation in total electrolyte concentrations or the activity ratios of specific ion components of the soil solution can be adequately resolved when water to soil ratios vary from field moisture contents to ratios >1. This is the main limitation to the use of water extracts as models of soil solution.

> **(Wolt, 1994)**

12.2.1 Laboratory Methods

Many laboratory methods have been developed for sampling the soil solution; only a few will be summarized here. The methods may be broadly categorized as aqueous extracts, column displacement, and pressure extraction. Each technique has advantages and limitations. (One must also consider proper sampling and handling techniques for the samples. For example, the simple act of air-drying the soil can have a profound effect on the soil chemical and microbiological properties [Bartlett and James, 1980; West et al., 1992].)

The steps to obtaining aqueous extracts include adding water to the soil to the point of saturation or beyond, equilibrating, and

removing solution. Equilibration times and separation techniques will depend upon the method used and desired application. If a saturated paste is prepared (United States Salinity Laboratory Staff, 1954) for the assessment of soil salinity, the paste is equilibrated for 16 h followed by vacuum removal of the soil solution. If equilibrium with soil solid phases is desired, equilibrations of hundreds to thousands of hours may be required (Kittrick, 1977; Schwab, 1989; Evans and Banwart, 2006; Sposito, 2008). The advantages of these techniques include ease of preparation, simple separation of soil and solution, and the ability to control many experimental parameters such as aeration and shaking. Limitations are unrealistic moisture contents, abrasion of soil surfaces during shaking, and uncertain impacts of wide soil:solution ratios on solution composition. This method has been applied with success in many instances (Lindsay, 1979; Berggren and Mulder, 1995).

Column displacement consists of forcing a fraction of the soil solution from the soil by leaching the soil with an aqueous solution (miscible displacement) or a water-insoluble organic solvent (immiscible displacement). These techniques have a long history, perhaps beginning with Thompson (1850) and Way (1850) leaching ammonium sulfate solutions through columns of soil and finding that the ammonium had been replaced by calcium. Schloesing (1866) used miscible displacement to obtain a sample of the soil solution; this method is quite similar to modern miscible displacement techniques (Adams, 1974). In all column displacement procedures, moist soil is packed into columns to a desired bulk density, sealing the surface by mechanically dispersing the clays, and leaching with the displacing liquid. The soil solution is collected in fractions until the displacing liquid appears in the leachate. This system may be modified to include pressure from the top (Ross and Bartlett, 1990) or vacuum applied to the bottom of the column (Wolt and Graveel, 1986).

The advantages of column displacement methods are obtaining soil solution from a soil at field moisture conditions, no requirement for grinding or shaking, and maintaining the dissolved gas content (except in the vacuum modification). Disadvantages include (1) generally short incubation times to prevent unusual microbial growth or changes in moisture content; (2) lack of knowledge of the fraction of the "true" soil solution that is obtained because of chemical and physical heterogeneity of the columns and because a large fraction of the water-filled pore space being occupied by the diffuse double layer; and (3) (in the case of immiscible displacement) an unknown influence of organic solvents on the displacement of ions.

Pressure extraction is the use of positive pressure, vacuum (negative pressure), or force applied by a centrifuge to remove the soil solution. For soils with moisture contents below saturation, the centrifuge method is more efficient than the positive pressure or vacuum methods. The efficiency of the latter two methods can be increased by employing a displacing solution, similar to some of the techniques described immediately above. Positive pressure methods generally refer to the use of a pressure membrane (Richards, 1941). A cylinder, equipped with small-pore filters on the inlet and outlet, is packed with the moist soil. Pressure of approximately 2 MPa (20 bar) is applied to the inlet, and the solution is collected

from the outlet. The centrifugation method, similarly, requires specialized equipment. The apparatus usually consists of a centrifuge tube to contain the moist soil; a permeable frit, filter, or ceramic beneath the soil; and a small volume at the bottom of the apparatus for collection of the extruded solution (Elkhatib et al., 1986). Large samples (up to 1 kg) usually are centrifuged at low speed, but smaller samples can be subjected to much greater speed and result in a greater fraction of the soil water being collected. An immiscible liquid also may be placed on the top of a small sample prior to centrifugation, a modification of immiscible displacement.

The pressure-membrane approach has not been used as widely as the centrifuge method because of lower yields, requirement for large samples, and specialty apparatus. The high-speed centrifuge method is an excellent choice because it uses equipment that is normally available in a soil chemistry laboratory, yields a sample very quickly (<1 h), and is amenable to a large number of samples. The disadvantages are similar to other displacement-type methods: short incubation times, an unknown fraction of the "true" soil solution that is sampled, uncertain impact of organic solvents (if used), and potential changes in the dissolved gas composition. The centrifugation method was used by van Hees et al. (2001) to infer the importance of organic matter and exchange reactions in the chemistry of Al in soils.

12.2.2 Field Methods

As is often the case in soil science, field methods of sampling the soil solution are more challenging than laboratory methods, particularly when one is interested in obtaining samples that are truly representative. Field methods cover a wide range of configurations: block or monolithic lysimeters; zero-tension or pan lysimeters; and porous-cup vacuum samplers.

Monolithic lysimeters are large blocks of soil (undisturbed or refilled) contained in a structure that has some means of collecting leachates. The apparatus are labor intensive and expensive to construct but are generally placed in a typical field setting to allow growth of vegetation while measuring a variety of soil parameters. The leachate collection system may be free drainage (i.e., zero-tension or air-entry potential) or use a vacuum system to impose a moderate moisture tension. Soil in contact with a free-drainage collection system must be very near saturation to allow water to move from the column to the collection area. The thickness of the near-saturated zone above the collection area will be dependent upon the texture of the soil, and the saturated condition will impact the chemistry, microbiology, and physics of the soil. Addition of a vacuum system, while adding to the complexity and cost of construction, will overcome the saturation problem. Unfortunately, the vacuum will at least partially "degas" the collected leachates.

Water flow through large lysimeters will affect directly the chemical composition of the leachates. In refilled versions, the original structure is destroyed, and, in the absence of plants, preferential flow within the soil is eliminated. However, this is unrepresentative of natural conditions. Block-type "undisturbed" lysimeters preserve much of the original soil structure, and water movement will be similar to the original field condition. In both block and refilled
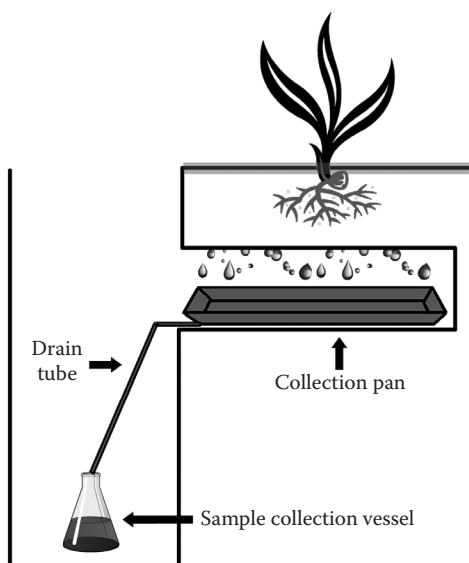
FIGURE 12.2 A zero-tension, pan lysimeter in a typical field context. The collection vessel and drain tube are located in an excavated trench cut into the soil along the side of the sample area.

lysimeters, wall flow (movement of water in the spaces between the lysimeter walls and the soil) can be an important form of preferential transport of water and solutes (Till and McCabe, 1976). Wall flow can be minimized through careful design and construction, and its relative impact is lessened as the ratio of surface area of the side walls to the total volume of soil increases.

Zero-tension lysimeters include any device installed in the soil that collects water by free drainage of the soil above including pans, troughs, funnels, and plates. These lysimeters usually are installed from a trench. A slot is excavated from the trench into the soil that allows installation of the lysimeters (Figure 12.2). Leachate collects in the lysimeter, and the sample is removed by means of a hose or tube leading to the trench or to the soil surface. This lysimeter design is useful for monitoring the solutes that move with water. The assumption is made that the soil disturbance from digging the trench and installing the pan will have little impact on water flow and solute composition. However, water will not flow into the pan until the soil is nearly saturated. Water can accumulate above the interface between the soil and lysimeter and cause incoming water to move laterally away from the lysimeter. Modifications to this procedure such as putting the lysimeter under tension (Cole et al., 1961) can overcome this problem in part.

Porous-cup vacuum lysimeters (Figure 12.3) are widely used in sampling the soil solution. Their design is simple, and they are easily installed using standard soil sampling equipment. Entire assemblies are readily built or available commercially. A length of PVC pipe is fitted with and glued to a ceramic cup. A rubber stopper with two air lines is placed in the opposite end of the pipe. The lysimeter is buried in the soil with the ceramic cup facing down and the air lines extending to the soil surface. One air line is used for drawing a vacuum, either with a hand pump or with a portable electric pump. The lines are sealed, and the lysimeter is allowed to draw the soil solution into the ceramic
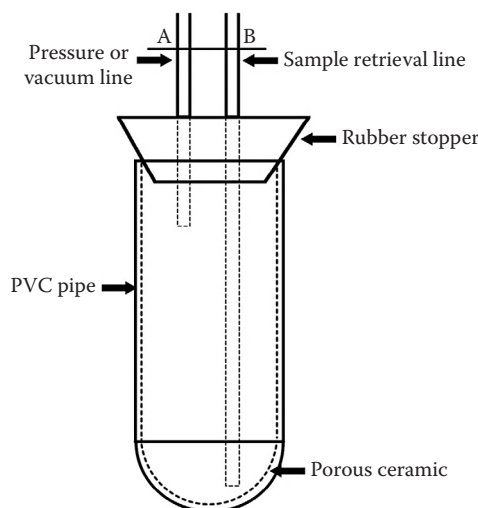
FIGURE 12.3 Design of a typical porous-cup vacuum lysimeter.

cup. (A continuous vacuum may be applied, or the vacuum may be applied intermittently.) The second air line extends to the bottom of the cup and is used for retrieving the collected sample either by drawing a vacuum on the retrieval line (marked B in Figure 12.3) or by applying pressure on the other line (A).

Porous-cup lysimeters have been criticized as field samplers of the soil solution (Caron et al., 1999). The ceramic cup has the potential to retain analytes or to contaminate the sample (Wood, 1973), particularly for inorganic species. Likewise, the PVC tube can retain organic compounds of interest. The area of soil that is sampled is not known (Warrick and Amoozegar, 1977), but will be influenced by the vacuum applied (Morrison and Lowry, 1990), the texture of the soil, the method of installation, and soil moisture content. Unlike the other field methods described above, vacuum pore water samplers will operate only when the operator engages them. Unless the porous cup is in saturated soil, the soil solution will not flow into it until vacuum is applied.

Each field method has positive and negative aspects. The monolithic lysimeters are, in essence, a field laboratory with all parameters either controllable or measurable. However, they are difficult and expensive to construct and maintain, and wall effects can be dominant. Zero-tension lysimeters are less expensive but require significant disturbance to the soil adjacent to the site, and water flow patterns may be altered immediately above the interface between soil and lysimeter. Vacuum pore water samplers are the least expensive of all field methods but collect the smallest fraction of the soil solution. In all cases, the soil solution will be defined by the specifics of the method used.

12.3 Thermodynamics of the Soil Solution

In discussing the thermodynamics of aqueous systems, strong distinctions are made between chemical equilibrium and nonequilibrium (kinetic) systems. Overlap exists between the statics and dynamics of solutions both on theoretical and

practical basis. In this section, only equilibrium thermodynamics will be addressed. Kinetics are addressed in Chapter 13. The assumption of equilibrium may be applied safely to only a fraction of the reactions occurring in soils; many reactions do not achieve equilibrium before new perturbations are imposed. Despite serious practical limitations, equilibrium models are important because "they are simpler in that they require less information, but they are nevertheless powerful when applied within their proper limits" (Stumm and Morgan, 1996).

Soil solutions are open systems in which energy and matter are exchanged readily with the surrounding environment. Experimentalists often are more familiar with closed systems in which such exchanges do not occur and rigorous mass and energy balances can be applied. Thus, the study of soils in natural settings can require some adjustments in design, execution, and interpretation. Open systems will be assumed in all derivations discussed later unless otherwise noted.

Chemical thermodynamics of aqueous solutions have been developed to various degrees of depth and detail, depending upon the application. Stumm and Morgan (1996) provide an excellent, in-depth discussion of equilibrium thermodynamics for aqueous systems. Wolt (1994) provided a discussion of similar depth but with the ultimate application to soil solutions. Sposito (1981) dedicated an entire textbook to the subject of the thermodynamics of soil solutions. The reader is referred to these references for a thorough discussion of chemical thermodynamics applied to aqueous solutions. In this chapter, important thermodynamic laws and equations will be stated and discussed; they will not be derived.

12.3.1 Fundamentals, Units, and Variables

As mentioned previously, true equilibrium is achieved in only a fraction of the reactions that occur in soil solutions but is a powerful tool when properly applied. Some of the reasons for investigating the application of equilibrium thermodynamics to systems that are frequently not in equilibrium include determining whether some or all of the components of the solution are in equilibrium; comparing the measured system with systems in equilibrium; quantifying the energy of disequilibrium (i.e., the energy input necessary to achieve equilibrium); and calculating the effects of temperature on equilibria. The equations necessary to obtain these goals are presented in the following sections.

TABLE 12.3 Variables, Thermodynamic Functions, and Equations of State

Variable	Units	Function	Equation of State
Temperature (T)	K	Enthalpy (H)	$dH = T\,dS + V\,dP$
Entropy (S)	J deg^{-1}	Helmholtz free energy (A)	$dA = -S\,dT - P\,dV$
Pressure (P)	kPa	Gibbs free energy (G)	$dG = -S\,dT + V\,dP$
Volume (V)	L	Internal energy (E)	$dE = T\,dS - P\,dV$
Chemical potential	J mol^{-1}		
Quantity	mol		

Important variables and functions and associated units are given in Table 12.3. Nearly all thermodynamic equations relative to soil solutions are derived from the four principles of thermodynamics and the first and second laws of thermodynamics (Stumm and Morgan, 1996). The four principles of thermodynamics establish an absolute temperature scale, define the internal energy of a system, and describe the relationship between entropy and temperature.

12.3.2 First and Second Laws of Thermodynamics

The first law of thermodynamics for equilibrium systems of fixed compositions,

$$dE = dq - dw, \tag{12.1}$$

represents the changes in internal energy (E) as affected by heat transferred to the system (q) and work done by the system (w). For a reversible process, the second law is given by

$$dS_{sys} = \frac{dq}{T} \tag{12.2}$$

and relates the temperature and heat transferred to a system with entropy (S). Other forms of this equation exist for the system plus its surroundings as well as irreversible changes in a system.

From the basic (not differentiated) equations $H = E + PV$ and $G = E + PV - TS$, one can obtain the important relationship, $G = H - TS$, where G is the Gibbs free energy, H is enthalpy, P is pressure, and V is volume. For a finite state change at constant pressure and temperature,

$$\Delta G = \Delta H - T\Delta S. \tag{12.3}$$

This equation, a restatement of the second law of thermodynamics in terms of state functions of the system, is used in the application of thermodynamics to numerical solutions of chemical problems.

12.3.3 Associated Thermodynamic Relationships

The development of theoretical thermodynamics can be quite detailed, but its utility can be realized only if expressed in measurable variables. For example, enthalpy change (dH) in Table 12.3 can be described as $dH = dq + V\,dP$; if $dP = 0$ (constant pressure process), then $dH = dq_P$. This expression is particularly useful in determining temperature effects because constant pressure heat capacity (C_P) is measurable and is given by

$$C_P = \left(\frac{dq}{dT}\right)_P. \tag{12.4}$$

In the absence of external work, $q_P = \Delta H$.

Under conditions of constant pressure and constant temperature the state of a system is characterized by dG. For irreversible changes in a system,

$$dG - V\,dP + S\,dT < 0, \tag{12.5}$$

and for reversible changes,

$$dG - V\,dP + S\,dT = 0. \tag{12.6}$$

The above equations are applicable only for systems of constant chemical composition, $dn_i = 0$. When the composition of only species i is allowed to change,

$$dG = \left(\frac{\partial G}{\partial T}\right)_{P,n_j} dT + \left(\frac{\partial G}{\partial P}\right)_{T,n_j} dP + \sum_i \left(\frac{\partial G}{\partial n_i}\right)_{P,T,n_j} dn_i. \tag{12.7}$$

12.3.4 Chemical Potential and Free Energy

The chemical potential, μ, of a species i is defined as

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{P,T,n_j}. \tag{12.8}$$

Employing the relationships $(\partial G/\partial T)_{P,nj} = -S$ and $(\partial G/\partial P)_{T,nj} = V$ yields the equation

$$dG = -S\,dT + V\,dP + \sum \mu_i\,dn_i, \tag{12.9}$$

and for a single-phase system at constant temperature and pressure,

$$dG = \sum \mu_i\,dn_i. \tag{12.10}$$

For a multiphase system, dG for the entire system is obtained by summing $\mu_i\,dn_i$ over all phases. When equilibrium is established for all reactions and phases, $dG = 0$.

These basic equations can be manipulated further to yield equations for chemical potentials of components in various phases.

The resulting relationships depend upon assumptions made during the derivation. Equations for gases and electrolytes are summarized in Table 12.4. As an example derivation for an ideal gas, consider the partial free energy of this gas with respect to P at constant T and n:

$$\left(\frac{\partial G}{\partial P}\right)_{T,n} = V \tag{12.11}$$

and, thus,

$$\left(\frac{\partial G}{\partial P}\right)_{T,n} = \frac{nRT}{P}. \tag{12.12}$$

If the second equation is rearranged and integrated from G^o (standard free energy) to G and from P^o (standard pressure) to P, the result is

$$G - G^o = nRT\ln\left(\frac{P}{P^o}\right). \tag{12.13}$$

Differentiation with respect to n and recalling the definition for chemical potential gives

$$\mu_i = \mu_i^o + RT\ln\left(\frac{P_i}{P^o}\right), \tag{12.14}$$

where μ_i^o is the chemical potential for the ideal gas in its standard state (e.g., as defined by the conditions given in Table 12.4) at $T = 298.15\,K$ and $P = 101.33\,kPa$ (1 atm). Similar derivations for real gases, condensed phases that obey Raoult's law and condensed phases that obey Henry's law, are summarized by Wolt (1994).

Sposito (1994) discussed the standard states for phases and elements relevant to the study of soil solutions.

TABLE 12.4 Expressions for the Chemical Potential of Components in Gas and Condensed Phases as Influenced by the Assumptions of the Behavior of the Component

Phase	Chemical Potential	Coefficients	Standard State
Ideal gas	$\mu_i = \mu_i^o + RT\ln(P_i/P^o)$		298.15 K, 101.33 kPa; pure ideal gas at 101.33 kPa
Nonideal gas	$\mu_i = \mu_i^o + RT\ln(P_i\lambda_i)$	$\lambda_i \equiv$ fugacity coefficient	298.15 K, 101.33 kPa; pure nonideal gas at 101.33 kPa, $\lambda_i = 1$
Condensed phase: follows Henry's law	$\mu_i = \mu_i^o + RT\ln(C_i)$	$C_i \equiv P_i/K_{Hi}$	Pure condensed phase component, $C_i^o = 1$
Condensed phase: does not follow Henry's law	$\mu_i = \mu_i^o + RT\ln(C_i\lambda_i)$	$\lambda_i \equiv$ Henry's law coefficient	Pure condensed phase component, $C_i^o = 1$ and $\lambda_i = 1$
Condensed phase: follows Raoult's law	$\mu_i = \mu_i^o + RT\ln(\chi_i)$	$\chi_i = P_i/P^o$	Pure condensed phase component, $\chi_i^o = 1$
Condensed phase: does not follow Raoult's law	$\mu_i = \mu_i^o + RT\ln(\chi_i\lambda_i)$	$\lambda_i \equiv$ Raoult's law coefficient	Pure condensed phase component, $\chi_i^o = 1$ and $\lambda_i = 1$
Solute in ideal, dilute solution	$\mu_i = \mu_i^o + RT\ln(C_i)$		Pure condensed phase component, $C_i^o = 1$
Solute in nonideal solution	$\mu_i = \mu_i^o + \mu RT\ln(C_i\gamma_i)$	$\gamma_i \equiv$ single ion activity coefficient	Pure condensed phase component, $C_i^o = 1$ and $\gamma_i = 1$
Solute in nonideal solution	$\mu_i = \mu_{\pm}^o + \nu RT\ln(C_{\pm}\gamma_{\pm})$	$\gamma_{\pm} \equiv$ mean activity coefficient $\nu =$ stoichiometric coefficient	Pure condensed phase component, $C_{\pm} = 1$ and $\gamma_{\pm} = 1$

Sources: Wolt, J.D. 1994. Soil solution chemistry. Applications to environmental science and agriculture. John Wiley & Sons, New York; Stumm, W., and J.J. Morgan. 1996. Aquatic chemistry. Chemical equilibria and rates in natural waters. Wiley Interscience, New York.

12.3.5 Chemical Potential and Activities in Ideal Solutions

Perhaps the most important expressions for chemical potential in dealing with soil solutions are those for solutions of electrolytes. A classical case is that of NaCl dissolved in water. Sodium chloride is a strong electrolyte, indicating that it is dissociated fully with no NaCl complexes in the aqueous solution. The chemical potential of this solution is the sum of the chemical potentials for the Cl^- and Na^+ species:

$$\mu_{NaCl} = \mu_{Na^+} + \mu_{Cl^-}. \tag{12.15}$$

In an ideal, dilute solution, the equations for the chemical potentials of the individual ions with concentrations C_{Na^+} and C_{Cl^-} would be

$$\mu_{Na^+} = \mu_{Na^+}^o + RT \ln(C_{Na^+}), \tag{12.16}$$

$$\mu_{Cl^-} = \mu_{Cl^-}^o + RT \ln(C_{Cl^-}). \tag{12.17}$$

The standard potential for an aqueous solution of NaCl ($\mu_{NaCl(aq)}^o$) is defined as

$$\mu_{NaCl(aq)}^o \equiv \mu_{Na^+}^o + \mu_{Cl^-}^o, \tag{12.18}$$

and the total chemical potential is

$$\mu_{NaCl(aq)} = \mu_{NaCl(aq)}^o + RT \ln(C_{Na^+} C_{Cl^-}). \tag{12.19}$$

If the concentration of the salt is defined as C_{NaCl} and realizing that $C_{NaCl} = C_{Na^+} = C_{Cl^-}$, then

$$\mu_{NaCl} = \mu_{NaCl}^o + RT \ln(C_{NaCl})^2. \tag{12.20}$$

12.3.6 Activities and Activity Coefficients: Nonideal Solutions

For nonideal solutions, the concept of activity and activity coefficient is introduced. A nonideal solution of NaCl would have a chemical potential of

$$\mu_{NaCl(eq)} = \mu_{NaCl(aq)}^o + RT \ln(a_{Cl^-} a_{Na^+}) \tag{12.21}$$

or

$$\mu_{NaCl(eq)} = \mu_{NaCl(aq)}^o + RT \ln(C_{Cl^-} \gamma_{Cl^-} C_{Na} + \gamma_{Na^+}), \tag{12.22}$$

in which a_i is the activity of component I and γ_I is the activity coefficient of component i. The product $a_{Na^+} a_{Cl^-}$ is designated by a_{NaCl}, which is the activity of the aqueous solute, and its mean activity is $a_\pm = (a_{Na^+} a_{Cl^-})^{1/2}$, which can be determined experimentally. Similarly, the mean activity coefficient, $\gamma_\pm$, is equal to $(\gamma_{Na^+} \gamma_{Cl^-})^{1/2}$; $\gamma_{Na^+} \gamma_{Cl^-}$ can be determined experimentally by measuring water vapor pressures over NaCl solutions of varying concentrations.

12.3.7 Equilibrium Constants

The Gibbs free energy of a reaction can be related to the system composition through the expressions for chemical potential:

$$\mu_I = \mu_i^o + RT \ln a_i \tag{12.23}$$

and ΔG

$$\Delta G = \sum_i \upsilon_i \mu_i, \tag{12.24}$$

in which υ_i is the stoichiometric coefficient of a component i in the reaction. Combining these equations gives

$$\Delta G = \sum_i \upsilon_i \mu_i^o + RT \sum_i \upsilon_i \ln(a_i) \tag{12.25}$$

or

$$\Delta G = \Delta G^o + RT \ln \prod_i (a_i)^{u_i}, \tag{12.26}$$

where

$$\Delta G^o = \sum_i \upsilon_i \mu_i^o, \tag{12.27}$$

ΔG^o is the standard Gibbs free energy change of the reaction
Π_i is the quotient of concentrations of products over reactants

Consider the reaction,

$$aA + bB \leftrightarrow cC + dD, \tag{12.28}$$
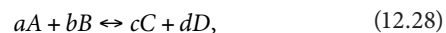
in which
  a and b are the stoichiometric coefficients of reactants A and B
  c and d are the stoichiometric coefficients of products C and D

The expression for Π_i would be

$$\left[\frac{a_C^c a_D^d}{a_A^a a_B^b} \right].$$

The expression often is given the symbol Q. Thus,

$$\Delta G = \Delta G^o + RT \ln Q. \tag{12.29}$$

At equilibrium, $\Delta G = 0$, $Q = K$ (the equilibrium constant), and

$$\Delta G^o = -RT \ln K. \tag{12.30}$$

Both K and Q are written in terms of activities (a_i), but activities of individual species are not measurable. However, both can be written in terms of concentrations using the relationship between activity and concentration, $a_i = C_i \gamma_1$:

$$K = \frac{a_C^c a_D^d}{a_A^a a_B^b} = \frac{(C_C \gamma_C)^c (C_D \gamma_D)^d}{(C_A \gamma_A)^a (C_B \gamma_B)^b}. \tag{12.31}$$

12.3.8 Effects of Temperature and Pressure

The general expression for the effect of temperature on the free energy of reaction is

$$\left(\frac{\partial \Delta G^o / T}{\partial T} \right)_P = -\frac{\Delta H^o}{T^2}, \tag{12.32}$$

and the van't Hoff equation for the corresponding equilibrium constant is

$$\frac{d \ln K}{dT} = \frac{\Delta H^o}{RT^2}. \tag{12.33}$$

The temperature dependence of enthalpy of reaction is

$$H_2 - H_1 = \int_{T_1}^{T_2} C_p \, dT. \tag{12.34}$$

If ΔH^o is independent of temperature,

$$\ln \frac{K_2}{K_1} = \frac{\Delta H^o}{R} \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \tag{12.35}$$

or, when the heat capacity ($\Box C_p^o$) is independent of temperature,

$$\ln \frac{K_2}{K_1} = \frac{\Delta H^o}{R} \left(\frac{1}{T_1} - \frac{1}{T_2} \right) + \frac{\Delta C_P^o}{R} \left(\frac{T_1}{T_2} - 1 - \ln \frac{T_1}{T_2} \right) \tag{12.36}$$

or

$$\ln K = B - \left(\frac{\Delta H_o}{RT} \right) + \frac{\Delta C_P^o}{R} \ln T, \tag{12.37}$$

where B and ΔH_o are constants (Stumm and Morgan, 1996). When $\Box C_p^o$ is a function of temperature, the form of the final equation will reflect the temperature-dependent expression for the heat capacity.

The effects of pressure on free energy and equilibrium constants are handled in a fashion similar to temperature (Stumm

and Morgan, 1996). The general expression, when ΔV^o is independent of pressure,

$$\left(\ln \frac{K_P}{K_1} \right)_P = -\frac{\Delta V^o (P - 1)}{RT}. \tag{12.38}$$

With specific reference to aqueous solutions,

$$\mu_i = \mu_i^o + RT \ln \gamma_i C, \tag{12.39}$$

$$\left(\frac{\partial \ln K}{\partial P} \right)_{T,C} = -\frac{\Delta V^o}{RT}, \tag{12.40}$$

$$\left(\frac{\partial \ln \gamma_i}{\partial P} \right)_{T,C} = -\frac{\overline{V}_i - \overline{V}_i^o}{RT}, \tag{12.41}$$

where

$\overline{V}_i$ is the partial molar volume
$\overline{V}_i^o$ is the standard partial molar volume of species i

12.3.9 Single-Ion Activity Coefficients

The mean activity coefficient of a salt in solution, $\gamma_\pm$, is measurable by experimental methods. However, the activity coefficient of a single ion, such as γ_{Na^+} or γ_{Cl^-}, is not measurable and must be estimated by theoretical models. Such a model was provided by the Debye–Huckel theory, which combined thermodynamic and electrostatic expressions to describe the interaction between charged species in solution. In its first configuration, the theory was based upon the assumption that the ions act as point charges. The resulting equation, the Debye–Huckel limiting law, was

$$\log \gamma_i = -AZ_i^2 (I^{0.5}), \tag{12.42}$$

where

Z_i is the valence of the ion
I is the ionic strength of the solution $\left(I = 0.5 \sum C_i Z_i^2 \right)$
A is related to the dielectric constant for water and has a value of 0.509 at 298.15 K and 101.33 kPa

Activity coefficients calculated from this equation begin to deviate from measurements when $I > 0.005 \, \text{mol L}^{-1}$. The Debye–Huckel theory was extended to greater ionic strengths ($I = 0.1 \, \text{mol L}^{-1}$) by adding terms that account for the spatial interaction of the ions:

$$\log \gamma_i = \frac{-AZ_i^2 I^{0.5}}{1 + \beta a_i^o I^{0.5}}, \tag{12.43}$$

in which β is a constant that depends upon the nature of the solvent and temperature. For water, β is equal to 0.328×10^8 at 298.15 K and 101.33 kPa. The ion size parameter, a_i^o, ranges from

(2.5 to 9) × 10⁻⁸ and must be obtained from a compilation of values (Kielland, 1937). The Davies equation is a further modification of the Debye–Huckel and is applicable when $I < 0.5\,\mathrm{mol\,L^{-1}}$:

$$\log \gamma_i = -AZ_i^2 \left(\frac{I^{0.5}}{1 + I^{0.5}} - 0.3I \right). \tag{12.44}$$

The advantages of the Davies over other equations are the applicability to higher ionic strengths and the elimination of the necessity of using β values.

12.3.10 Electrochemical Potential

The standard electrochemical potential of a oxidation/reduction reaction, E^o, is the potential of the reaction relative to the oxidation of $H_2(g)$ to H^+ in aqueous solution. The balanced chemical reaction of such a cell involving the reduction of Cu^{2+} to the metal would be

$$Cu^{2+}(aq) + H_2(g) \leftrightarrow 2H^+ + Cu(s). \tag{12.45}$$
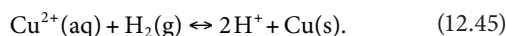
The half-cell reactions are conveniently written as follows, with the implicit understanding that the hydrogen half-cell is always present:

$$Cu^{2+}(aq) + 2e^- \leftrightarrow Cu(s). \tag{12.46}$$
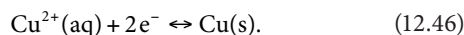
For any given reaction, the Nernst equation may be derived:

$$E_H = E_H^o + \frac{2.303RT}{nF}\log \frac{\prod_i \{ox\}^{n_i}}{\prod_j \{red\}^{n_j}}, \tag{12.47}$$

where
  E_H is the measured potential
  E_H^o is the standard potential of the cell
  R is the gas constant
  T is temperature in degrees K
  n is the number of electrons involved in the reaction
  F is the Faraday constant
  Π designates the product of either the reactants or products in the equation

For any oxidation/reduction reaction, the relationship between the electrode potential and pe, negative logarithm of the electron activity, $-\log(e^-)$, may be derived as

$$pe = \frac{F}{2.303RT}E_H, \tag{12.48}$$

or $pe = E_H/59.2$ when E_H is in mV and determined at 298.15 K. As with H^+, aqueous solutions do not contain free electrons, and the concentration of solvated electrons is vanishingly small. Nevertheless, pe is a very convenient parameter in manipulating equilibrium equations and plotting data.
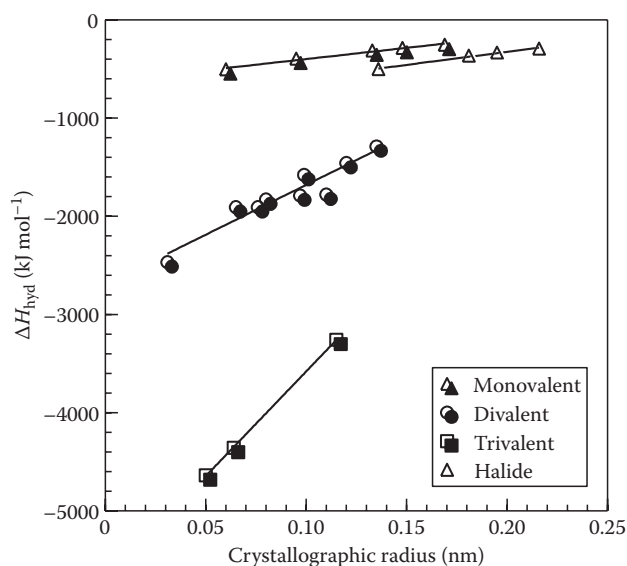
FIGURE 12.4 Relationship between crystallographic radius and enthalpy of hydration of monovalent, divalent, and trivalent cations and monovalent halide anions. The drawn lines represent linear regressions: $R^2 = 0.98$ for monovalent cations; $R^2 = 0.90$ for divalent cations; $R^2 = 0.99$ for trivalent cations; and $R^2 = 0.98$ for monovalent anions. (Data from Bohn, H., B. McNeal, and G. O'Connor. 1985. Soil chemistry. Wiley Interscience, New York; Morris, D.F.C. 1969. Ionic radii and enthalpy of hydration of ions. Struct. Bond. 63:157–159.)

12.3.11 Heat of Hydration of Ions

One of the most important reactions for ions in solution is hydration, the electrostatic interaction between the polar water molecules and the charged ion. When an ion is released into aqueous solution, heat is released as water molecules form a somewhat ordered structure around the ion. The water that surrounds the ion tends to insulate the charged species from other ions in solution. In infinitely dilute solutions, this solvation effect completely isolates the ions from interacting with each other.

The heat released during solvation of an ion by water is the heat of hydration, $\Delta H_{\text{hydration}}$. The strength of the water–ion interaction increases with increasing valence because higher charged ions have the capacity to react with more water molecules. Within a group of ions of the same valence, $\Delta H_{\text{hydration}}$ decreases (becomes more negative) linearly with decreasing crystallographic radius (Figure 12.4). Thus, a small monovalent ion such as Li^+ releases more heat upon hydration than the much larger Cs^+ ion. A direct result of this is that Li^+ also has a greater hydrated radius than Cs^+, which is partly responsible for some of the differences in strength of retention of these ions by cation exchange sites in soil (see Chapter 17).

12.4 Interactions of Gases with the Soil Solution

Chemical reactions between gases and the liquid aqueous phase are important not only in the soil solution but also in biological systems, surface water, groundwater, and the atmosphere.

For example, acid rain is a long-recognized problem that results from the combustion of fossil fuels. The oxidation of C, S, and N in the fuels generates several gaseous oxides including CO_2, NO_2, NO, SO_2, and SO_3. When these oxides dissolve in water, they generate acids: H_2CO_3, HNO_3, HNO_2, H_2SO_3, and H_2SO_4. The extent to which the gaseous oxides dissolve in the water can be described by Henry's law referenced in Section 12.3.4 and given in Table 12.4.

12.4.1 Henry's Law

The equilibrium distribution of a species between the gas phase and the aqueous is given by Henry's law. The expression is based upon the thermodynamic parameter, chemical potential, and requires the determination of a partitioning coefficient for each gas. The thermodynamic expression is

$$f_A = K a_A, \qquad (12.49)$$

where

　K is a constant
　f_A is the fugacity of the gas
　a_A is the activity of the species in the aqueous phase

The transition between the above thermodynamic equation and a usable expression with measurable terms is made simpler if one first assumes the condition of dilute solutions and low concentrations in the gas phase. Under these circumstances, Henry's law may be written in one of two ways, either of which is correct. In the first expression, the partitioning coefficient (H) is dimensionless:

$$\frac{\left[A(\text{aq})\right]}{\left[A(\text{g})\right]} = H. \qquad (12.50)$$

The units of concentration for the gaseous and aqueous species must be the same (e.g., mol L^{-1}). In the second form of Henry's law, the partitioning coefficient is not dimensionless:

$$\frac{\left[A(\text{aq})\right]}{P_A} = K_H. \qquad (12.51)$$

If the units of concentration for the aqueous component are mol L^{-1} and the partial pressure is in atmospheres, then K_H must have units of mol L^{-1} atm^{-1}. The conversion between H and K_H is

$$K_H = \frac{H}{RT}. \qquad (12.52)$$

Table 12.5 contains a compilation of K_H values for important gases in soils and other settings.

　Although Henry's law dictates that the solubility of gases in water is a linear function only of their partial pressures, many of the dissolved species react further with water to form acids that are subject to deprotonation. Thus, pH is an important

TABLE 12.5 Henry's Law Constants (K_H) for Important Gas–Water Reactions

Gas	$K_{H,298.15K}$ (mol L^{-1} atm^{-1})	Gas	$K_{H,298.15K}$ (mol L^{-1} atm^{-1})
H_2	7.78×10^{-4}	O_2	1.27×10^{-3}
N_2	6.53×10^{-4}	N_2O	2.42×10^{-2}
NO	1.92×10^{-3}	CO	9.77×10^{-4}
CO_2	3.39×10^{-2}	H_2S	1.02×10^{-1}
SO_2	1.36	Cl_2	9.31×10^{-2}
CH_4	1.41×10^{-3}	NH_3	5.71×10^{1}
O_3	1.04×10^{-4}		

　Constants were calculated from the data of Gevantman (2001) with the exception of NH_3 (Sposito, 2008).

controlling variable for the total dissolved component in aqueous solution. Calculations also can be complicated by whether the system contains an infinite sink of the gas (open system) or if the gas is limited (closed system). These calculations are handled in detail by Stumm and Morgan (1996).

12.4.2 Volatile Organic Compounds

Many organic compounds are subject to loss from solid, liquid, or aqueous phases through volatilization. As with inorganic gases, the tendency for volatile organic compounds to partition between the aqueous phase and the atmosphere can be described by Henry's law. Henry's law constants can be calculated by measuring aqueous and gaseous phase concentrations in systems at equilibrium (Table 12.6).

12.4.3 Rates of Dissolution of Gases in Water

Quantification of equilibrium distributions of volatile compounds between the atmosphere and aqueous phase is more powerful when accompanied by an understanding of the rates of reactions. Transfer of a gas across the liquid/gas interface can be approximated by a diffusion model with two diffusion films

TABLE 12.6 Water Solubilities, Vapor Pressures, and Henry's Law Constants for Selected Organic Compounds

Compound	Water Solubility (mol L^{-1})	Vapor Pressure, P_A (atm)	K_H (mol L^{-1} atm^{-1})
Hexane	7.0×10^{-4}	0.25	2.8×10^{-3}
n-Octane	5.8×10^{-6}	1.8×10^{-2}	3.1×10^{-4}
Dieldrin	5.8×10^{-7}	6.6×10^{-9}	88
Lindane	2.6×10^{-5}	8.3×10^{-8}	313
Naphthalene	2.6×10^{-4}	1.0×10^{-4}	2.6
Benzene	2.3×10^{-2}	0.12	0.19
Toluene	5.6×10^{-3}	3.7×10^{-2}	0.15
Biphenyl	4.9×10^{-5}	7.5×10^{-5}	0.65
Dimethyl sulfide	0.35	0.63	0.56

　Source: Stumm, W., and J.J. Morgan. 1996. Aquatic chemistry. Chemical equilibria and rates in natural waters. Wiley Interscience, New York.

　A more comprehensive compilation is available in Staudinger and Roberts (1996).

(liquid phase and gas phase) assuming that the bulk phases are well mixed (Stumm and Morgan, 1996). The diffusion films are assumed to be within a very small distance of the interface.

The flux across the films of thickness z is expressed by Fick's law:

$$F = -D\frac{dc}{dz}. \tag{12.53}$$

The units for flux, F, will depend upon the units of concentration (c), diffusion coefficient (D), and thickness; assuming concentrations in mol m^{-3}, D in m^2 s^{-1}, and distance in m, then F will have units of mol m^{-2} s^{-1}. A steady state will have been achieved when the flux across the two films is equal. The thickness of the air diffusion film is assumed to be z_{air}; the thickness of water diffusion film is z_{water}; the concentrations in the air and water are c_{air} and c_{water}; the concentrations in the air–water interface is c_{a-w}; and the concentration in the water–air interface is c_{w-a}. Therefore, at steady state,

$$F = -\frac{D_w}{z_w}(c_{w-a} - c_w) = -\frac{D_a}{z_a}(c_a - c_{a-w}). \tag{12.54}$$

Assuming that the transfer across the interfaces is much faster than any ensuing chemical reactions that may occur, Henry's law may be applied to the concentrations in the interface regions (using the dimensionless constant from Section 12.4.1):

$$H = \frac{c_{w-a}}{c_{a-w}} = K_H RT. \tag{12.55}$$

Substituting into the steady-state equation,

$$F = \frac{D_a}{z_a}\left(\frac{c_{w-a}}{H} - c_a\right) = \frac{D_w}{z_w}(c_w - c_{w-a}). \tag{12.56}$$

Using data from Lovelock et al. (1993 in Stumm and Morgan, 1996), one can calculate that the flux of freon across the water–gas interface in marine environments is 1.5×10^{-9} g cm^{-2} year^{-1}. Extended derivations and similar calculations can be made for other chemical systems after making allowances for chemical reactions. For example, the steady-state flux of $CO_2(g)$ from a lake into the atmosphere is 6×10^{-9} mol cm^{-2} s^{-1} at pH 6.7, total alkalinity of 3×10^{-3} mol L^{-1}, and 298.15 K.

12.5 Acid–Base Reactions in the Soil Solution

Acid and base reactions are the most fundamental and often the most important in soil solutions. The weathering of primary minerals often generates alkaline conditions, and natural and anthropogenic activities can generate acidic conditions. The rates and extents of many chemical reactions are dependent upon soil solution pH including biological activity, mineral

dissolution, partitioning of some gases, and bioavailability of critical nutrient elements. This section will address some of the important acid–base concepts and discuss the numerical handling of related equilibria.

12.5.1 Fundamentals of Acid–Base Chemistry in Soil Solution

The central component of acid–base reactions is the proton or hydrogen ion, H$^+$. This ion does not actually exist as H$^+$ in aqueous solutions but is hydrated to form H_3O^+, $H_7O_3^+$, $H_9O_4^+$, etc. For the sake of simplicity in representing the equilibria involving the proton, the symbol H$^+$ will be used.

According to the Bronsted–Lowry concept, an acid is any substance that donates a proton to another substance. Similarly, a Bronsted–Lowry base is any substance that accepts a proton from another substance. Using this definition, hundreds of reactions in soil solutions are examples of acids and bases. For example, bicarbonate ion dissociates readily to form carbonate and H$^+$:

$$HCO_3^- \leftrightarrow CO_3^{2-} + H^+. \tag{12.57}$$
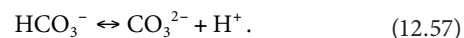
In this case, bicarbonate is acting as an acid. Bicarbonate may also act as a base, accepting a proton to form carbonic acid:

$$HCO_3^- + H^+ \leftrightarrow H_2CO_3^o. \tag{12.58}$$

Although the discussions of Bronsted–Lowry theory for acids and bases presented here will be limited to aqueous systems, the theory is applicable to all solvents in which protons may be exchanged, including ammonia, sulfuric acid, and ethanol.

Metal ions in aqueous solutions are readily solvated and, as such, exist as hydrates rather than bare ions. The number of water molecules surrounding a cation in the first hydration layer will be dependent upon the ionic radius of the cation and its charge. The water in the hydration sphere tends to act as a weak acid, donating a proton to the solution and (in essence) contributing a hydroxyl ion to the metal:

$$\left[Fe(H_2O)_6\right]^{3+} \leftrightarrow \left[Fe(H_2O)_5OH\right]^{2} + H^+. \tag{12.59}$$
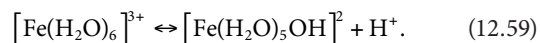
The acidity of the water associated with the hydrolysis reaction increases with increasing valence and decreasing ionic radius of the central cation.

Another acid–base concept was formulated by G.N. Lewis (Lewis and Randall, 1923). A Lewis acid accepts a pair of electrons from a Lewis base. All Bronsted–Lowry acids and bases are also Lewis acids and bases, but the Lewis definition encompasses more reactions. In the neutralization reaction of H$^+$ with OH$^-$ to give water, a lone pair of electrons on the hydroxyl is donated to hydrogen; thus, this reaction fits both acid and base definitions. When an orthophosphate ion reacts with Fe(III) in the structure of goethite, the Fe(III) accepts an electron pair from the oxygen

on the phosphate group to form the bond. This fits the Lewis definition of an acid–base reaction but is clearly not a Bronsted–Lowry acid–base pair.

12.5.2 Calculations for Acid–Base Equilibria

Graphical representations of acid–base equilibria can take several forms, but the underlying calculations are built on the same theoretical foundation. Whether the system is simple or complex, the approach is the same, although organizing and executing the computations can be challenging for large systems. The first step is to identify the participating components, assemble the pertinent equations, identify the master variables, and solve the equations in terms of the master variables. A simple system, aqueous solution of carbonate, will be used as an example with more complex systems developed later in this chapter.

In this system, the species in solution would be OH^-, H^+, $H_2CO_3^o$, HCO_3^-, and CO_3^{2-} with $CO_2(g)$ in the gas phase. The defining reactions and equilibrium constants are given in Table 12.7. The equilibrium constants (K^o) are given at zero ionic strength, 298.15 K, and 101.33 kPa (1 atm) and are taken from Lindsay (1979). By defining the equilibrium constants at zero ionic strength, it is assumed that all species are given in terms of activities. In Figure 12.5, activities of the three solution species in equilibrium with 0.0003 atm $CO_2(g)$ are plotted as a function of pH. The predominant solution species below pH 6.33 is $H_2CO_3^o$ at an activity of $10^{-4.98}$. Between pH 6.33

TABLE 12.7   Controlling Equations for Equilibria Involving $CO_2(g)$ and $H_2O(l)$

Equation	$\log K^o$
$CO_2(g) + H_2O(l) \leftrightarrow H_2CO_3^o$ (aq)	−1.46
$H_2CO_3^o \leftrightarrow HCO_3^- + H^+$	−6.36
$H_2CO_3^o \leftrightarrow CO_3^{2-} + 2H^+$	−16.69

Source: Lindsay, W.L., *Chemical Equilibria in Soils*, Wiley Interscience, New York, 1979.
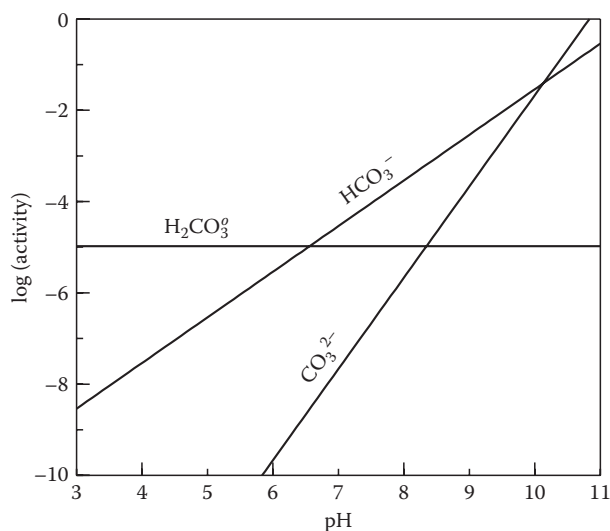
FIGURE 12.5 Activities of carbonate species as affected by pH assuming equilibrium with P_{CO_2} = 0.0003 atm.
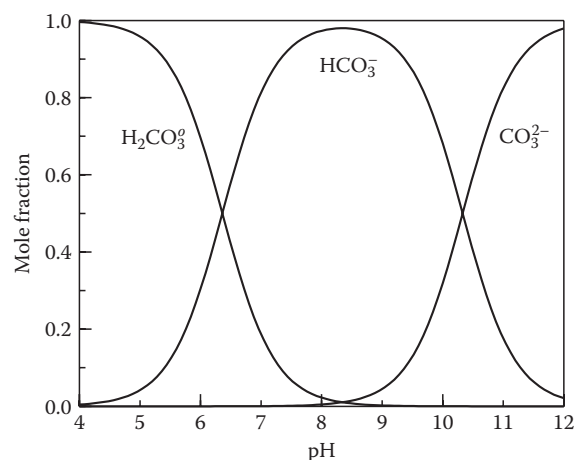
FIGURE 12.6   Mole fraction distribution of carbonate species as a function of pH. For all calculations, activity coefficients are assumed to be unity (γ_i = 1.0).

and 10.36, the dominant carbonate species is HCO_3^- with CO_3^{2-} being present at the greatest activities above pH 10.36. However, the total carbonate in solution approaches molar quantities near pH 10 in equilibrium with carbon dioxide at this pressure, concentrations rarely seen in natural solutions. Thus, the kinetics of $CO_2(g)$ dissolution probably predominate over equilibrium predictions at these high pH values in open systems.

Information similar to Figure 12.5 can be provided without the overlying assumption of an open system or predicting equilibrium activities. Figure 12.6 is a mole fraction distribution diagram depicting the relative concentrations of the three carbonate species as a function of pH. The mole fraction is defined as the ratio of the concentrations of a given species to the sum of the concentrations of all the carbonate species. For CO_3^{2-},

$$\text{Mole fraction(carbonate)} = \frac{[CO_3^{2-}]}{[H_2CO_3^o] + [HCO_3^-] + [CO_3^{2-}]}.$$

(12.60)

The equilibrium expressions for these species in Table 12.4 are substituted appropriately and solved across pH (Lindsay, 1979). The resulting diagram (assuming that γ_i = 1) is applicable to either open or closed systems and requires equilibrium only among the solution species; equilibrium between the aqueous and gaseous phases is not required. As Figures 12.5 and 12.6 illustrates that bicarbonate ion is dominant between pH 6.36 and 10.33 with carbonic acid dominant below pH 6.36 and carbonate above pH 10.33.

In the examples above, the simple system of carbon dioxide in water was considered. Multiple component systems or other complexities can be handled in the same fashion.

12.5.3 Buffering Capacity of Soil Solutions

Because of the many acids and bases present in the soil and dissolved in the soil solution, the acid–base chemistry of soils is
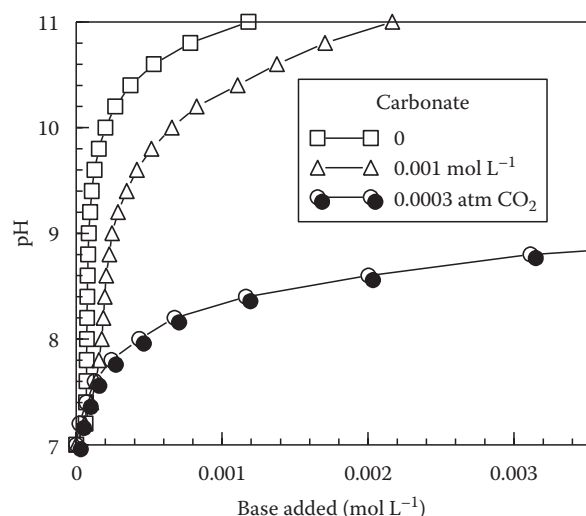
FIGURE 12.7 Theoretical titration curves of an aqueous solution containing pure water ($\square$) in a closed system (no access to atmospheric $CO_2(g)$), 0.001 M total carbonate in a closed system ($\triangle$), or an open system in equilibrium with atmospheric $CO_2(g)$ at 30 Pa (0.0003 atm) (●). In all systems, the assumption is made that $\gamma_i = 1$ and the volume remains constant.

highly complicated. One of the results of this complex system is that the pH of soils and, to a smaller extent, soil solutions does not change greatly in response to inputs of acid or base. The ability of a system to resist pH changes is called the buffering capacity or buffering intensity and can be determined experimentally by titrating the solution and measuring the pH at each increment of acid or base addition. A simple, qualitative illustration of the buffering power of carbonate in water is illustrated in Figure 12.7 in which aqueous solutions are theoretically titrated with base from pH 7 to 11. In the first case, only pure water is titrated, and the solution shows little capacity to resist pH change; with each increment of base, a large pH change is noted. In the second case, a solution containing 0.001 M total carbonate is initially adjusted to pH 7 and titrated with NaOH to pH 11. In this theoretical titration, activity coefficients are ignored ($\gamma_i = 1$), but the volume is assumed to remain constant. The carbonate solution has a significant resistance to pH change (buffering capacity) compared with the pure solution. The third case in Figure 12.7 is an open system in equilibrium with atmospheric $CO_2(g)$ assuming 0.0003 atm (30 Pa). The buffering capacity of this solution is low when pH < 8, but at greater pH values, the dissolution of CO_2 into the solution radically increases the buffering.

Buffering may be quantified mathematically by defining buffering as the change in pH induced by the addition of acid or base to a solution:

$$\text{Buffering} = \frac{\Delta C_{\text{base}}}{\Delta \text{pH}} = -\frac{\Delta C_{\text{acid}}}{\Delta \text{pH}}. \tag{12.61}$$

Thus, the buffering is mathematically defined as the change in pH induced by an increment of acid or base (in mol L^{-1}) added to the system. For a monoprotic acid (HA) dissolved in water,

the buffering can be shown to be represented as (Stumm and Morgan, 1996)

$$\text{Buffering} = 2.3\left([H^+] + [OH^-] + \frac{[HA][A^-]}{[HA] + [A^-]}\right). \tag{12.62}$$

This equation defines the buffering of the system at any point in the titration of the monoprotic acid, HA. Similar expressions can be derived for mixtures of acids or for polyprotic acids.

12.5.4 Soluble Organic Acids

The chemistry of soluble organic acids in soil solutions has not been studied to the extent of inorganic acids, although detailed reviews exist (e.g., Stevenson, 1967). The relative lack of information is not because of the lack of importance of soluble organic in soils but is a reflection of the difficulty in quantitatively characterizing the organic components in solution and the dynamic nature of soil organic molecules. Many simple organic acids have been identified in solution ranging from formic and acetic acids to more complex aromatic acids such as catechin (McKeague et al., 1986). The aliphatic carboxylic acids are degraded very rapidly, but the aromatic acids tend to be more persistent.

For organic acids that can be identified, the theoretical approach to defining their chemistry is identical to that of inorganic acids. Extensive lists of acidity constants have been compiled (Martell and Smith, 2004), and these constants may be used to generate activity and mole fraction diagrams similar to Figures 12.5 and 12.6. The concepts of buffering intensity and acid neutralizing capacity of organic acids are applied in identical fashion as those of inorganic acids.

Methods also exist for quantifying the acid–base behavior of more complex organic mixtures. A direct approach is to titrate the soil solution or extract, plot the titration curve, and identify the characteristic buffering regions (Dudley and McNeal, 1987; Sposito, 2008). The resulting curves can be modeled based on hypothetical mixtures of organic acid groups, or the acid neutralizing capacity and formation functions may be determined. There is general agreement that complex organic matter in soil and in soil solution is a broad mixture of carboxylic and benzoic acids with a wide range of acidity constants.

12.6 Formation of Soluble Complexes

12.6.1 Types of Complexes

A solution complex is the close association between a central molecular component (such as a cation) with other atoms or molecules. Very often, metals or other positively charged species act as the central component attracting neutral or negatively charged ligands. In the hydrolysis of Fe^{3+} to form $Fe(H_2O)_5OH^{2+}$, the Fe^{3+} cation is acting as the central unit with water and hydroxyl acting as the ligands. This is a special case of complex formation called a solvation complex. Two other categories of complexes can be formed depending upon the strength of

bonding between the central unit and the ligand. If the inter-action between the cation and ligand is strong enough that the ligand displaces the solvation sphere, the resulting association is termed an inner-sphere complex. Typical examples of inner-sphere complexes are AlF_n^{3-n} and Fe(III)citrate. If the attraction between the cation and ligands are not strong enough to displace the hydration layer, an outer-sphere complex is formed or, sometimes, an ion pair. An inner-sphere complex can form if the heat evolved during the association exceeds the energy needed to displace the hydration sphere. If the heat evolved is less than the energy necessary to displace the water, then the hydration sheath remains intact and an outer-sphere complex results.

If a ligand occupies more than one coordination site in the complex, the ligand is referred to as multidentate. For example, two of the oxygens in oxalate can bond simultaneously with a central Fe(III) ion. Such a complex is called a chelate, and the process is chelation. The multidentate nature of the interaction can significantly increase the strength of the bond, and the resulting formation constant is much higher than ordinary complexes.

12.6.2 Hard and Soft Acid–Base Rules

Other useful concepts in predicting the association between cations and ligands are the hard and soft acid–base rules (Pearson, 1963). Hard acids and bases are those species that tend to have smaller radii and are not readily deformable (Table 12.8). Soft acids and bases are often larger and more polarizable (have a more easily deformed electron sheath). The tendency is that hard acids bond preferentially with hard bases, and soft acids bond preferentially with soft bases. Thus, Ca^{2+} would tend to bond with phosphate and carbonate rather than chloride and sulfide.

12.6.3 Rates of Formation of Solution Complexes

The rates of formation of solution complexes are often rapid, establishing equilibrium very quickly. Certain reactions, however, proceed very slowly. For example, the reaction

$$Al^{3+} + F^- \leftrightarrow AlF^{2+} \qquad (12.63)$$

requires nearly 20 min to proceed half way to completion. This is in contrast to the formation of $MnSO_4^o$, which requires 10^{-5} s.

TABLE 12.8 Hard and Soft Acids and Bases

	Hard	Intermediate	Soft
Bases	F^-, CO_3^{2-}, Cl^-, OH^-, CH_3COO^-, PO_4^{3-}, SO_4^{2-}, NH_3, $R–NH_2$, H_2O, $R–OH$, NO_3^-	SO_3^{2-}, NO_2^-, $C_6C_5NH_2$, Br^-	S^{2-}, CN^-, I^-, $R–SH$, SCN^-, $S_2O_3^{2-}$
Acids	H^+, Li^+, Na^+, K^+, Mg^{2+}, Ca^{2+}, Sr^{2+}, Al^{3+}, La^{3+}, Si^{4+}, Zr^{4+}, Th^{4+}, Th^{4+}, Cr^{3+}, Mn^{3+}, Fe^{3+}	Mn^{2+}, Fe^{2+}, Cu^{2+}, Zn^{2+}, Pb^{2+}, Bi^{3+}, Ni^{2+}, SO_2	Ag^+, Cu^+, Cd^{2+}, Hg^{2+}, Cs^+

Source: Pearson, R.G. 1963. Hard and soft acids and bases. J. Am. Chem. Soc. 85:3533.

Hydrolysis reactions generally are quite rapid for the monovalent and divalent cations but can proceed slowly for the higher charged cations. The rate of these reactions can be approximated by examining the rate of water exchange from a hydrated cation, $Me(H_2O)_x^{y+}$ (Stumm and Morgan, 1996):

$$Me(H_2O)_x^{y+} + H_2O^* \leftrightarrow Me(H_2O)_x^{y+} H_2O^*, \qquad (12.64)$$

where H_2O^* is the water being exchanged into the solvation complex. The forward rate constant in this expression is k_1, and the rate constant for the reverse reaction is k_{-1}. The exchange reaction is completed in this equation with a rate constant of k_{-w}:

$$Me(H_2O)_x^{y+} H_2O^* \rightarrow Me(H_2O)_{x-1}(H_2O^*)^{y+}. \qquad (12.65)$$
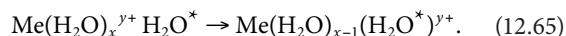
The rate of this reaction is

$$\frac{d\left[Me(H_2O)_{x-1}(H_2O^*)^{y+}\right]}{dt} = k_{-w}\left[Me(H_2O)_x^{y+}(H_2O)^*\right]. \qquad (12.66)$$

At steady state,

$$\frac{d\left[Me(H_2O)_x^{y+}(H_2O^*)\right]}{dt}$$

$$= k_1[Me(H_2O)_x^{y+}][H_2O^*] - (k_{-1} + k_{-w})\left[Me(H_2O)_x^{y+} H_2O^*\right] = 0. \qquad (12.67)$$

After further manipulation and assuming that $k_{-1} \gg k_{-w}$,

$$\frac{d\left[Me(H_2O)_{x-1}(H_2O^*)^{y+}\right]}{dt} = k_{-w}K_{OS}\left[Me(H_2O)_x^{y+}\right][H_2O^*], \qquad (12.68)$$

where $K_{OS} = k_1/k_{-1}$ is the outer-sphere complex formation equilibrium constant. Thus, the rate constant for the exchange of water in the hydration shell is estimated by $k_{-w}K_{OS}$. The same development can be used to describe the kinetics of complex formation with a ligand, L, rather than H_2O^*. The resulting equation is

$$\frac{d\left[Me(H_2O)_{x-1}L^{y+}\right]}{dt} = k_{-w}K_{OS}\left[Me(H_2O)_x^{y+}\right][L]. \qquad (12.69)$$

Omitting the waters of hydration results in the standard notation:

$$\frac{d[MeL]}{dt} = k[Me][L], \qquad (12.70)$$

where $k = k_{-w}K_{OS}$. This development illustrates the importance of the equilibrium constant for the outer-sphere complex (i.e., the energetics of the reaction) as well as the kinetics of exchange of waters of hydration surrounding the metal.

Stumm and Morgan (1996) compiled the values of k_{-w} for water exchange reactions for several metals. For Pb^{2+}, Hg^{2+}, Cu^{2+}, Ca^{2+}, Cd^{2+}, La^{3+}, Zn^{2+}, Mn^{2+}, Co^{2+}, $Fe(OH)_2^+$, and $Fe(OH)_4^-$, the values were $k_{-w} \geq 10^6$ s^{-1}. Values of k_{-w} for other metals were 2×10^2 for Fe^{3+}, 1.0 for Al^{3+}, and 5×10^{-7} for Cr^{3+}.

Sequential reactions can be "coupled" and solved. Exact analytical solutions to these models can be obtained (Sposito, 1994), but it is typical to use experimental observations to help establish simplifying assumptions.

The formation of the solution complex $CdCl^+$ in the presence of 10^{-3} mol L^{-1} Cl can be used to illustrate the application of these equations:

$$Cd(H_2O)_x^{2+} + Cl^- \leftrightarrow Cd(H_2O)_{x-1}Cl^+ + H_2O. \quad (12.71)$$
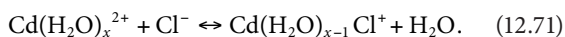
The rate of reaction can be given by

$$\frac{d\left[Cd(H_2O)_{x-1}Cl^+\right]}{dt} = K_{OS}k_{-w}\left[Cd(H_2O)_x^{2+}\right][Cl^-]$$

$$= -\frac{d\left[Cd(H_2O)_x^{2+}\right]}{dt}, \quad (12.72)$$

where

$k_{-w} = 10^{8.48}$ (s^{-1})
$K_{OS} = 10^{1.98}$ (mol L)$^{-1}$
$(k_{-w})(K_{OS}) = 10^{10.46}$ (L mol^{-1} s^{-1})

Substituting these values into the above equation and simplifying by not expressing the waters of hydration,

$$-\frac{d[Cd^{2+}]}{dt} = (10^{10.46})(10^{-3})[Cd^{2+}] = 10^{7.46}[Cd^{2+}]. \quad (12.73)$$

For a first-order reaction such as this, the half-life of the reaction (i.e., the time required for the reaction to proceed half way to completion) is equal to (ln 2) k^{-1}. Thus, the half-life for this reaction is $0.693/2.88 \times 10^7$ or 2.40×10^{-8} s. Sposito (1989) tabulated the half-lives for several solution reactions, and the values ranged from 10^{-9} s for the formation of $MnSO_4^o$ to 10^3 s for the formation of AlF^{2+}.

The relationships presented in this section demonstrate that most reactions in the soil solution are rapid enough that equilibrium can be achieved during the course of most experiments. Even the slowest example given above, the formation of AlF^{2+} in 10^3 s, will be nearly complete in less than 1 h. However, the rate of certain oxidation/reduction reactions can proceed very slowly, particularly reactions that involve oxyanions (such as arsenates or chromate) or that are not microbially catalyzed. These reactions can take days to reach completion, if at all. Therefore, knowledge of equilibrium predictions alone may be of limited utility if the kinetics of reaction are unfavorable.

12.7 Application of Thermodynamic and Equilibrium Concepts to Soil Solutions

The chemical composition and dynamics of the soil solution are reflections of all the processes, which depend upon the aqueous phase: biological activity, mineral dissolution/precipitation, adsorption/desorption, physical transport, and anthropogenic inputs. Wolt (1994) described the soil solution as "a window to chemically reacting soil systems where the intensity and distribution of chemicals in the soil aqueous phase represents the integration of multiple physical, chemical, and biological processes occurring concurrently within the soil environment." The information provided by the chemical composition is further enhanced by the knowledge of time trends and the application of kinetic and thermodynamic theories.

12.7.1 Speciation and Single-Ion Activity Determinations

The determination of free ion activities (or concentrations) for soil solution components provides a powerful interpretative tool for chemical reactions in soils. A limited number of ion activities can be determined directly (such as the H^+ activity using a glass electrode), but the rest must be approximated through rigorous calculations.

The steps in calculating ion activities and the distribution of solution species include obtaining and analyzing the soil solution, identifying the important solution species that can form, assembling relevant equilibrium constants and reactions for the formation of these species, obtaining an analytical or numerical solution for the series of equations, and calculating activity coefficients. All the steps in the calculations can be handled by any of the many geochemical models currently available (discussed later); however, understanding the chemical concepts behind the models and how the activities are determined will be helpful in properly executing the computer models and obtaining the best possible data. The following sections provide a brief review of the important chemical reactions and activity calculations.

12.7.1.1 Hydrolysis, Complexation, and Oxidation/Reduction Reactions

To help illustrate the steps involved in activity calculations, soluble Fe will be used as an example. The first step is to determine the important reactions in solution, and this can be done with the aid of some of the compilations of thermodynamic data for aqueous systems, such as Garrels and Christ (1965), Lindsay (1979), Sadiq and Lindsay (1979), and Martell and Smith (2004). Many of the metals undergo hydrolysis reactions, and Fe(III) and Fe(II) have many important hydrolysis species as indicated in Table 12.9. The distribution of these Fe(III) and Fe(II) hydrolysis species as a function of pH (Figure 12.8) shows that the predominant species will be strongly dependent upon pH with mole fractions ranging from 0.0 to over 0.9. Examination of Table 12.9 indicates that the complexes listed must be considered if $(Cl^-) > 10^{-2}$, $(F^-) > 10^{-7}$,

TABLE 12.9 Reactions and Equilibrium Constants for Selected Solution Complexes of Iron

Reaction	$\log K^o$
Fe(III) hydrolysis	
$Fe^{3+} + H_2O \leftrightarrow FeOH^{2+} + H^+$	−2.19
$Fe^{3+} + 2H_2O \leftrightarrow Fe(OH)_2^+ + 2H^+$	−5.69
$Fe^{3+} + 3H_2O \leftrightarrow Fe(OH)_3^o + 3H^+$	−13.09
$Fe^{3+} + 4H_2O \leftrightarrow Fe(OH)_4^- + 4H^+$	−21.59
Fe(II) hydrolysis	
$Fe^{2+} + H_2O \leftrightarrow FeOH^+ + H^+$	−6.74
$Fe^{2+} + 2H_2O \leftrightarrow Fe(OH)_2^o + 2H^+$	−16.04
Redox	
$Fe^{3+} + e^- \leftrightarrow Fe^{2+}$	13.04
Complexation	
$Fe^{3+} + Cl^- \leftrightarrow FeCl^{2+}$	1.48
$Fe^{3+} + F^- \leftrightarrow FeF^{2+}$	6.00
$Fe^{3+} + SO_4^{2-} \leftrightarrow FeSO_4^{2-}$	4.15
$Fe^{3+} + H_2PO_4^- \leftrightarrow FeHPO_4^- + H^+$	3.71
$Fe^{2+} + SO_4^{2-} \leftrightarrow FeSO_4^o$	2.20

Source: Lindsay, W.L. 1979. Chemical equilibria in soils. Wiley Interscience, New York.
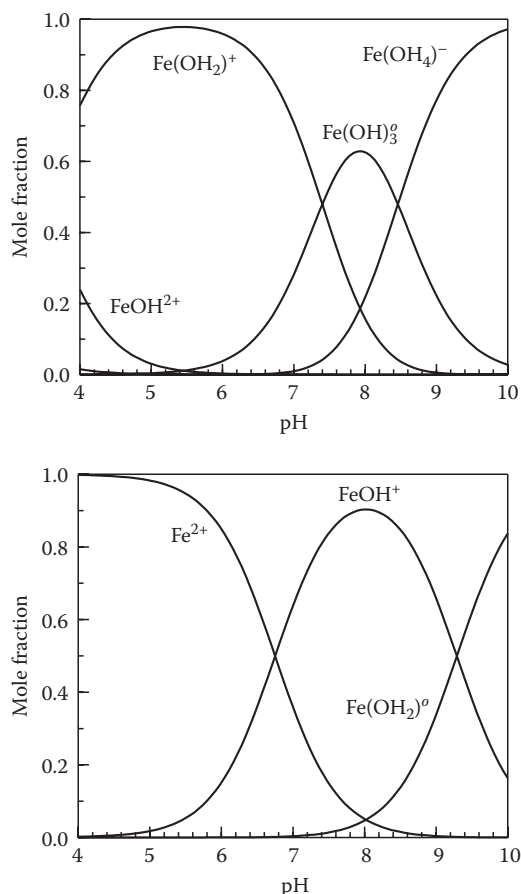
FIGURE 12.8 Distribution of Fe(III) and Fe(II) hydrolysis species in solution as a function of pH.

$(SO_4^{2-}) > 10^{-5}$ mol L^{-1}. The constants in Table 12.9 and their corresponding equations can be used to solve for ionic activities.

12.7.1.2 Solution Composition Example

Consider a soil solution with the following composition: $0.2\,\mu$mol L^{-1} total Fe, 20 mmol L^{-1} total Cl, 10 μmol L^{-1} total F, 1 mmol L^{-1} total sulfate, ionic strength 0.01 mol L^{-1}, pH 6.5, and a redox potential of 600 mV. This is enough information to estimate the Fe^{2+} and Fe^{3+} activities. The solution is fairly complex and requires a few assumptions to make it solvable. The first assumption will be that the ligands will not form complexes with any cations in solution of other Fe^{2+} and Fe^{3+}. This is a poor assumption because each ligand forms a number of complexes with other cations, particularly Al^{3+}, Ca^{2+}, and Mg^{2+}. The second assumption is that the only species of Fe to be considered are those given in Table 12.9.

The approach given here is to identify all potential variables, establish the same number of independent equations as the number of variables, and solve the equations simultaneously. The system variables are the activities of H^+, OH^-, e^-, Fe^{3+}, Fe^{2+}, Cl^-, SO_4^{2-}, F^-, $FeOH^{2+}$, $Fe(OH)_2^+$, $Fe(OH)_3^o$, $Fe(OH)_4^-$, $FeOH^+$, $Fe(OH)_2^o$, $FeCl^{2+}$, FeF^{2+}, $FeSO_4^+$, and $FeSO_4^o$, for a total of 18 variables. The relevant equations include 11 complexation equations from Table 12.9 (all equations except for the formation of $FeHPO_4^+$), K_w (water dissociation equation), pH = 6.5, Eh = 600 mV (pe = 10.1), and the four mass balance equations for the components:

$$\sum Fe = [Fe^{3+}] + [Fe^{2+}] + [FeOH^{2+}] + [Fe(OH)_2^+] + [Fe(OH)_3^o]$$
$$+ [FeOH^+] + [Fe(OH)_2^o] + [FeCl^{2+}]$$
$$+ [FeF^{2+}] + [FeSO_4^+] + [FeSO_4^o], \tag{12.74}$$

$$\sum SO_4^{2-} = [SO_4^{2-}] + [FeSO_4^+] + [FeSO_4^o], \tag{12.75}$$

$$\sum Cl = [Cl^-] + [FeCl^{2+}], \tag{12.76}$$

$$\sum F = [F^-] + [FeF^{2+}]. \tag{12.77}$$

In the mass balance expressions, brackets [x] represent concentrations of species x. Activity coefficients must be calculated for each species to tie together the mass balance equations with the formation equations. Thus, there are a total of 18 equations for the 18 variables. Many methods exist for solving this system of equations including back substitution and matrix algebra. These systems can be solved by hand, but solutions are reached more rapidly using computers.

12.7.2 Geochemical Models

Over the past decades, scientists and research groups have recognized the need for computer programs to solve chemical

equilibrium systems, and several computer models have been established. Models vary widely in their construction and composition as well as their intended use (Baham, 1984; Melchior and Bassett, 1990; van der Lee and de Windt, 2001; Hummel, 2005). Many of the models are based upon single-ion activities, single-ion activity coefficients, and formation constants; Pitzer's equations for applications at high ionic strength; or a free energy minimization approach.

12.7.2.1 Formation Constant Approach

In Section 12.7.1.2, an example was given outlining the formation constant approach to determine single-ion activities in soil solutions. Several series of published models take this same approach. The WATEQ series (Truesdall and Jones, 1974; Ball et al., 1979) was developed by the U.S. Geological Survey to compute single-ion activities to predict the fate of critical elements in geochemical environments. Emphasis was placed on obtaining a well-documented thermodynamic database with the best available equilibrium constants; a reaction would be kept out of the database if the constant was estimated or suspected to be of poor quality. The mathematical solution method was back substitution and required recompilation of the programming code if new reactions were added or if constants were changed. The MINEQL series (Westall et al., 1976) was developed using the mathematical approach of REDEQL (Morel and Morgan, 1972) to solve equilibrium problems and provide information about trends in observed data. The mathematical solution method is rapid and efficient, but the thermodynamic database was not rigorously reviewed. The model SOILCHEM (Sposito and Coves, 1988) was an extension of the MINEQL model with emphasis on expanding the database. Some smaller models have been developed, such as SOILSOLN (Wolt, 1989) and CALPHOS (Adams, 1971), for very specific (but limited) applications. SOILSOLN was designed as a teaching tool with an easy-to-use interface and embedded graphics.

12.7.2.2 High Ionic Strength Models

The general approach in the models described in Section 12.7.2.1 is the combination of equilibrium constant expressions with mass balances. The impact of ionic strength (and activity coefficients) is disregarded or single-ion activity coefficients are calculated. The single-ion activity coefficient is not rigorously defined in thermodynamic theory, and the equations used to approximate activity coefficients are subject to error, particularly at high ionic strength. A special approach must be taken for aqueous solutions with high-salt concentrations (ionic strength greater than 0.5 mol L^{-1}). One such approach is to use the Pitzer equations (Pitzer, 1979) for calculation of single-ion activity coefficients, and at least two published models currently use this approach. Felmy and Weare (1995) discussed their model and experimental results of simple systems with high ionic strengths. The agreement between theory and experimental determinations is promising. The C-SALT model (Smith et al., 1995) also uses the Pitzer equations. The model has been tested against experimental data.

The Pitzer equations require a relatively large number of parameters that are specific to the ions and systems under consideration. The number of required parameters increases rapidly as the systems become more complex. The Pitzer, Davies, and Debye–Hückel equations calculate single-ion activity coefficients. Unfortunately, these activity coefficients cannot be defined unambiguously because they are dependent upon a clear, thermodynamic definition of single-ion concentrations. As stated by Sposito (1994) when discussing the activity coefficients for a metal ion, γ_M, and a ligand, γ_L, "For γ_M and γ_L to have chemical significance, the species molalities, m_M and m_L, must have a well-defined operational meaning. Thus, the single-ion activity coefficient has no meaning apart from the set of operational procedures used to define ionic species and to determine their concentrations in an aqueous solution." Although the determination of single-ion activities can provide useful information, single-ion activity coefficients cannot be experimentally determined and do not have a thermodynamic foundation.

12.7.2.3 Solving Example Problem with Visual MINTEQ

The geochemical model, MINTEQ, and later MINTEQA2 (Allison et al., 1990), were developed by merging the mathematics and computer code from MINEQL with the database from the WATEQ series. Although the MINTEQ database is not as extensive as SOILCHEM's, all values were fully documented as to their source and reason for selection. Visual MINTEQ was developed to provide a graphic interface for MINTEQA2 version 4 (Gustafsson, 2009) as well as expanding the database. Visual MINTEQ provides solution speciation, mass transfer through precipitation and dissolution of solid phases, ion adsorption, ion exchange, redox, and humic complexation models, and gas phase calculations.

The hypothetical analytical data for the example problem in Section 12.7.1.2 were entered into a Visual MINTEQ input file, and the program executed. The resulting activities (Table 12.10)

TABLE 12.10   Activities of Selected Aqueous Species Calculated Using Visual MINTEQ Version 2.61 for the Example Solution Composition Given in Section 12.7.1.2

Species	Log (Activity)	Species	Log (Activity)
Fe^{3+}	−14.00	Fe^{2+}	−11.11
$Fe(OH)^{2+}$	−9.52	$FeOH^+$	−14.00
$Fe(OH)_2^+$	−6.75	$Fe(OH)_2^o$	−18.60
$Fe(OH)_3^o$	−9.50	$FeSO_4^o$	−11.89
$Fe(OH)_4^-$	−10.70	FeF^+	−14.94
FeF^{2+}	−13.00	$FeCl^+$	−13.05
FeF_2^+	−13.43	F^-	−5.05
FeF_3^o	−15.43	SO_4^{2-}	−3.18
$FeSO_4^+$	−12.92	Cl^-	−1.74
$FeCl^{2+}$	−14.26		

Source: Gustafsson, J.P. 2009. Visual MINTEQ version 2.61. http://www.lwr.kth.se/English/OurSoftware/vminteq/, accessed on May 16, 2011.

show that the hydrolysis species of Fe(III) and Fe(II) are much more important than the chloro, fluoro, and sulfate species. This is because these ligands react only with the unhydrolyzed cations, and, in the case of Fe(III), the activity of Fe^{3+} is very low. In the case of Fe(II), the ligand concentrations simply are not high enough to become the predominant solution species. (It is to be noted that the program was executed with the specification that supersaturated solid phases would not be allowed to precipitate. Many solid phases were supersaturated, and allowing their precipitation would have changed the results.)

The comparison of the Fe(III) and Fe(II) species also is worth noting. In Table 12.9, the constant for the equilibrium between Fe^{3+} and Fe^{2+} suggests that Fe^{2+} will be present are higher activities than Fe^{3+} when the pe < 13 (770 mV). In this case, Fe^{2+} is nearly 1000 times greater than Fe^{3+}, but the Fe(III) hydrolysis species are by far the predominant Fe solution species. At this pH, the oxidation/reduction potential would have to be pe < 6.0 (355 mV) before the predominant Fe(II) species is present at a greater activity than the predominant Fe(III) species.

12.7.3 Oxidation/Reduction Reactions

Chemical reactions involving oxidation and reduction are important in all biological systems, and soils are no exception. In all soils, localized areas can be found that are highly oxidized and highly reduced. Areas of reduction generally are associated with small microsites that are saturated with water and have restricted $O_2(g)$ diffusion. Microbial activity within these sites rapidly depletes the available oxygen, and the redox potential of the soil solution decreases. The ability to predict the impact of changes in redox on soil solution composition is a useful tool.

The first step in handling oxidation/reduction equilibria in soil solution is defining the limits of stability of aqueous systems. If a solution is subject to a high oxidizing potential, then the water spontaneously can be degraded into oxygen:

$$0.5H_2O(l) = H^+ + e^- + 0.25O_2(g) \qquad (12.78)$$
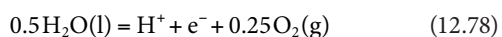
with an equilibrium constant expression:

$$K^o = \frac{(H^+)(e^-)\left(O_2(g)\right)^{0.25}}{(H_2O)^{0.5}} = 10^{-20.78}. \qquad (12.79)$$

Substituting, rearranging, and solving for pe in terms of pH,

$$pe = 20.78 + 0.25\log(O_2) - pH. \qquad (12.80)$$

The pe at which water spontaneously decomposes to oxygen gas would correspond to $O_2(g)$ pressure of 1 atm or pe = 20.78 − pH. Using the combined parameter, pe + pH, the upper oxidizing limit of the stability field of water would be

$$pe + pH = 20.78. \qquad (12.81)$$

A strongly reducing potential can generate hydrogen gas:

$$H^+ + e^- \leftrightarrow 0.5H_2(g). \qquad (12.82)$$
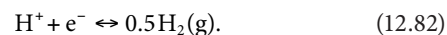
Again solving for pe,

$$pe = -0.5\log\left(H_2(g)\right) - pH. \qquad (12.83)$$

At 1 atm $H_2(g)$, this reduces to pe = −pH. Again converting to the combined parameter, pe + pH,

$$pe + pH = 0. \qquad (12.84)$$

In a plot of pe versus pH, the region between the lines pe = −pH and pe = 20.78 − pH represents the stability field for water. Water is thermodynamically unstable at potentials outside this region. Another important aspect of these equations is the equilibrium partial pressures of $H_2(g)$ and $O_2(g)$ at typical redox potentials in soil solutions. For example, highly reduced soils seldom have redox potentials below pe + pH = 4, corresponding to $P_{H_2} = 10^{-8}$ atm. Measured redox potentials in fully oxidized soils seldom exceed pe + pH 18, and equilibrium at this potential would require $10^{-11.23}$ atm $O_2(g)$. If equilibrium were rapidly established with these gases, then the partial pressures of hydrogen and oxygen would be expected to be quite small in all situations. An equilibrium approach can be used to investigate many redox reactions in soil solutions. The Fe(III)/Fe(II) system was discussed in a prior section, and many other redox active species exist in soil solutions including Cu^{2+}, Mn^{2+}, and CrO_4^{2-}. Chromium is an environmentally important metal and has an interesting redox behavior. The Cr(III) species is considered environmentally less hazardous than the Cr(VI), which is closely regulated. Published equilibrium constants can be used to determine the redox potential at pH 7 at which the predominant Cr(III) species, $Cr(OH)_2^+$, converts to the predominant Cr(VI) species, CrO_4^{2-}. From the MINTEQ database, the following reaction and constant were obtained:

$$CrO_4^{2-} + 6H^+ + 3e^- \Leftrightarrow Cr(OH)_2^+ + 2H_2O \quad \log K^o = 67.38. \qquad (12.85)$$
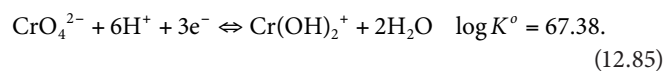
The expression for the equilibrium constant, converted to logarithms, would be

$$\log\left[\frac{\left(Cr(OH)_2^+\right)}{(CrO_4^{2-})}\right] + 6\,pH + 3\,pe = 67.38. \qquad (12.86)$$

The redox potential at which Cr(III) converts to Cr(VI) occurs when the predominant species of each oxidation state of Cr have equal activities, $(Cr(OH)_2^+) = (CrO_4^{2-})$. The ratio of these activities would equal unity, the logarithm of which is 0. Therefore, after dividing by 3,

$$2\,pH + pe = 22.46 \qquad (12.87)$$

or, solving for pe + pH,

$$pe + pH = 22.46 - pH \qquad (12.88)$$

and, at pH = 7,

$$pe + pH = 15.46. \qquad (12.89)$$

This is a moderately reducing condition for soil solutions, roughly the same redox potential at which denitrification occurs (Lindsay, 1979).

12.7.4 Successful Applications of Geochemical Modeling to Soil Solutions

The application of equilibrium thermodynamics to soil solutions will have several requirements for success, and the definition of success will depend upon individual interpretations. The kinetics of some reactions in soil solutions were discussed in Section 12.6.3; most reactions were found to proceed quickly, but others are quite slow. Therefore, care must be taken to apply equilibrium thermodynamics only to the systems having enough time to equilibrate, and this is dependent upon the method of sample collection. For example, miscible displacement methods often have equilibration periods of less than 24 h, which is not long enough for some solid-phase equilibria or redox reactions involving oxyanions. However, most of the field methods should be adequate.

The success of applying a model to a chemical system will increase as more information is known. One of the important areas of progress in geochemical modeling in recent years has been the addition of aluminum–phosphate speciation and empirical models for evaluating the impact of soluble organic matter.

A critical aspect of employing a geochemical model is experimental validation. Models have been validated by published laboratory tests followed by extending the model to similar systems in natural settings. Felmy and Weare (1995) incorporated Pitzer's equations into an ion-interaction model, the purpose of which was to allow chemical speciation in high ionic strength solutions. The Davies equation for activity coefficients was found to inadequately predict $\gamma_\pm$ for many salts at ionic strengths as low as 0.2 mol L^{-1} for 1:1 electrolytes and 0.05 mol L^{-1} for 2:1 electrolytes (Figure 12.9). The parameters for the ion-interaction model were obtained experimentally and applied to several systems, including $Na_2B_4O_7$–Na_2SO_4–H_2O, $Na_2B_4O_7$–Na_2CO_3–H_2O, and $NaBO_2$–$NaCl$–H_2O systems. Salt concentrations were varied, and other activities measured. The agreement between modeling predictions and experimental results were ±10%, even in the $NaBO_2$–$NaCl$–H_2O system (Figure 12.10) with very high borate concentrations and ionic strengths as high as 14 mol L^{-1}.

The effect of redox on the solubility of Fe in soil and aqueous systems has been a subject of study for decades. The presence of mixed valence state oxides has been observed and implied, and extensive equilibrium studies have been published. One of the
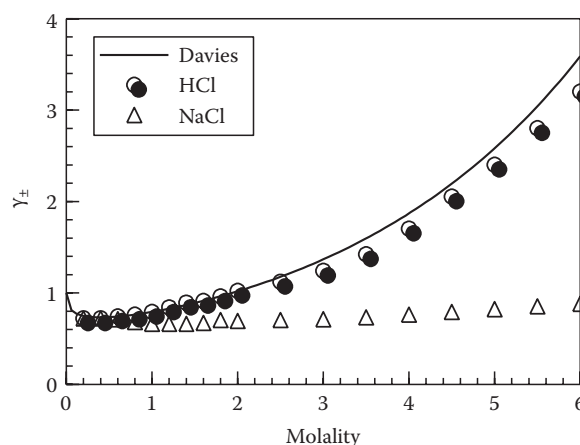
FIGURE 12.9 Changes in experimentally determined mean activity coefficients ($\gamma_\pm$) for various electrolytes compared with those generated by the Davies equation. Measured values are symbols, and predicted values are lines. (From Felmy, A.R., and J.H. Weare. 1995. The development and application of aqueous thermodynamic models: The specific ion-interaction approach. *In* R. Loeppert et al. (eds.) Chemical equilibrium and reaction models. Soil Science Society of America Special Publications No. 42. ASA, Madison, WI.)
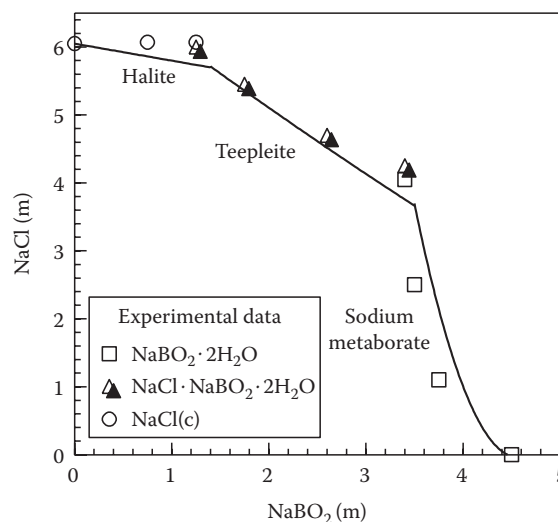
FIGURE 12.10 The system $NaBO_2$–$NaCl$–H_2O at 20°C. Experimental data of Skvortsov et al. (1976) are shown as symbols, and the model predictions of Felmy and Weare (1995) are represented as lines.

more thorough studies involved the controlled reduction of ferric oxides (Brennan and Lindsay, 1998). Figure 12.11 is a redrafting of the experimental data in conjunction with the solubility constants of the solid phases from the Visual MINTEQ database. The implications of the original publication are unchanged despite some significant differences between the MINTEQ constants and those used by Brennan and Lindsay (1998). The observed solubilities of Fe^{2+} as a function of redox had a slope consistent with a solid phase with a stoichiometry of $Fe(II)Fe(III)_2O_4$. Although the molecular structure and Fe valence states of the resultant precipitate were not verified, the authors
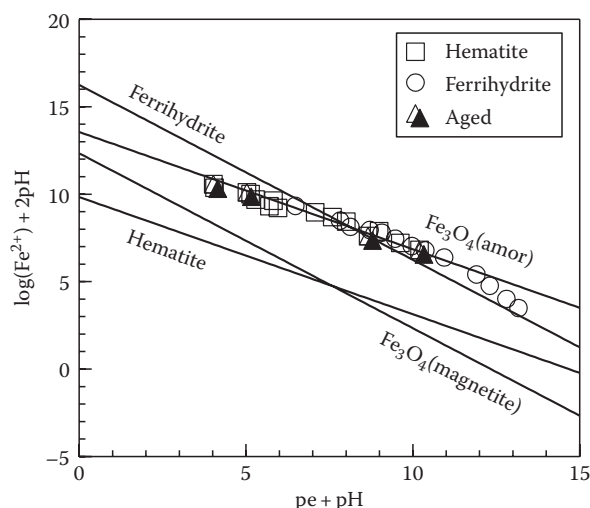
FIGURE 12.11 Oxidation/reduction experiments under controlled conditions with different starting materials or different incubation periods. The open squares began with hematite, the open circles began with ferrihydrite, and the filled triangles were aged for 15 days. (From Brennan, E.W., and W.L. Lindsay. 1998. Reduction and oxidation effect on the solubility and transformation of iron oxides. Soil Sci. Soc. Am. J. 62:930–937.)

suggested that Fe solubility was being controlled by an amorphous form of magnetite, Fe_3O_4(amorphous).

12.7.5 Limitations to Applying Geochemical Models to Soil Solutions

Four major limitations exist in the application of equilibrium geochemical models to soil solutions: the dynamic (nonequilibrium) nature of soils, poorly defined equilibria, limitations in the analytical chemistry, and lack of knowledge of how to accurately handle soluble organic compounds. The ever-changing nature of soils naturally leads to nonequilibrium conditions. Although equilibrium within the soil solution may be attained quickly, solid and gas phases often establish equilibrium at a slow rate, if at all. Unfortunately, these systems have a very strong impact on the composition of the soil solution. Understanding the degree of nonequilibrium and the rate at which a steady state is being approached can be discerned by sampling the soil solution over time and monitoring the progress of the chemical reactions. Suarez (1995) discussed the merits of a nonequilibrium approach to modeling the carbonate system.

The problem of poorly defined equilibria arises when information about the system is lacking or if the database accompanying the geochemical model does not contain all the equilibrium constants to fully characterize important reactions. If data are severely lacking, then a speciation approach may not be possible. However, estimating experimental parameters or equilibrium constants may be an acceptable alternative. Limitations in analytical chemistry (quantification limits, capability to analyze certain species) must be dealt with in the same way.

As discussed in previous sections, soluble organic compounds are among the most important components in the soil solution,

and we have the least thermodynamic information about them. These compounds can be very complex and contain a wide range of functional groups and molecular chemical environments, making it difficult to approximate complexation constants. Thus, a decision must be made either to ignore the effects of soluble organics completely because their impact cannot be quantified exactly or to apply estimates of the constants. The latter approach has been applied with some success, including the addition of an organic component to Visual MINTEQ.

12.8 Current Status and Future Research Directions

This review of the chemistry of the soil solution assumed a distinct theoretical approach and did not focus on some of the excellent, recent advances in other aspects of the science. Topics not covered include colloidal chemistry (both inorganic and organic), spectroscopic approaches to identifying important complexes in the soil solution, salinity, xenobiotics in soil solutions, and heavy metals. Most of these topics are covered in other chapters.

The study of the soil solution has its origins in agriculture, and a great portion of soil chemical research remains agriculturally oriented. Many challenges exist in which soil chemists will play a pivotal role including managing soil acidity, growing crops in arid regions with limited or poor quality irrigation water, adapting to the inevitable trend toward lower quality fertilizers, and contributing to the development of truly sustainable agricultural systems.

Although soil chemistry has its history and future firmly rooted in agriculture, soil chemists are increasingly finding a niche in other, less traditional research areas. Geochemists, environmental engineers, hydrologists, and chemical engineers are finding that soil chemists make excellent research partners in solving environmental and production problems. Working knowledge of and the ability to use new technologies have allowed soil chemists to explore new areas: molecular-level spectroscopy to quantify the chemistry of surfaces, complexation reactions in solution, and chemical associations in both the solid and solution phases; analytical instrumentation to quantify ultratrace quantities of contaminants; and molecular genetics of microorganisms to help answer difficult ecological and environmental questions.

References

Adams, F. 1971. Ionic concentrations and activities in soil solution. Soil Sci. Soc. Am. Proc. 35:420–426.

Adams, F. 1974. Soil solution, p. 441–481. *In* E.W. Carson (ed.) The plant root and its environment. University Press of Virginia, Charlottesville, Virginia.

Allison, J.D., D.S. Brown, and K.J. Novo-Gradac. 1990. MINTEQA2/PRODEFA2, a geochemical assessment model for environmental systems: Version 3.00 user's manual. EPA-600/3-91-021. U.S. EPA, Athens, GA.

Baham, J. 1984. Prediction of ion activities in soil solutions: Computer equilibrium modeling. Soil Sci. Soc. Am. J. 48:525–531.

Ball, J.W., E.A. Jenne, and D.K. Nordstrom. 1979. WATEQ—A computerized chemical model for trace and major element speciation and mineral equilibria of natural waters, p. 813–835. *In* E.A. Jenne (ed.) Chemical modeling in aqueous systems. ACS Symposium Series No. 93. American Chemical Society, Washington, DC.

Bartlett, R., and B. James. 1980. Studying dried, stored soil samples—Some pitfalls. Soil Sci. Soc. Am. J. 44:721–724.

Berggren, D., and J. Mulder. 1995. The role of organic matter in controlling aluminum solubility in acidic mineral soil horizons. Geochim. Cosmochim. Acta 59:4167–4180.

Bohn, H., B. McNeal, and G. O'Connor. 1985. Soil chemistry. Wiley Interscience, New York.

Brennan, E.W., and W.L. Lindsay. 1998. Reduction and oxidation effect on the solubility and transformation of iron oxides. Soil Sci. Soc. Am. J. 62:930–937.

Caron, A., S. Ben Jemia, J. Gallichand, and L. Trepanier. 1999. Field bromide transport under transient state: Monitoring with time domain reflectometry and porous cup. Soil Sci. Soc. Am. J. 63:1544–1553.

Cole, D.W., S.P. Gessel, and E.E. Held. 1961. Tension lysimeter studies of ion and moisture movement in glacial till and coral atoll soils. Soil Sci. Soc. Am. Proc. 25:321–325.

Dudley, L.M., and B.L. McNeal. 1987. A model for electrostatic interaction among charged sites of water-soluble organic polyions: I. Description and sensitivity. Soil Sci. 143:329–340.

Elberling, B., and H. Matthiesen. 2007. Methodologically controlled variations in laboratory and field pH measurements in waterlogged soils. Eur. J. Soil Sci. 58:207–214.

Elkhatib, E.A., O.L. Bennett, V.C. Baligar, and R.J. Wright. 1986. A centrifugation method for obtaining soil solution using an immiscible liquid. Soil Sci. Soc. Am. J. 50:297–299.

Evans, K.A., and S.A. Banwart. 2006. Rate controls on the chemical weathering of natural polymineralic material. I. Dissolution behavior of polymineralic assemblages determined using batch and unsaturated column experiments. Appl. Geochem. 21:352–376.

Felmy, A.R., and J.H. Weare. 1995. The development and application of aqueous thermodynamic models: The specific ion-interaction approach. *In* R. Loeppert et al. (eds.) Chemical equilibrium and reaction models. Soil Science Society of America Special Publications No. 42. ASA, Madison, WI.

Garrels, R.M., and C.L. Christ. 1965. Solutions, minerals, and equilibria. Freeman, Cooper & Co., San Francisco, CA.

Geuantman, L. 2001. *In* CRC Handbook of Chemistry and Physics, 82 Edn., Sec. 8, CRC Press Boca Raton, FL.

Gustafsson, J.P. 2009. Visual MINTEQ version 2.61. http://www.lwr.kth.se/English/OurSoftware/vminteq/, accessed on May 16, 2011.

Hummel, W. 2005. Solubility equilibria and geochemical modeling in the field of radioactive waste disposal. Pure Appl. Chem. 77:631–641.

Kielland, J. 1937. Individual activity coefficients of ions in aqueous solutions. J. Am. Chem. Soc. 59:1675–1678.

Kittrick, J.A. 1977. Mineral equilibria and the soil system, p. 1–25. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Lewis, G.N., and M. Randall. 1923. Thermodynamics and the free energy of chemical substances. McGraw-Hill, New York.

Lindsay, W.L. 1979. Chemical equilibria in soils. Wiley Interscience, New York.

Lovelock, J.E. and R.J. Maggs. 1993. Halogenated hydrocarbons in and over the Atlantic. Nature. 241:194–196.

Martell, A.S., and R.M. Smith. 2004. NIST critically selected stability constants of metal complexes database; NIST standard reference database 46, version 8.0. NIST, Gaithersburg, MD.

McBride, M.B. 1994. Environmental chemistry of soils. Oxford University Press, New York.

McKeague, J.A., M.V. Cheshire, F. Andreux, and J. Berthelin. 1986. Organo-mineral complexes in relation to pedogenesis, p. 549–592. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and microbes. SSSA, Madison, WI.

Melchior, D.C., and R.L. Bassett. 1990. Chemical modeling of aqueous systems II, ACS Symposium Series 416. American Chemical Society, Washington, DC.

Morel, F.M.M., and J.J. Morgan. 1972. A numerical method for computing equilibrium in aqueous chemical systems. Environ. Sci. Technol. 6:58–67.

Morris, D.F.C. 1969. Ionic radii and enthalpy of hydration of ions. Struct. Bond. 63:157–159.

Morrison, R.D., and B. Lowry. 1990. Effect of cup properties, sampler geometry and vacuum on the sampling rate of porous cup samplers. Soil Sci. 149:308–316.

Pearson, R.G. 1963. Hard and soft acids and bases. J. Am. Chem. Soc. 85:3533.

Pitzer, K.S. 1979. Theory: Ion interaction approach, p. 157–208. *In* R.M. Pytkowicz (ed.) Activity coefficients in electrolyte solutions. CRC Press, Boca Raton, FL.

Richards, L.A. 1941. A pressure-membrane extraction apparatus for soil solution. Soil Sci. 51:377–386.

Ross, D.S., and R.J. Bartlett. 1990. Effects of extraction methods and sample storage on properties of solutions obtained from forested spodosols. J. Environ. Qual. 19:108–113.

Sadiq, M., and W.L. Lindsay. 1979. Selection of standard free energies of formation for use in soil chemistry. Colorado State University Experiment Station Technical Bulletin No. 134. Colorado State University, Fort Collins, CO.

Schloesing, T. 1866. Sur l'anlse des pricipes solubles de la terre vegetale. Compt. Rend. Acad. Sci. 63:1007.

Schwab, A.P. 1989. Manganese–phosphate solubility relationships in an acid soil. Soil Sci. Soc. Am. J. 53:1654–1660.

Schwab, A.P. 1992. Chemical and physical characterization of soils. *In* L.E. Erickson, S.C. Grant, and J.P. McDonald (eds.) Conf. Proc. Hazard. Subst. Res. University of Colorado, Boulder, Co., p. 326–344.

Skvortsov, V.G., R.S. Tsekhanskii, and A.M. Gavrilov. 1976. System of sodium metaborates–sodium chloride–water at 20 degrees C. Zh. Neorg. Khim. 21:583–585.

Smith, G.R., K.K. Tanji, R.G. Burau, and J.J. Jurinak. 1995. C-salt, a chemical equilibrium model for multicomponent solutions. *In* R. Loeppert et al. (eds.) Chemical equilibrium and reaction models. Soil Science Society of America Special Publication No. 42. ASA, Madison, WI.

Sposito, G. 1981. The thermodynamics of soil solutions. Oxford University Press, New York.

Sposito, G. 1989. The chemistry of soils. Oxford University Press, New York.

Sposito, G. 1994. Chemical equilibria and kinetics in soils. Oxford University Press, New York.

Sposito, G. 2008. The chemistry of soils. 2nd edn. Oxford University Press, New York.

Sposito, G., and J. Coves. 1988. SOILCHEM: A computer program for the calculation of chemical equilibria in soil solutions and other natural water systems. Kearney Foundation of Soil Science, University of California, Riverside, CA.

SSSA. 2009. Glossary of soil science terms. (Available online with updates at https://www.soils.org/sssagloss/index.php, accessed on May 16, 2011.)

Staudinger, J., and P.V. Roberts. 1996. A critical review of Henry's law constants for environmental applications. Crit. Rev. Environ. Sci. Technol. 26:205–297.

Stevenson, F.J. 1967. Organic acids in soil, p. 119–146. *In* A.D. McLaren and G.H. Peterson (eds.) Soil biochemistry. Wiley, New York.

Stumm, W., and J.J. Morgan. 1996. Aquatic chemistry. Chemical equilibria and rates in natural waters. Wiley Interscience, New York.

Suarez, D.L. 1995. Carbonate chemistry in computer programs and application to soil chemistry. *In* R. Loeppert et al. (eds.) Chemical equilibrium and reaction models. Soil Science Society of America Special Publication No. 42. ASA, Madison, WI.

Thompson, H.S. 1850. On the absorbent power of soils. R. Agric. Soc. Eng. J. 11:68–74.

Till, A.R., and T.P. McCabe. 1976. Sulfur leaching and lysimeter characterization. Soil Sci. 121:44–47.

Truesdall, A.H., and B.F. Jones. 1974. WATEQ, a computer program for calculating chemical equilibria of natural waters. U.S. Geol. Surv. 2:223.

United States Salinity Laboratory Staff. 1954. Diagnosis and improvement of saline and alkali soils. USDA handbook no. 60. USDA, Washington, DC.

van der Lee, J., and L. De Windt. 2001. Present state and future directions of modeling of geochemistry in hydrogeological systems. J. Contam. Hydrol. 47:265–282.

van Hees, P., U. Lundstrom, R. Danielsson, and L. Nyberg. 2001. Controlling mechanisms of aluminum in soil solution—An evaluation of 180 podzolic soils. Chemosphere 45:1091–1101.

Warrick, A.W., and A. Amoozegar-Fard. 1977. Soil water regimes near porous cup water samplers. Water Resour. Res. 13:203–207.

Way, J.T. 1850. On the power of soils to absorb manure. R. Agric. Soc. Eng. J. 11:313–379.

West et al. 1992. Microbial activity and survival in soils dried at different rates. Aust. J. Soil Res. 30:209–222.

Westall, J.C., J.L. Zachary, and F.M.M. Morel. 1976. MINEQL—A computer program for the calculation of chemical equilibrium composition of aqueous systems. Technical Note No. 18. MIT, Cambridge, MA.

Wolt, J.D. 1989. SOILSOLN: A program for teaching equilibria modeling of soil solution composition. J. Agron. Educ. 18:40–42.

Wolt, J.D. 1994. Soil solution chemistry. Applications to environmental science and agriculture. John Wiley & Sons, New York.

Wolt, J.D., and J.G. Graveel. 1986. A rapid routine method for obtaining soil solution using vacuum displacement. Soil Sci. Soc. Am. J. 50:602–605.

Wood, W.G. 1973. A technique using porous cups for water sampling at any depth in the unsaturated zone. Water Resour. Res. 9:468–488.

13

Kinetics and Mechanisms of Soil Chemical Reactions

Donald L. Sparks
University of Delaware

13.1 Introduction

Since its inception in the mid-1850s, soil chemistry has focused on the macroscopic, equilibrium aspects of soil chemical reactions (CRs) and processes. From these studies, much was learned about important soil chemical processes, including sorption, desorption, precipitation, complexation, dissolution, and oxidation/reduction. However, such investigations do not convey information on reaction rates or reaction mechanisms. In the past three decades, as concerns and interests about soil and water quality have increased, soil and environmental chemists, environmental and chemical engineers, and geochemists have increasingly realized that reactions in subsurface environments are time dependent. Thus, to accurately predict the fate, mobility, speciation, and bioavailability of environmentally important plant nutrients, trace elements, radionuclides, and organic chemicals in soils, one must understand the kinetics and mechanisms of the reactions.

While major progress has been made in better understanding the kinetics of soil chemical processes, much uncertainty remains. In part, this is due to the complex, heterogeneous nature of natural materials such as soils. However, with the development of kinetic techniques that can be used to measure a wide range of timescales, time-dependent models that can describe both CR and mass transfer processes, and the employment of state-of-the-art in situ spectroscopic and microscopic surface techniques in combination with rate studies, major advances are being made in understanding the kinetics and mechanisms of

soil CRs. Arguably, this will be a major leitmotif in soil chemistry research for decades to come.

In this review, the application of chemical kinetics to heterogeneous systems, such as soils and soil components (clay minerals, organic matter, metal hydr(oxides), and humic substances), with emphasis on sorption/release processes will be discussed. A critical review of kinetic models that can be used to describe reaction rates on heterogeneous surfaces will be covered. The review will also present discussions on the rates of important soil CRs and processes including inorganic and organic sorption/desorption, dissolution, and redox. For additional details on these topics and other aspects of kinetics of soil chemical and geochemical processes, the reader should consult a number of recent books and monographs (Sparks, 1989, 1995, 2002, 2005; Sparks and Suarez, 1991; Stumm, 1992; Schwarzenbach et al., 1993; Sposito, 1994; Grossl et al., 1997; Matocha et al., 2005; Borda and Sparks, 2008).

13.2 Timescales of Soil Chemical Processes

A variety of CRs occur in soils and often in combination with one another. Reaction timescales can vary from microseconds to milliseconds for many ion association and some ion exchange, sorption, and redox reactions to years for many mineral solution and mineral crystallization phenomena and for some sorption/release reactions (Figure 13.1). Ion association reactions include ion pairing, inner- and outer-sphere complexation,
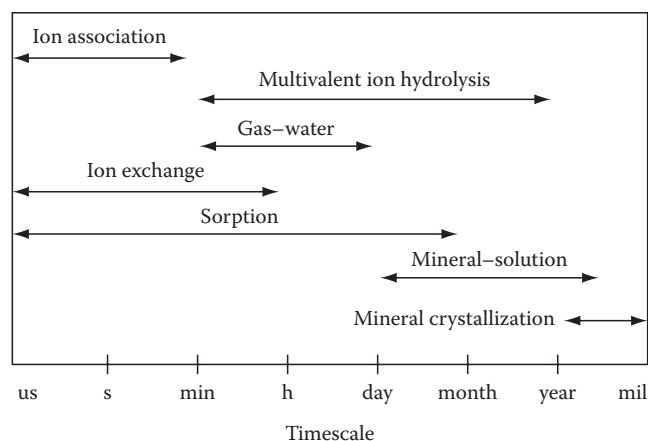
FIGURE 13.1 Time ranges required to attain equilibrium by different types of reactions in soil environments. (Reprinted with permission from Amacher, M.C. 1991. Methods of obtaining and analyzing kinetic data, p. 19–59. *In* D.L. Sparks and D.L. Suarez (eds.) Rates of soil chemical processes. SSSA, Madison, WI.)

and chelation in solution. Gas–water reactions involve gaseous exchange across the air–liquid interface. Ion exchange reactions occur when cations and anions are adsorbed (outer-sphere complexation) and desorbed from soil surfaces by electrostatic attractive forces. Ion exchange reactions are reversible and stoichiometric. Sorption reactions can involve adsorption processes, including partitioning, outer-sphere and inner-sphere complexation, and multinuclear complexation (e.g., surface precipitation). Mineral–solution reactions include precipitation/dissolution of minerals and coprecipitation reactions in which small constituents become a part of mineral structures (Sparks, 1989; Amacher, 1991).

The type of soil component can drastically affect the reaction rate. For example, sorption reactions are often more rapid on clay minerals, such as kaolinite and smectites, than on vermiculitic and micaceous minerals. This is in large part due to the availability of sites for sorption. For example, kaolinite has readily available planar external sites, and smectites have primarily internal sites that are also quite available for retention of sorbates. Thus, sorption reactions on these soil constituents are often quite rapid, even occurring on timescales of seconds and milliseconds (Sparks, 1989, 2005).

On the other hand, vermiculite and micas have multiple sites for retention of metals and organics, including planar, edge, and interlayer sites, with some of the latter sites being partially to totally collapsed. Consequently, sorption and desorption reactions on these sites can be slow, tortuous, and mass transfer controlled. Often, an apparent equilibrium may not be reached even after several days or weeks. Thus, with vermiculite and mica, sorption can involve two to three different reaction rates: high rates on external sites, intermediate rates on edge sites, and low rates on interlayer sites (Jardine and Sparks, 1984; Comans and Hockley, 1992).

Metal sorption reactions on oxides, hydroxides, and humic substances depend on the type of surface and metal being

studied, but the CR rate appears to be rapid. For example, CR rates of metals and oxyanions on goethite occurred on millisecond timescales (Zhang and Sparks, 1989, 1990a, 1990b; Grossl et al., 1994, 1997). Reaction rates of redox processes can also be very rapid. For example, Parikh et al. (2008) found that 50% of the initial reaction process for As(III) oxidation on Mn oxides occurred in 1 min. Half-times for divalent Pb, Cu, and Zn sorption on peat ranged from 5 to 15 s (Bunzl et al., 1976). A number of studies have shown that heavy metal sorption on oxides (Bruemmer et al., 1988; Ainsworth et al., 1994; Scheidegger and Sparks, 1996a; Scheidegger et al., 1998) and clay minerals (Lövgren et al., 1990) increases with longer residence times. The mechanism for these lower reaction rates is not well understood but has been ascribed to diffusion phenomena, sites of lower reactivity, and surface nucleation/precipitation (Scheidegger and Sparks, 1996b; Sparks, 1998, 1999a, 1999b). Recent findings on slow metal retention rates and mechanisms at the mineral/water interface will be discussed later.

Sorption/desorption of metals and organic chemicals on soils is often very slow, which has been attributed to diffusion into micropores of inorganic minerals and into humic substances, retention on sites of varying reactivity, and surface nucleation/precipitation (Sparks 1989, 1999a, 1999b, 2002, 2005; Scheidegger and Sparks, 1996b; Matocha et al., 2005; Borda and Sparks, 2008). These reactions will be discussed in more detail later.

13.3 Application of Chemical Kinetics to Heterogeneous Surfaces

The study of chemical kinetics, even in homogeneous systems, is complex and often arduous. When one attempts to study the kinetics of reactions in heterogeneous systems such as soils, sediments, and even soil components, such as clay minerals, hydrous oxides, and humic substances, the difficulties are greatly magnified. This is largely due to the complexity of soils that are made up of a mixture of inorganic and organic components. These components often interact with each other and display different types of sites with various reactivities for inorganic and organic sorptives. Moreover, the variety of particle sizes and porosities in soils and sediments further adds to their heterogeneity. In most cases, both chemical kinetics and multiple transport processes are occurring simultaneously. Thus, the determination of chemical kinetics, which can be defined as the investigation of rates of CRs and of the molecular processes by which reactions occur where transport is not limiting (Gardiner, 1969), is extremely difficult, if not impossible, in heterogeneous systems. In these systems, one is studying kinetics, which is a generic term referring to time-dependent or nonequilibrium processes. Thus, apparent and nonmechanistic rate laws and rate parameters are determined (Skopp, 1986; Sparks, 1989).

13.3.1 Rate-Limiting Steps

Both transport and CR processes can affect the reaction rates in the subsurface environment. Transport processes include
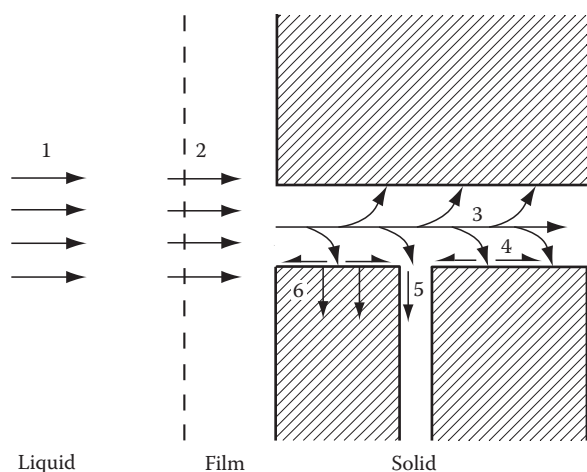
Liquid   Film   Solid

FIGURE 13.2 Transport processes in solid–liquid soil reactions—nonactivated processes: (1) transport in the soil solution, (2) transport across a liquid film at the solid–liquid interface, (3) transport in a liquid-filled macropore—activated processes, (4) diffusion of a sorbate at the surface of the solid, (5) diffusion of a sorbate occluded in a micropore, (6) diffusion in the bulk of the solid. (Reprinted from Aharoni, C., and D.L. Sparks. 1991. Kinetics of soil chemical reactions: A theoretical treatment, p. 1–18. *In* D.L. Sparks and D.L. Suarez (eds.) Rates of soil chemical processes. SSSA, Madison, WI.)

(Aharoni and Sparks, 1991) the following: (1) transport in the solution phase; (2) transport across a liquid film at the particle/liquid interface (film diffusion [FD]); (3) transport in liquid-filled macropores (>2 nm), all of which are nonactivated diffusion processes and occur in mobile regions; (4) particle diffusion (PD) processes, which include diffusion of sorbate occluded in micropores (<2 nm, pore diffusion) and along pore wall surfaces (surface diffusion); and (5) diffusion processes in the bulk of the solid, all of which are activated diffusion processes (Figure 13.2). Pore and surface diffusion within the immediate region can be referred to as interaggregate (interparticle) diffusion while that in the solid is intraaggregate diffusion. The actual CR at the surface, for example, adsorption, is usually instantaneous. The slowest of the CR and transport processes is rate limiting.

13.3.2 Rate Laws

There are two important reasons for investigating the rates of soil chemical processes (Sparks, 1989, 1995, 1998, 1999a, 1999b, 2002) (1) to determine how rapidly reactions attain equilibrium and (2) to infer information on reaction mechanisms. One of the most important aspects of chemical kinetics is the establishment of a rate equation or law. By definition, a rate law is a differential equation. For the following reaction (Bunnett, 1986; Sparks, 1989, 1995, 1998, 1999a, 1999b, 2002),

$$aA + bB \rightarrow yY + zZ, \qquad (13.1)$$
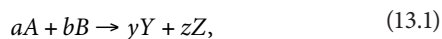
the rate is proportional to some power of the concentrations of reactants *A* and *B* and/or other species (*C*, *D*, etc.) in the

system, and *a*, *b*, *y*, and *z* are stoichiometric coefficients and are assumed to be equal to 1 in the discussion that follows on rate laws. The power to which the concentration is raised may equal 0 (i.e., the rate is independent of that concentration), even for reactant *A* or *B*. Rates are expressed as a decrease in reactant concentration or an increase in product concentration per unit time. Thus, the rate of conversion of reactant *A* above, which has a concentration [*A*] at any time *t*, is $(-d[A]/(dt))$ while the rate with regard to product *Y* having a concentration [*Y*] at time *t* is $(d[Y]/(dt))$.

The rate expression for Equation 13.1 is therefore as follows:

$$\frac{d[Y]}{dt} = -\frac{d[A]}{dt} = [A]^{\alpha}[B]^{\beta}, \qquad (13.2)$$

where
  k is the rate constant
  α and β are the orders of the reaction with respect to reactants *A* and *B*, respectively, and can be referred to as a partial orders for the total reaction

These orders are experimentally determined and not necessarily integral numbers. The sum of all the partial orders (α and β) is the overall order (*n*) of the total reaction and may be expressed as follows:

$$n = \alpha + \beta + \cdots \qquad (13.3)$$

Once the values of α, β, etc., are determined experimentally, the rate law is defined. Reaction order provides only information about the manner in which rate depends on concentration. Order does not mean the same as molecularity, which concerns the number of reactant particles (atoms, molecules, free radicals, or ions) entering into an elementary reaction. One can define an elementary reaction as one in which no reaction intermediates have been detected or need to be postulated to describe the CR on a molecular scale. An elementary reaction is assumed to occur in a single step and to pass through a single transition state (Bunnett, 1986).

To demonstrate that a reaction is elementary, one can use experimental conditions that are different from those employed in determining the reaction rate law. For example, if one conducted kinetic studies using a flow technique with set steady-state flow rates, one could see if reaction rate and rate constants changed with flow rate. If they did, one would not be determining mechanistic rate laws (see definition below).

Rate laws serve three purposes: (1) they assist one in predicting the reaction rate, (2) mechanisms can be proposed, and (3) reaction orders can be ascertained. There are four types of rate laws that can be determined for soil chemical processes (Skopp, 1986): mechanistic, apparent, transport with apparent, and transport with mechanistic. Mechanistic rate laws assume that only chemical kinetics are operational and transport phenomena are not occurring. Consequently, it is difficult to determine mechanistic rate laws for most soil chemical systems due

to the heterogeneity of the system caused by different particle sizes, porosities, and types of retention sites. There is evidence that with some kinetic studies, using chemical relaxation techniques (Sparks, 1989; Sparks and Zhang, 1991), that mechanistic rate laws are determined or closely approximated since the agreement between equilibrium constants calculated from both kinetic and equilibrium studies are comparable (Hachiya et al., 1984; Tang and Sparks, 1993). Recent studies using molecular scale in situ spectroscopic methods such as attenuated total reflectance Fourier transform infrared (ATR–FTIR) and quick x-ray absorption spectroscopy (QXAS; Parikh et al., 2008; Ginder-Vogel et al., 2009; Landrot et al., 2010) to measure rapid initial reaction rates also show that chemical kinetics are being measured, and, thus, mechanistic rate laws can be proposed.

Since natural materials are heterogeneous and transport processes often affect the reaction rate, apparent rate laws are usually determined for such systems. Apparent rate laws include both chemical kinetics and transport-controlled processes. Thus, soil structure, stirring, mixing, and flow rate all would affect the kinetics. Transport with apparent rate laws emphasizes transport-limited phenomena. One often assumes first-order or zero-order reactions (see the following discussion on reaction order). In determining transport with mechanistic rate laws, one attempts to describe simultaneously transport-controlled and chemical kinetics phenomena. One is thus trying to explain accurately both the chemistry and physics of the system.

13.3.3 Determination of Reaction Order and Rate Constants/Coefficients

There are three basic ways to determine rate laws and rate constants/coefficients (Bunnett, 1986; Skopp, 1986; Sparks, 1989, 1995, 1998, 1999a, 1999b, 2002): (1) initial rates, (2) directly using integrated equations and graphing the data, and (3) using nonlinear least-squares analysis.

Let us assume the following elementary reaction between species *A*, *B*, and *Y*:

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} Y. \tag{13.4}$$
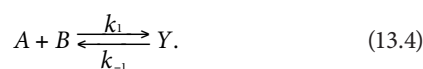
A forward reaction rate law can be written as

$$\frac{d[A]}{dt} = -k_1[A][B], \tag{13.5}$$

where
k_1 is the forward rate constant
α and β (see Equation 13.2) are each assumed to be 1

The reverse reaction rate law for Equation 13.4 is

$$\frac{d[A]}{dt} = +k_{-1}[Y], \tag{13.6}$$

where k_{-1} is the reverse rate constant.

Equations 13.5 and 13.6 are only applicable far from equilibrium where back or reverse reactions are insignificant. If both forward and reverse reactions are occurring, Equations 13.5 and 13.6 must be combined such that

$$\frac{d[A]}{dt} = -k_1[A][B] + k_{-1}[Y]. \tag{13.7}$$

Equation 13.7 applies the principle that the net reaction rate is the difference between the sum of all reverse reaction rates and the sum of all forward reaction rates.

One way to ensure that back reactions are not important is to measure initial rates. The initial rate is the limit of the reaction rate as time reaches zero. With an initial rate method, one plots the concentration of a reactant or product over a short reaction time period during which the concentrations of the reactants change so little that the instantaneous rate is hardly affected. Thus, by measuring initial rates, one could assume that only the forward reaction in Equation 13.4 predominates. This would simplify the rate law to that given in Equation 13.5, which, as written, would be a second-order reaction, first order in reactant *A* and first order in reactant *B*. Equation 13.5, under these conditions, would represent a second-order irreversible elementary reaction. To measure initial rates, one must have available a technique that can measure rapid reactions, such as a chemical relaxation method or a rapid molecular scale spectroscopic technique and an accurate analytical detection system to determine product concentrations. Recently, Parikh et al. (2008) used a novel real-time ATR–FTIR spectroscopic technique to measure the initial rates of As(III) oxidation on Mn oxides. Using a rapid scan technique, IR spectra were collected with a time resolution of up to 2.55 s. Ginder-Vogel et al. (2009) and Landrot et al. (2010) have employed state-of-the-art QXAS to investigate initial rates of As(III) oxidation on hydrous Mn oxide and Cr(III) oxidation on hydrous Mn oxide, respectively. In the study of Ginder-Vogel et al. (2009), QXAS scans were collected every 0.98 s, while scans were collected every 0.75 s in the study of Landrot et al. (2010).

Integrated rate equations can also be used to determine rate constants/coefficients. If one assumes that reactant *B* in Equation 13.5 is in large excess of reactant *A*, which is an example of the method of isolation to analyze kinetic data, and $Y_o = 0$, where Y_o is the initial concentration of product *Y*, Equation 13.5 can be simplified to

$$\frac{d[A]}{dt} = -k_1'[A], \tag{13.8}$$

where $k_1' = k_1[B]$.

The first-order dependence of [*A*] can be evaluated using the integrated form of Equation 13.8 using the initial conditions at $t = 0$, $A = A_o$:

$$\log[A]_t = \log[A]_o - \frac{k_1' t}{2.303}. \tag{13.9}$$
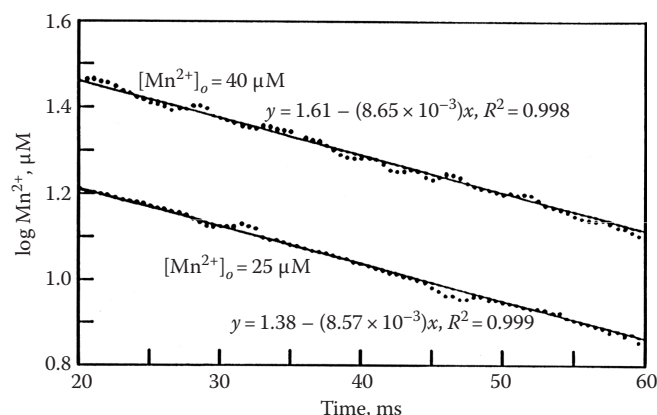
FIGURE 13.3 Initial reaction rates depicting the first-order dependence of Mn^{2+} sorption as a function of time for initial Mn^{2+} concentrations ($[Mn^{2+}]_o$) 25 and 40 μM. (Reprinted from Fendorf, S.E. et al. 1993. Soil Sci. Soc. Am. J., 57:57–62. With permission of the Soil Science Society of America.)

The half-time ($t_{1/2}$) for the above reaction is equal to $0.693/k_1'$ and is the time required for half of reactant A to be consumed.

If a reaction is first order, a plot of log $[A]_t$ versus t should result in a straight line with a slope = $k_1'/2.303$ and an intercept of log $[A]_o$. An example of first-order plots for Mn^{2+} sorption on δ-MnO_2 at two initial Mn^{2+} concentrations, $[Mn^{2+}]_o$ (25 and 40 μM), is shown in Figure 13.3. One sees that the plots are linear at both concentrations, which would indicate that the sorption process is first order. The $[Mn^{2+}]_o$ values obtained from the intercepts in Figure 13.3 were 24 and 41 μM, which are in good agreement with the two $[Mn^{2+}]_o$ values, and the rate constants were 3.73×10^{-3} and 3.75×10^{-3} s^{-1}, respectively. The fact that the rate constants do not significantly change with concentration is a good indication that the reaction in Equation 13.8 is first order under the imposed experimental conditions.

It is dangerous to conclude that a particular reaction order is correct, based simply on the conformity of data to an integrated equation. As illustrated above, multiple initial concentrations that vary considerably should be employed to see if the rate is independent of concentration. One should also test multiple integrated equations. It may also be useful to show that reaction rate is not affected by a species whose concentration does not change considerably during an experiment, such as substances not consumed or present in large excess (Bunnett, 1986; Sparks, 1989, 1991, 1995, 1998, 1999a, 1999b).

Least-squares analysis can also be used to determine rate constants/coefficients. With this method, one fits the best straight line to a set of points that are linearly related as $y = mx + b$, where y is the ordinate and x is the abscissa datum point, respectively. The slope (m) and the intercept (b) can be calculated by least-squares analysis.

When kinetic data are plotted, curvature may be observed due to an incorrect assumption of reaction order. If first-order kinetics is assumed and the reaction is really second order, downward curvature results. If second-order kinetics is assumed but the reaction is first order, upward curvature is observed. Curvature can also be due to fractional, third, higher, or mixed reaction order. Nonattainment of equilibrium often results in downward curvature. Temperature changes during the study can also cause curvature; thus, it is important that temperature be accurately controlled during a kinetic experiment.

13.4 Kinetic Models

13.4.1 Ordered Models

First-order kinetic models often describe reactions at the soil mineral/water interface. Both single first-order and multiple first-order reactions have been described by many investigators (Sparks, 1989, 1991, 2002; Sparks et al., 1993).

It is not uncommon to observe biphasic kinetics, namely, a rapid reaction rate followed by a much slower reaction rate. Such data can often be described by two first-order reactions. Some investigators have interpreted such biphasic kinetics to suggest reactions on two types of sites, such as external, readily accessible sites (Slope 1) and internal, difficult to access sites (Slope 2) (Jardine and Sparks, 1984; Comans and Hockley, 1992) or molecular sites of differing reactivity such as high-reactivity inner-sphere complex sites and low-reactivity outer-sphere complex sites (Grossl et al., 1994, 1997).

However, it is unsound to conclude anything about mechanisms based solely on multiple rate constants that are calculated from multiple slopes of kinetic plots. There are other ways to definitively ascertain reaction mechanisms, such as calculating energies of activation, elucidating rate-limiting steps through stopped flow and interruption approaches, using independent or direct methods to determine mechanisms such as spectroscopic and microscopic techniques, and employing blocking agents that are specific for certain reaction sites. An example of the latter approach is found in the research of Jardine and Sparks (1984) who studied K–Ca exchange on a Delaware soil at three temperatures and observed two apparent simultaneous first-order reactions at 283 and 298 K (Figure 13.4). They hypothesized that the first, more rapid reaction was predominantly due to adsorption on external planar sites of the organic matter and kaolinite in the soil. The slower reaction was ascribed to vermiculitic clay sites that promoted slow pore and surface diffusion. These hypotheses were seemingly validated using a large organic polymer, cetyltrimethylammonium bromide (CTAB), which because of its size is sterically hindered from internal sites. Thus, CTAB should only block external planar sites. When CTAB was applied to the soil, the first slope was eliminated, while the second slope was still present, suggesting multireactive sites.

While first-order models have been used widely to describe the kinetics of CRs on natural materials, a number of other simple kinetic models have also been employed. These include various ordered equations such as zero, second, and fractional order;
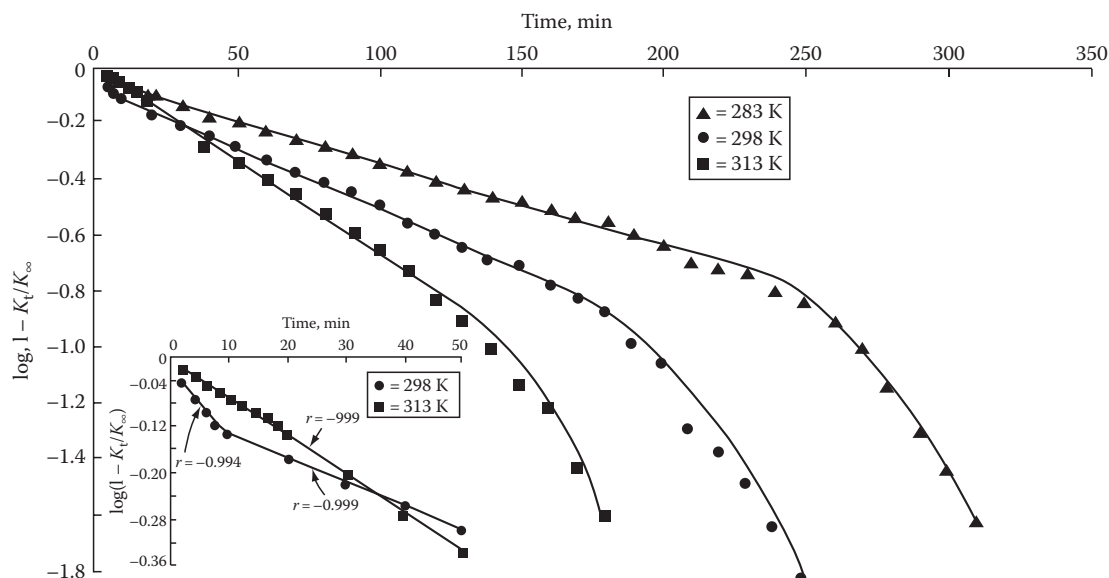
FIGURE 13.4 First-order kinetics for potassium adsorption at three temperatures on Evesboro soil with inset showing the initial 50 min of the first-order plots at 298 and 313 K. (Reprinted from Jardine, P.M., and D.L. Sparks. 1984. Potassium-calcium exchange in a multireactive soil system. I. Kinetics. Soil Sci. Soc. Am. J. 48:39–45. With permission of the Soil Science Society of America.)

TABLE 13.1　Linear Forms of Kinetic Equations Commonly Used in Environmental Soil Chemistry

Zero order[a]

$$[A]_t = [A]_o - k_1' t$$

First order

$$\log[A]_t = \log[A]_o - \frac{k_1' t}{2.303^c}$$

Second order[b]

$$\frac{1}{[A]_t} = \frac{1}{[A]_o} + kt$$

Elovich

$$q = \left(\frac{1}{\beta}\right) \ln(\alpha\beta) + \left(\frac{1}{\beta}\right) \ln t$$

Parabolic diffusion

$$\left(\frac{1}{t}\right)\left(\frac{Q_t}{Q_\infty}\right) = \frac{4}{\pi^{1/2}}\left(\frac{D}{r^2}\right)^{1/2} \frac{1}{t^{1/2}} - \frac{D}{r^2}$$

Power function

$$\ln q_t = \ln k + v \ln t$$

Source: Sparks, D.L., *Environmental Soil Chemistry*, Academic Press, San Diego, CA, 1995.

Terms in equations are defined in the text of the chapter.

[a] Describing the reaction $A \rightarrow Y$.

[b] $\ln x = 2.303 \log x$ is the conversion from natural logarithms (ln) to base 10 logarithms (log).

Elovich; power function or fractional power; and parabolic diffusion models. A brief discussion of some of these will be given; the final forms of the equations are given in Table 13.1. For more complete details and applications of these models, one may consult Sparks (1989, 1995, 1998, 1999a, 1999b, 2002).

13.4.2 Elovich Equation

The Elovich equation was originally developed to describe the kinetics of heterogeneous chemisorption of gases on solid surfaces (Low, 1960). It seems to describe a number of reaction mechanisms including bulk and surface diffusion and activation and deactivation of catalytic surfaces.

In soil chemistry, the Elovich equation has been used to describe the kinetics of sorption and desorption of various inorganic materials on soils (Sparks, 1989, 1995, 1998, 1999a, 1999b, 2002). It can be expressed as (Chien and Clayton, 1980)

$$q = \left(\frac{1}{\beta}\right) \ln(\alpha\beta) + \left(\frac{1}{\beta}\right) \ln t, \tag{13.10}$$

where

　q is the amount of sorbate per unit mass of sorbent at time t
　α and β are constants during any one experiment

A plot of q versus $\ln t$ should give a linear relationship if the Elovich equation is applicable with a slope of $(1/\beta)$ and an intercept of $(1/\beta)\ln(\alpha\beta)$. An application of Equation 13.10 to P sorption on soils is shown in Figure 13.5.

Some investigators have used the α and β parameters from the Elovich equation to estimate reaction rates. For example, it has been suggested that a decrease in β and/or an increase in α would increase reaction rate. However, this is questionable. The slope of plots using Equation 13.10 changes with the concentration of the sorptive and with the solution-to-soil ratio (Sharpley, 1983). Therefore, the slopes are not always characteristic of the soil but may depend on various experimental conditions.

Some researchers also have suggested that breaks or multiple linear segments in Elovich plots could indicate a changeover
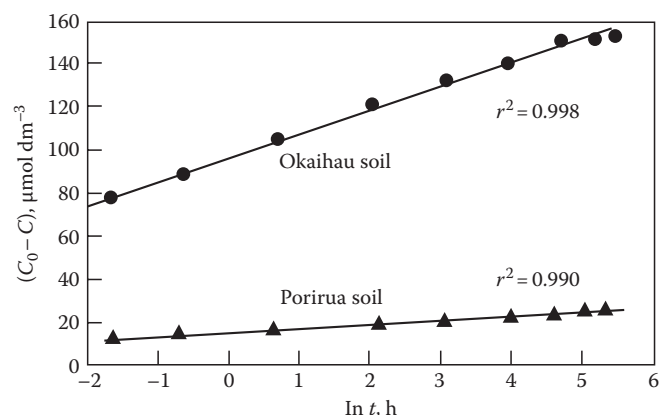
FIGURE 13.5 Plot of Elovich equation for phosphate sorption on two soils where C_0 is the initial phosphorus concentration in the soil solution at time t. The quantity (C_0-C) can be equated to q, the amount sorbed at time t. (Reprinted from Chien, S.H., and W.R. Clayton. 1980. Application of Elovich equation to the kinetics of phosphate release and sorption in soils. Soil Sci. Soc. Am. J. 44:265–268. With permission from the Soil Science Society of America.)

from one type of binding site to another (Atkinson et al., 1970). However, such mechanistic suggestions may not be correct (Sparks, 1989, 1995).

13.4.3 Parabolic Diffusion Equation

The parabolic diffusion equation is often used to suggest that diffusion-controlled phenomena are rate limiting. It was originally derived from radial diffusion in a cylinder where the ion concentration on the surface is constant, and initially, the ion concentration within the cylinder is uniform. It is also assumed that ion diffusion through the upper and lower faces of the cylinder is negligible. Following Crank (1976), the parabolic diffusion equation as applied to soils can be expressed as

$$\left(\frac{Q_t}{Q_\infty}\right) = \frac{4}{\pi^{1/2}}\frac{Dt^{1/2}}{r^2} - \frac{Dt}{r^2} - \frac{1}{3\pi^{1/2}}\frac{Dt^{3/2}}{r^2}, \quad (13.11)$$

where
  r is the radius of the cylinder
  Q_t is the quantity of diffusing substance that has left the cylinder at time t
  Q_∞ is the corresponding quantity after infinite time
  D is an apparent diffusion coefficient

For the relatively short times in most experiments, the third and subsequent terms may be ignored and thus

$$\frac{Q_t}{Q_\infty} = \frac{4}{\pi^{1/2}}\frac{Dt^{1/2}}{r^2} - \frac{Dt}{r^2}$$

or   (13.12)

$$\frac{1^c}{t}\frac{Q_t^m}{Q_\infty} = \frac{4}{\pi^{1/2}}\frac{Dt^{1/2}}{r^2}\frac{1}{t^{1/2}} - \frac{Dt}{r^2},$$

and thus, a plot of $Q_t/Q_\infty/t$ versus $1/t^{1/2}$ should give a straight line with a slope of

$$\frac{4}{\pi^{1/2}}\left(\frac{D}{r^2}\right)^{1/2}$$

and intercept of $-D/r^2$. Thus, if r is known, D may be calculated from both the slope and intercept.

The parabolic diffusion equation has successfully described metal reactions on soils and soil constituents (Chute and Quirk, 1967; Jardine and Sparks, 1984; Krishnamurti and Huang, 1992; Krishnamurti et al., 1997), feldspar weathering (Wollast, 1967), and pesticide reactions (Weber and Gould, 1966).

13.4.4 Fractional Power or Power Function Equation

This equation can be expressed as

$$q = kt^v, \quad (13.13)$$

where
  q is amount of sorbate per unit mass of sorbent at time t
  k and v are constants
  v is positive and <1

Equation 13.13 is empirical, except for the case where $v = 0.5$, when it is similar to the parabolic diffusion equation.

Equation 13.13 and various modified forms have been used by a number of researchers to describe the kinetics of reactions in natural materials (Kuo and Lotse, 1974; Havlin and Westfall, 1985).

13.4.5 $Z(t)$ and Diffusion Models

In a number of studies, several simple kinetic models have described rate data well, based on correlation coefficients and standard errors of the estimate (Chien and Clayton, 1980; Onken and Matheson, 1982; Sparks and Jardine, 1984; Allen et al., 1995). Despite this, there is often an inconsistent relation between the equation that gives the best fit and the physicochemical and mineralogical properties of the sorbent(s) being studied. Another problem with some of the kinetic models is that they are empirical and no meaningful rate parameters can be obtained.

Aharoni and Ungarish (1976) and Aharoni (1984) noted that some simple kinetic models are approximations of more general expressions within certain limited time ranges. They suggested a generalized empirical equation by examining the applicability of power function, Elovich, and first-order equations to experimental data. By writing these as the explicit functions of the reciprocal of the rate (Z), which is $(dq/dt)^{-1}$, one can show that a plot of Z versus t should be convex if the power function equation is operational (1 in Figure 13.6); linear if the Elovich
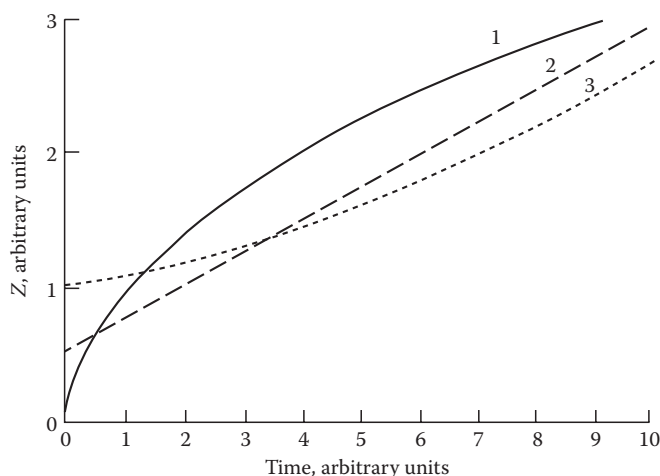
FIGURE 13.6 Plot of *Z* versus time implied by (1) power function model, (2) Elovich model, and (3) first-order model. The equations for the models were differentiated and expressed as explicit functions of the reciprocal of the rate, *Z*. (Reprinted with permission from Aharoni, C., and D.L. Sparks. 1991. Kinetics of soil chemical reactions: A theoretical treatment, p. 1–18. *In* D.L. Sparks and D.L. Suarez (eds.) Rates of soil chemical processes. SSSA, Madison, WI.)
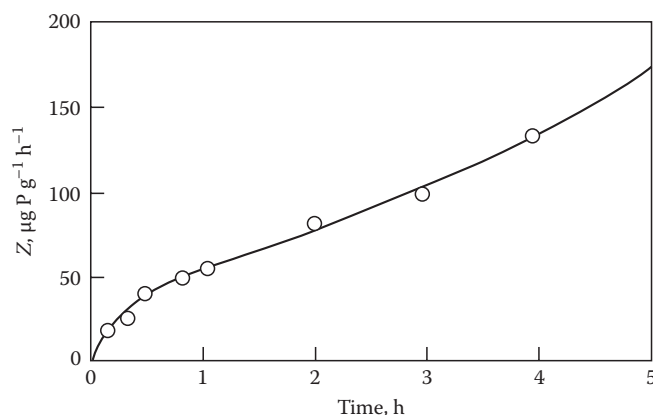
FIGURE 13.7 Sorption of phosphate by a Typic Dystrochrept soil plotted as *Z* versus time. The circles represent the experimental data of Polyzopoulos et al. (1986). The solid line is a curve calculated according to a homogeneous diffusion model. (Reprinted with permission from Aharoni, C., and D.L. Sparks. 1991. Kinetics of soil chemical reactions: A theoretical treatment, p. 1–18. *In* D.L. Sparks and D.L. Suarez (eds.) Rates of soil chemical processes. SSSA, Madison, WI.)

equation is appropriate (2 in Figure 13.6); and concave if the first-order equation is appropriate (3 in Figure 13.6). However, such plots for soil systems (Figure 13.7) are usually S-shaped, convex at small, concave at large, and linear at some intermediate *t* value suggesting that the reaction rate can best be described by the power function, first order and Elovich equations in these ranges of *t*, respectively. Thus, the S-shaped curve indicates that the above equations may be applicable, each over some limited time range.

One of the reasons a particular kinetic model appears to be applicable may be that the study is conducted during the time range when the model is most appropriate. While sorption, for example, decreases over many orders of magnitude before equilibrium is approached, with most methods and experiments, only a portion of the entire reaction is measured, and over this time range, the assumptions associated with a simple kinetic model (power function, Elovich, and first order) are valid. Aharoni and Suzin (1982a, 1982b) showed that the S-shaped curves could be well described using homogeneous and heterogeneous diffusion models. In homogeneous diffusion situations, the final and initial portions of the S-shaped curves (conforming to the power function and first-order equations, respectively) predominated, whereas in instances where the heterogeneous diffusion model was operational, the linear portion of the S-shaped curve that conformed to the Elovich equation predominated. Derivations of homogeneous and heterogeneous diffusion models can be found in Aharoni and Sparks (1991).

13.4.6 Implications of Diffusion Models

The finding that slower reactions at the soil mineral/liquid interface can be described by diffusional models indicates that the kinetics of chemical processes cannot be considered separately from physically limited transport phenomena. Thus, such a combination of processes cannot be treated using first- or other-order chemical kinetic equations. When one states that a reaction between the molecular species *A* and *B* is first order with respect to *A*, one assumes that the molecules of *A* have equal chances of participating in the reaction, and therefore, the rate is proportional to the concentration (C_A). This reasoning can be extended to a reaction between a sorbing surface and a sorptive. In this case, C_A refers to the number of reactive sites per unit area, which corresponds to the number of unoccupied sites per unit area ($1 - \theta_A$). However, by using first-order kinetics (or other order kinetics), one tacitly assumes that all the surface sites are potential reactants at any time and have an opportunity of participating in the sorption process. If one assumes that there are sites that cannot be reached directly from the fluid phase, but can be reached after the sorbate has undergone sorption and desorption at other sites, one cannot separate chemical kinetics from diffusion-limited kinetics. The overall kinetic process obeys a diffusion equation since diffusion is the rate-limiting process. However, the diffusion coefficient, which reflects the rate at which the sorbate jumps from one site to another, is determined by the rate of the CRs by which the sorbent–sorbate bonds are created and destroyed. Additionally, the activation energy for diffusion is equivalent to the activation energy of the CR.

13.4.7 Multiple Site Models

Based on the previous discussion, it is evident that simple chemical kinetic models such as ordered reaction, power function, and Elovich models may not be appropriate to describe

reactions in heterogeneous systems such as soils, sediments, and soil components. In these systems where there is a range of particle sizes and multiple retention sites, both chemical kinetics and transport phenomena are occurring simultaneously, and a fast reaction is often followed by a slower reaction(s). In such systems, nonequilibrium models that describe both chemical and physical nonequilibrium and that consider multiple components and sites are more appropriate. Physical nonequilibrium is ascribed to some rate-limiting transport mechanism such as FD or PD, while chemical nonequilibrium is due to a rate-limiting mechanism at the particle surface (CR). Nonequilibrium models include two-site, multiple site, radial diffusion (pore diffusion), surface diffusion, and multiprocess models (Table 13.2). Emphasis here will be placed on the use of these models to describe sorption phenomena.

The term sites can have a number of meanings (Brusseau and Rao, 1989): (1) specific, molecular scale reaction sites; (2) sites of differing degrees of accessibility (external and internal);

(3) sites of differing sorbent type (organic matter and inorganic mineral surfaces); and (4) sites with different sorption mechanisms. With chemical nonequilibrium sorption processes, the sorbate may undergo two or more types of sorption reactions, one of which is rate limiting. For example, a metal cation may sorb to organic matter by one mechanism and to mineral surfaces by another mechanism, with one of the mechanisms being time dependent.

13.4.7.1 Chemical Nonequilibrium Models

Chemical nonequilibrium models describe time-dependent reactions at sorbent surfaces. The one-site model is a first-order approach that assumes that the reaction rate is limited by only one process or mechanism on a single class of sorbing sites and that all sites are of the time-dependent type. In many cases, this model appears to describe soil CRs quite well. However, often it does not. This model would seem not appropriate for most heterogeneous systems since multiple sorption sites exist.

TABLE 13.2 Comparison of Sorption Kinetic Models

Conceptual Model	Fitting Parameters	Model Limitations
One-site model k_d $S \rightarrow C$	k_d	Cannot describe biphasic sorption/desorption
Two-site model $K_p D_{eff}$ $S' \rightarrow C'_s \rightarrow C$	k_d, K_p, X_1	May not describe the "bleeding" or slow, reversible, nonequilibrium desorption for residual sorbed compounds (Karickhoff, 1980)
Radial diffusion penetration retardation (pore diffusion) model (Wu and Gschwend, 1986) $K_p D_{eff}$ $S' \leftrightarrow C' \rightarrow C$	$D_{eff} = f(n,t)D_m n/(1-n)\rho_s K_p$	Cannot describe instantaneous uptake without additional correction factor; did not describe kinetic data for times greater than 10^3 min (Wu and Gschwend, 1986)
Dual-resistance surface diffusion model (Miller and Pedit, 1992) $D_s k_b$ $S' \rightarrow C'_s \rightarrow C$	D_s, k_b	Model calibrated with sorption data predicted more desorption than occurred in the desorption experiments (Miller and Pedit, 1992)
Multisite continuum compartment model (Connaughton et al., 1993) $F(t) = 1 - \dfrac{M(t)}{M} = 1 - \left(\dfrac{\beta}{\beta+1}\right)^\alpha$	α, β	Assumption of homogeneous, spherical particles and diffusion only in aqueous phase
Pore space diffusion model (Fuller et al., 1993) $\left(\in + \dfrac{S_a}{n} K_s C(r)^{(1-1/n)}\right)\dfrac{\partial C(r)}{\partial t} = D_e\left(\dfrac{\partial^2 C(r)}{\partial r^2} + \dfrac{2\partial C(r)}{r\partial r}\right)$	$D_e, \in, K_s, 1/n, F_{eq}$	

Source: Partially adapted from Connaughton, D.F., J.R. Stedinger, L.W. Lion, and M.L. Shuler. 1993. Description of time-varying desorption kinetics: Release of naphthalene from contaminated soils. Environ. Sci. Technol. 27:2397–2403.

S, concentration of the bulk sorbed contaminant (g g^{-1}); C, concentration of the bulk aqueous-phase contaminant (g mL^{-1}); k_d, first-order desorption rate coefficient (min^{-1}); S_2, concentration of the sorbed contaminant that is rate limited (g g^{-1}); S_p concentration of the contaminant that is in equilibrium with the bulk aqueous concentration (g g^{-1}); X_1, fraction of the bulk sorbed contaminant that is in equilibrium with the aqueous concentration; K_p, sorption equilibrium partition coefficient (mL g^{-1}); D_{eff}, effective diffusivity of sorbate molecules or ions in the particles (cm^2 s^{-1}); S', concentration of contaminant in immobile bound state (mol g^{-1}); C', concentration of contaminant free in the pore fluid (mol cm^{-3}); n, porosity of the sorbent (cm^3 of fluid cm^{-3}); D_m, pore fluid diffusivity of the sorbate (cm^2 s^{-1}); ρ_s, specific gravity of the sorbent (g cm^{-3}); $f(n,t)$, pore geometry factor; k_b, boundary layer mass transfer coefficient (m s^{-1}); r, radius of the spherical solid particle, assumed constant (m); ρ, macroscopic particle density of the solid phase (g m^{-3}); C_s, solution-phase solute concentration corresponding to an equilibrium with the solid-phase solute concentration at the exterior of the particle (g L^{-1}); D_s, surface diffusion coefficient (m s^{-1}); K_p, can be determined independently; K_p, D_m, and ρ_s can be determined independently; $F(t)$, fraction of mass released through time t; $M(t)$, mass remaining after time t; M, total initial mass; β, scale parameter necessary for determination of mean and standard deviation of k_s; α, shape parameter; $\in$, internal porosity of sorbent; $C(r)$, concentration of sorptive in the aqueous phase in the pore fluid at radial distance r; S_a is the surface of sorbent per unit volume of solid; $1/n$, the adsorption isotherm slope; K_a, adsorption isotherm intercept; D_{eq}, effective diffusion coefficient; a, radius of the aggregate; F_{eq}, equilibrium fraction of absorption sites.

The two-site (two compartment, two box) or bicontinuum model has been widely used to describe chemical nonequilibrium (Leenheer and Ahlrichs, 1971; Hamaker and Thompson, 1972; Karickhoff, 1980; Karickhoff and Morris, 1985; McCall and Agin, 1985; Jardine et al., 1992) and physical nonequilibrium (Nkedi-Kizza et al., 1984; Lee et al., 1988; van Genuchten and Wagenet, 1989; Table 13.2). This model assumes that there are two reactions occurring, one that is fast and reaches equilibrium quickly and a slower reaction that can continue for long time periods. The reactions can occur either in series or in parallel (Brusseau and Rao, 1989).

In describing chemical nonequilibrium with the two-site model, two types of sorbent sites are assumed. One site involves an instantaneous equilibrium reaction and the other, the time-dependent reaction. The former is described by an equilibrium isotherm equation while a first-order equation is usually employed for the latter.

With the two-site model, there are two adjustable or fitting parameters: the fraction of sites at local equilibrium (X_1), and the rate constant (k). A distribution (K_d) or partition coefficient (K_p) is determined independently from a sorption/desorption isotherm.

To account for the multiple sites that may exist in heterogeneous systems, Connaughton et al. (1993) developed a multisite compartment (continuum) model (Γ) that incorporates a continuum of sites or compartments with a distribution of rate coefficients that can be described by a gamma density function. A fraction of the sorbed mass in each compartment is at equilibrium with a desorption rate coefficient or distribution coefficient for each compartment or site (Table 13.2). The multisite model has two fitting parameters α, a shape parameter, and $1/\beta$, which is a scale parameter that determines the mean standard deviation of the rate coefficients. Figure 13.8 shows application of the Γ model to desorption of naphthalene from contaminated soils. The entire desorption process was described well with this model.

13.4.7.2 Physical Nonequilibrium Models

A number of models can be used to describe physical nonequilibrium reactions. Since transport processes in the mobile phase are not usually rate limiting, physical nonequilibrium models focus on diffusion in the immobile phase or interaggregate/diffusion processes such as pore and/or surface diffusion. The transport between mobile and immobile regions is accounted for in physical nonequilibrium models in three ways (Brusseau and Rao, 1989): (1) explicitly with Fick's law to describe the physical mechanism of diffusive transfer, (2) explicitly by using an empirical first-order mass transfer expression to approximate solute transfer, and (3) implicitly by using an effective or lumped dispersion coefficient that includes the effects of sink/source differences and hydrodynamic dispersion and axial diffusion.

A pore diffusion model (Table 13.2) has been used by a number of investigators to study sorption processes using batch systems (Wu and Gschwend, 1986; Steinberg et al., 1987; Ball
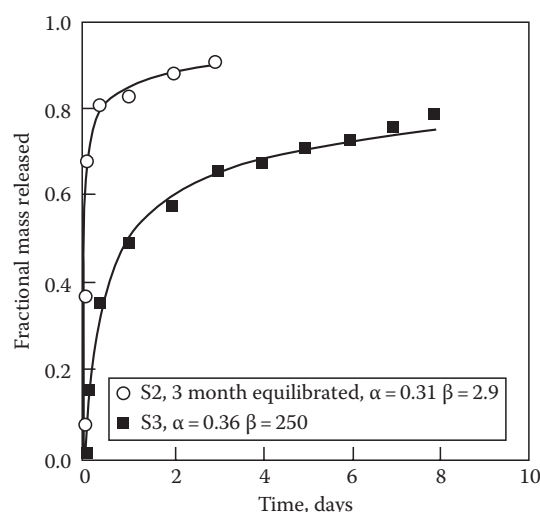
FIGURE 13.8 Mass fractional release of naphthalene from two soils, S2 and S3 (S2 is a freshly contaminated soil, reacted with naphthalene for 3 months and S3 is a field [aged]-contaminated soil) fitted with a multisite continuum compartment (Γ) model. (Reprinted from Connaughton, D.F., J.R. Stedinger, L.W. Lion, and M.L. Shuler. 1993. Description of time-varying desorption kinetics: Release of naphthalene from contaminated soils. Environ. Sci. Technol. 27:2397–2403. With permission American Chemical Society.)

and Roberts, 1991; Harmon et al., 1992; Pignatello et al., 1993). Wu and Gschwend (1986) successfully used the pore diffusion model to describe chlorobenzene congener sorption/desorption on soils and sediments. Figure 13.9 shows experimental and model fits for tetrachlorobenzene and pentachlorobenzene
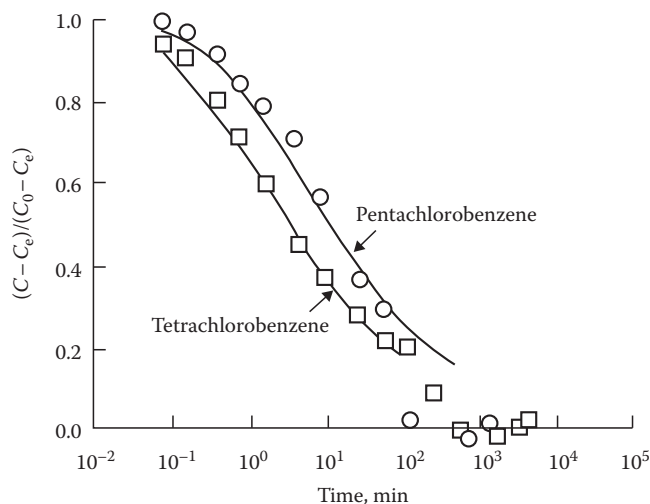
FIGURE 13.9 Experimental and model fitting results for pentachlorobenzene and tetrachlorobenzene sorption on Iowa soils where C is the dissolved concentration of organic chemical in the bulk solution, C_0 is the initial concentration, and C_e is the equilibrium concentration. The points represent experimental data and the solid lines represent fit of the data to the radial diffusion (pore diffusion) model. (Reprinted from Wu, S., and P.M. Gschwend. 1986. Sorption kinetics of hydrophobic organic compounds to natural sediments and soils. Environ. Sci. Technol. 20:717–725. With permission from American Chemical Society.)

sorption on soils. The sole fitting parameter in this model is the effective diffusion coefficient (D_e), which may be estimated a priori from chemical and colloidal properties. However, this estimation is only valid if the sorbent material has a narrow particle size distribution so that an accurate, average particle size can be defined. Moreover, in the pore diffusion model, an average representative D_e is assumed, which means there is a continuum in properties across an entire pore size spectrum. This is not a valid assumption for micropores in which there are higher adsorption energies of sorbates causing increased sorption. The increased sorption reduces diffusive transport rates and results in nonlinear isotherms for sorbents with pores less than several sorbate diameters in size. Other factors can cause reduced transport rates in micropores including steric hindrance, which increases as the pore size approaches the solute size and greatly increased surface area to pore volume ratios (which occurs as pore size decreases).

Another problem with the pore diffusion model is that sorption and desorption kinetics may have been measured over a narrow concentration range. This is a problem since a sorption/desorption mechanism in micropores at one concentration may be insignificant at another concentration.

Fuller et al. (1993) used a pore space diffusion model (Table 13.2) to describe arsenate adsorption on ferrihydrite that included a subset of sites whereby sorption was at equilibrium. A Freundlich model was used to describe sorption on these sites. Diffusion into the particle was described by Fick's second law of diffusion; homogeneous, spherical aggregates, and diffusion only in the aqueous phase were assumed. Figure 13.10 shows the fit of the model when sorption at all sites was controlled by intraaggregate diffusion. The fit was better when sites that had attained sorption equilibrium were included based on the assumption that there was an initial rapid sorption on external surface sites before intraaggregate diffusion.
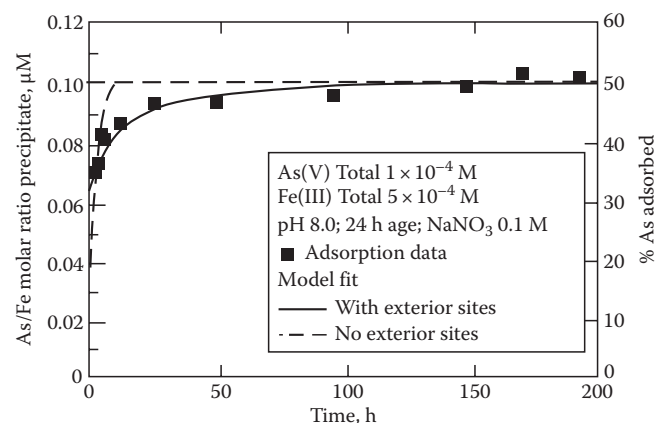
FIGURE 13.10 Comparison of pore space diffusion model fits of As(V) sorption with experimental data (dashed curve represents sorption where all surface sites are diffusion limited and the solid curve represents sorption on equilibrium sites plus diffusion-limited sites). (Reprinted from Fuller, C.C., J.A. Davis, and G.A. Waychunas. 1993. Surface chemistry of ferrihydrite: Part 2. Kinetics of arsenate adsorption and coprecipitation. Geochim. Cosmochim. Acta 57:2271–2282. With permission from Elsevier.)

Pedit and Miller (1995) have developed a general multiple particle class pore diffusion model that accounts for differences in physical and sorptive properties for each particle class (Table 13.2). The model includes both instantaneous equilibrium sorption and time-dependent pore diffusion for each particle class. The pore diffusion portion of the model assumes that solute transfer between the intraparticle fluid and the solid phases is fast vis-à-vis interparticle pore diffusion processes.

Surface diffusion models, assuming a constant surface diffusion coefficient, have been used by a number of researchers (Weber and Miller, 1988; Miller and Pedit, 1992). The dual resistance model (Table 13.2) combines both pore and surface diffusion.

13.5 Kinetics of Important Reactions on Soils and Soil Components

In the past several decades, numerous studies have been conducted on the kinetics of metal, oxyanion, radionuclide, plant nutrient, and organic CRs on natural materials. In this section, emphasis will be placed on the kinetics of sorption/desorption, precipitation/dissolution, and oxidation/reduction reactions on soils and soil components.

13.5.1 Sorption/Desorption Reaction Rates

13.5.1.1 Heavy Metals and Oxyanions

CRs of heavy metals and metalloids on soil components are rapid, occurring on a millisecond timescale. For such rapid reactions, chemical techniques such as pressure jump (p-jump) relaxation (Hayes and Leckie, 1986; Sparks, 1989; Sparks and Zhang, 1991; Grossl and Sparks, 1995; Sparks et al., 1996) and molecular scale methods such as ATR–FTIR and QXAS (Parikh and Chorover, 2007; Ginder-Vogel et al., 2009; Landrot et al., 2010) should be used.

The use of p-jump relaxation to measure the kinetics of ion sorption/desorption on metal oxide surfaces was pioneered by several Japanese chemists. Their research includes some of the following sorption/desorption kinetic studies: divalent metal ion (Hachiya et al., 1984), phosphate (Mikami et al., 1983a), chromate (Mikami et al., 1983c), and uranyl (Mikami et al., 1983b) sorption reactions on γ-Al$_2$O$_3$. Hayes and Leckie (1986) were the first to use p-jump relaxation to study sorption/desorption kinetics of a metal ion contaminant (Pb^{2+}) on goethite (α-FeOOH). Other successive studies monitored the rapid sorption/desorption kinetics of molybdate (Zhang and Sparks, 1989), sulfate (Zhang and Sparks, 1990a), selenate and selenite (Zhang and Sparks, 1990b), Cu^{2+} (Grossl et al., 1994), and arsenate and chromate (Grossl et al., 1997) on goethite. Additional studies have investigated borate sorption/desorption kinetics on pyrophyllite (Keren et al., 1994) and on γ-Al$_2$O$_3$ (Toner and Sparks, 1995).

Details of many of these studies are summarized in Hayes and Leckie (1986), Sparks (1989, 1995), Sparks and Zhang (1991), and Sparks et al. (1996) and will not be detailed here. A study

of Grossl et al. (1997) will be summarized to illustrate rapid CR rates of two environmentally important oxyanions (chromate and arsenate) on goethite. A double relaxation was observed for both arsenate and chromate sorption/desorption over a pH range of 6.5–7.5 for arsenate and 5.5–6.5 for chromate, respectively (Figures 13.11 and 13.12). Based on the double relaxations, a two-step process, resulting in the formation of an inner-sphere bidentate surface complex (Figure 13.13) was proposed. The first step involves an initial ligand exchange reaction of the aqueous oxyanion (H_2AsO or $HCrO_4$) with goethite, forming an inner-sphere monodentate surface complex that produces signals associated with fast τ values. The succeeding step involves a second ligand exchange reaction, resulting in the formation of an inner-sphere bidentate surface complex that produces the signal associated with slow τ values.

To determine if the mechanism displayed in Figure 13.13 was plausible and consistent with the kinetic data, the following linearized rate equations relating reciprocal relaxation time values (τ^1) to the concentrations of reactive species were used:

$$\tau_{\text{fast}}^{-1} + \tau_{\text{slow}}^{-1} = k_1([\text{XOH}] + [\text{ion species}]) + k_{-1} + k_2 + k_{-2}, \quad (13.14)$$

$$\tau_{\text{fast}}^{-1} \cdot \tau_{\text{slow}}^{-1} = k_1[k_2 + k_{-2}]([\text{XOH}] + [\text{ion species}]) + k_1 k_{-2}, \quad (13.15)$$

where the ion species are H_2AsO_4 or $HCrO_4$. The derivation of these equations was obtained from Bernasconi (1976) and is based on the two-step reaction system (A + B $\leftrightarrow$ C $\leftrightarrow$ D). If the mechanism portrayed in Figure 13.13 is accurate, then a plot of $\tau_{\text{fast}}^1 + \tau_{\text{slow}}^{-1}$ and $\tau_{\text{fast}}^1 \cdot \tau_{\text{slow}}^{-1}$ as a function of the concentration term ([XOH] + [ion species]) should be linear. Plots of Equations 13.14 and 13.15 were linear for both arsenate and chromate, suggesting that the proposed mechanism was plausible (Figures 13.14 and 13.15).
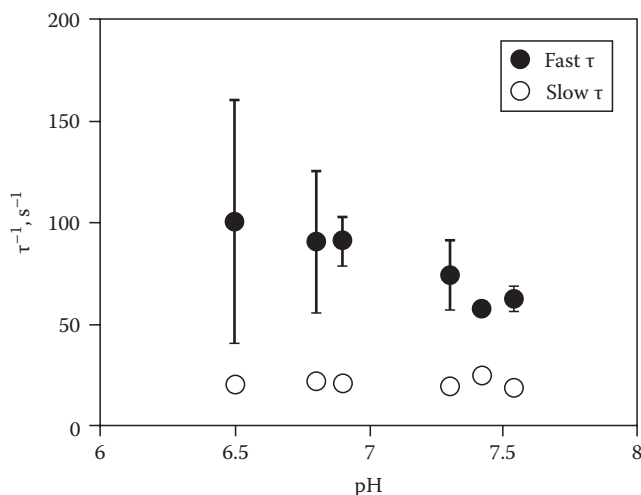
FIGURE 13.11 τ^{-1} Values determined from p-jump experiments for arsenate adsorption/desorption on goethite, as a function of pH. (Reprinted from Grossl, P.R., M. Eick, D.L. Sparks, S. Goldberg, and C.C. Ainsworth. 1997. Arsenate and chromate retention mechanisms on goethite. 2. Kinetic evaluation using a pressure-jump relaxation technique. Environ. Sci. Technol. 31:321–326. With permission from American Chemical Society.)
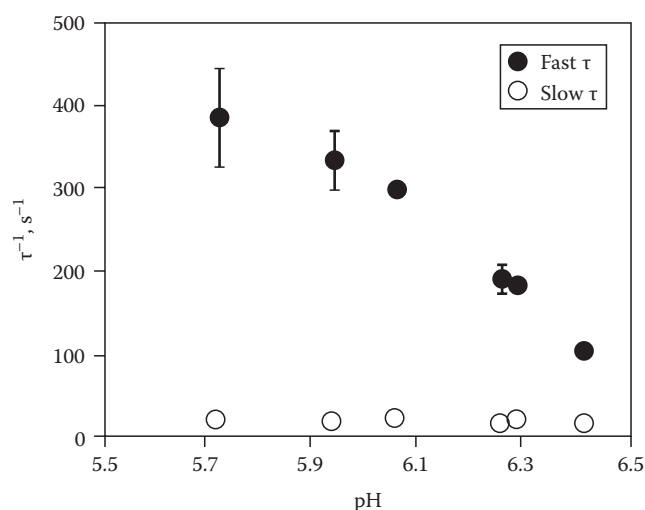
FIGURE 13.12 τ^{-1} Values determined from p-jump experiments for chromate adsorption/desorption on goethite, as a function of pH. (Reprinted from Grossl, P.R., M. Eick, D.L. Sparks, S. Goldberg, and C.C. Ainsworth. 1997. Arsenate and chromate retention mechanisms on goethite. 2. Kinetic evaluation using a pressure-jump relaxation technique. Environ. Sci. Technol. 31:321–326. With permission from American Chemical Society.)
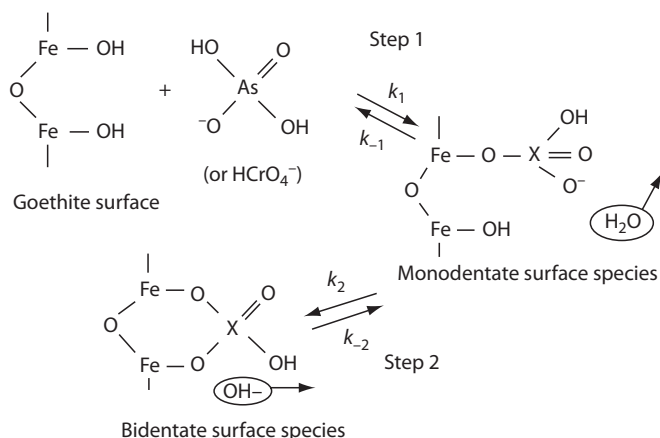
FIGURE 13.13 Proposed mechanism for oxyanion adsorption/desorption on goethite. The X represents either As(V) or Cr(VI). (Reprinted from Grossl, P.R., M. Eick, D.L. Sparks, S. Goldberg, and C.C. Ainsworth. 1997. Arsenate and chromate retention mechanisms on goethite. 2. Kinetic evaluation using a pressure-jump relaxation technique. Environ. Sci. Technol. 31:321–326. Copyright American Chemical Society.)

From the plots in Figures 13.14 and 13.15, forward and reverse rate constants were obtained for the sorption and desorption reactions of both the monodentate and bidentate steps where k_1 = slope (Figure 13.14); k_{-1} = intercept (Figure 13.14) − slope (Figure 13.15)/slope (Figure 13.14); k_2 = intercept (Figure 13.14) − k_{-1} − k_{-2}; and k_{-2} = intercept (Figure 13.15)/k_{-1}. The calculated rate constants for both chromate and arsenate adsorption/desorption on goethite are listed in Table 13.3.

Overall, the forward rate constants associated with the formation of the inner-sphere oxyanion/goethite surface complexes were more rapid than the reverse rate constants. Therefore, the
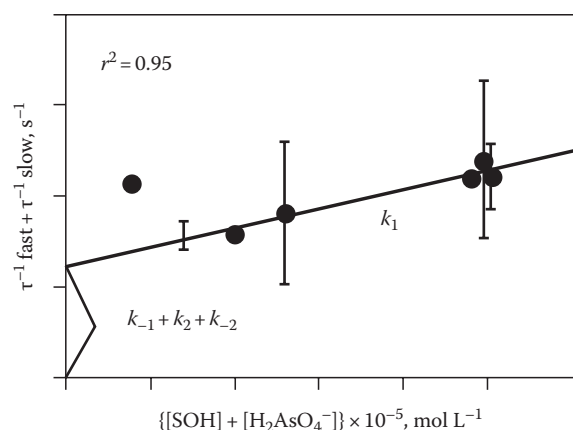
FIGURE 13.14 Evaluation of the linearized rate Equation 13.14 for the mechanism displayed in Figure 13.13 for arsenate. (Reprinted from Grossl, P.R., M. Eick, D.L. Sparks, S. Goldberg, and C.C. Ainsworth. 1997. Arsenate and chromate retention mechanisms on goethite. 2. Kinetic evaluation using a pressure-jump relaxation technique. Environ. Sci. Technol. 31:321–326. With permission from American Chemical Society.)
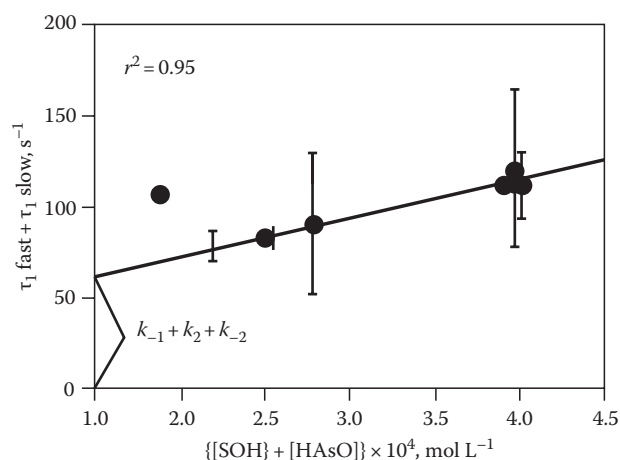
FIGURE 13.15 Evaluation of the linearized rate Equation 13.15 for the mechanism displayed in Figure 13.13 for arsenate. (Reprinted from Grossl, P.R., M. Eick, D.L. Sparks, S. Goldberg, and C.C. Ainsworth. 1997. Arsenate and chromate retention mechanisms on goethite. 2. Kinetic evaluation using a pressure-jump relaxation technique. Environ. Sci. Technol. 31:321–326. With permission from American Chemical Society.)

rate-limiting steps were the reverse reactions. The equilibrium constants listed in Table 13.3 were calculated using the rate constants for each reaction step in the proposed mechanism (Figure 13.13) from the following relationship:

$$K_{eq} = \frac{k_1}{k_{-1}}. \tag{13.16}$$

The calculated equilibrium constant for Step 1 for arsenate was $10^{5.35}$ and for Step 2 was $10^{0.26}$, while the calculated K_{eq} for Step 1 for chromate was $10^{3.7}$ and for Step 2 was $10^{0.4}$. The sorption of both oxyanions and subsequent formation of inner-sphere surface complexes are thermodynamically favorable, with the exception of the equilibrium constant for the second step associated

TABLE 13.3 Calculated Rate Constants for Chromate and Arsenate Adsorption/Desorption on Goethite

	Step 1	Step 2
Arsenate	$k_1 = 10^{63}$ L mol^{-1} s^{-1}	$k_2 = 15\,s^{-1}$
	$k_{-1} = 8\,s^{-1}$	$k_{-2} = 8\,s^{-1}$
	$k_{eq} = 10^{5.35}$ L mol^{-1} s^{-1}	$k_{eq} = 10^{0.26}$
Chromate	$k_1 = 10^{5.8}$ L mol^{-1} s^{-1}	$k_2 = 16\,s^{-1}$
	$k_{-1} = 129\,s^{-1}$	$k_{-2} = 38\,s^{-1}$
	$k_{eq} = 10^{3.7}$ L mol^{-1} s^{-1}	$k_{eq} = 10^{-0.4}$

Source: Reprinted from Grossl, P.R., M. Eick, D.L. Sparks, S. Goldberg, and C.C. Ainsworth. 1997. Arsenate and chromate retention mechanisms on goethite. 2. Kinetic evaluation using a pressure-jump relaxation technique. Environ. Sci. Technol. 31:321–326. With permission from American Chemical Society.

with chromate sorption (slightly less than 1). Thus, the monodentate chromate/goethite surface complex is slightly favored over the bidentate surface complex. This is in agreement with spectroscopic data obtained from x-ray absorption fine structure (XAFS) analyses (Fendorf et al., 1997), which indicate a mixture of both monodentate and bidentate arsenate, and chromate surface complexes; but at low-surface coverage, a greater proportion of chromate is associated with the monodentate complex than the bidentate complex. The results from both kinetic and XAFS experiments suggest that arsenate is more likely to form an inner-sphere surface complex with goethite than chromate.

While p-jump relaxation techniques are useful for measuring rapid reaction rates on soil components, the rate constants are calculated from linearized rate equations, which include parameters that are determined from equilibrium and modeling studies. Consequently, the rate "constants" are not directly determined.

The ideal way to measure rapid reaction rates that may often comprise a major portion of the total reaction process for many soil chemical processes, particularly on soil components is to use real-time molecular scale techniques such as ATR–FTIR and QXAS. With these techniques, one can not only measure reaction rates on millisecond timescales but also at the same time couple these measurements with spectroscopic analyses of the reaction mechanism (Parikh et al., 2008; Ginder-Vogel et al., 2009; Landrot et al., 2010).

While the initial sorption of heavy metals is rapid, with the CR step occurring on second or millisecond timescales, further sorption may be slow (Figure 13.16) occurring over timescales of days and longer. Such behavior has been observed for some Fe oxides. This slow sorption has been ascribed to several mechanisms, including interparticle or intraparticle diffusion in pores and solids, sites of low reactivity, and surface precipitation/nucleation (Sparks, 1998; Strawn and Sparks, 2000). With heterogeneous soils and even soil components, there may be a continuum between the three sorption mechanisms, for example, between adsorption and surface precipitation/nucleation. While it has generally been assumed that adsorption in comparison with surface precipitation/nucleation is much more rapid, recent
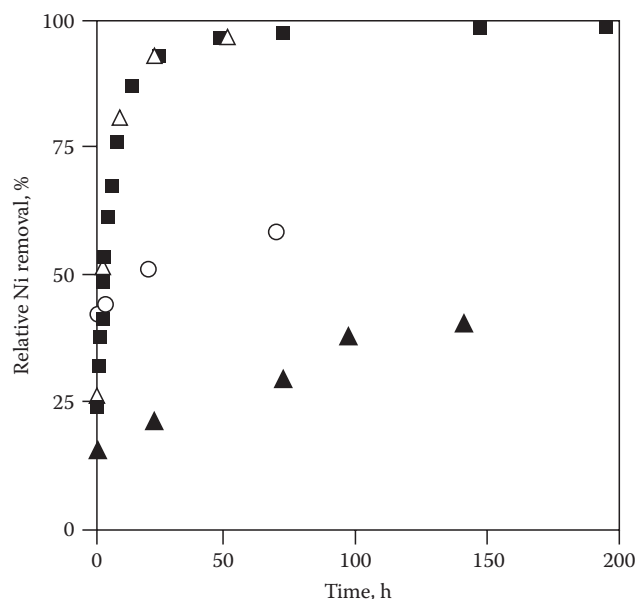
FIGURE 13.16 Kinetics of Ni sorption [%] on pyrophyllite (■), kaolinite (△), gibbsite (▲), and montmorillonite (○) from a 3 mM Ni solution at pH = 7.5 and an ionic strength $I = 0.1$M (NaNO$_3$). The last sample of each experiment was collected and analyzed by XAFS. (Reprinted from Scheidegger, A.M., G.M. Lamble, and D.L. Sparks. 1997. Spectroscopic evidence for the formation of mixed-cation hydroxide phases upon metal sorption on clays and aluminum oxides. J. Colloid Interface Sci. 186:118–128. With permission from Elsevier.)

studies (Scheidegger et al., 1998) that will be discussed later, have shown that surface precipitation/nucleation processes can occur on timescales as short as 15 min, which indicates that sorption and nucleation processes can occur simultaneously. However, in some cases, depending on reaction conditions and the metal involved, a particular sorption mechanism can dominate.

Obviously, an important factor affecting the degree of slow sorption/desorption of metals (and for that matter also of organic chemicals) is the time period the sorbate has been in contact with the sorbent (residence time). Bruemmer et al. (1988) studied Ni^{2+}, Zn^{2+}, and Cd^{2+} sorption on goethite, a porous Fe oxide that has defect structures in which metals can be incorporated to satisfy charge imbalances. Bruemmer et al. (1988) found, at pH 6, that as reaction time increased from 2 h to 42 days (at 293 K), adsorbed Ni^{2+} increased from 12% to 70% of total adsorption, and total Zn^{2+} and Cd^{2+} adsorption over this time increased by 33% and 21%, respectively. The kinetic reactions could be well described using a Fickian diffusion model. Metal uptake was hypothesized to occur by a three-step mechanism: (1) adsorption of metals on external surfaces, (2) solid-state diffusion of metals from external to internal sites, and (3) metal binding and fixation at positions inside the goethite particle.

Ainsworth et al. (1994) studied the adsorption/desorption of Co^{2+}, Cd^{2+}, and Pb^{2b+} on hydrous Fe(III) as a function of aging and metal oxide residence time. Oxide aging did not cause hysteresis of metal cation sorption/desorption. Aging the oxide with the metal cations resulted in hysteresis with Cd^{2+} and Co^{2+} but little with
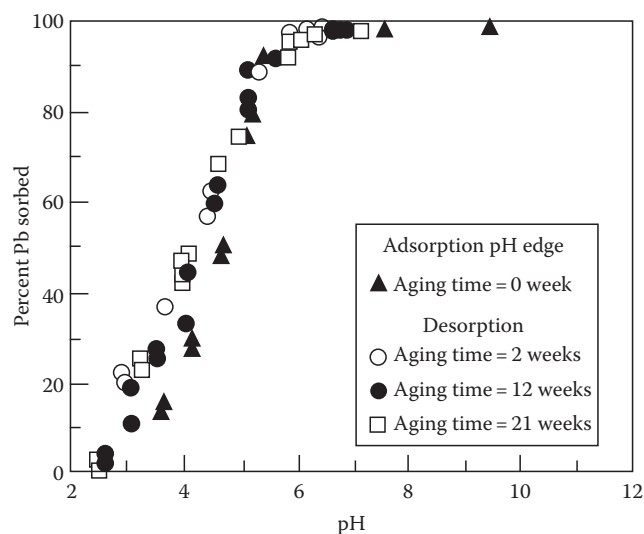
FIGURE 13.17 Fractional sorption–desorption of Pb^{2+} to hydrous Fe(III) as a function of pH and HFO–Pb^{2+} aging time. (Reprinted from Ainsworth, C.C., J.L. Pilon, P.L. Gassman, and W.G. Van Der Sluys. 1994. Cobalt, cadmium, and lead sorption to hydrous iron oxide: Residence time effect. Soil Sci. Soc. Am. J. 58:1615–1623. With permission from the Soil Science Society of America.)

Pb^{2+}. With Pb^{2+} between pH 3 and 5.5 there was slight hysteresis over a 21 week aging process (hysteresis varied from <2% difference between sorption and desorption to ~10%). At pH 2.5, Pb^{2+} desorption was complete within a 16 h period and was not affected by aging time (Figure 13.17). However, with Cd^{2+} and Co^{2+}, extensive hysteresis was observed over a 16 week aging period and the hysteresis increased with aging time (Figures 13.18 and 13.19). After 16 weeks of aging, 20% of the Cd^{2+} and 53% of the Co^{2+} were not desorbed, and even at pH 2.5, hysteresis was observed. The extent
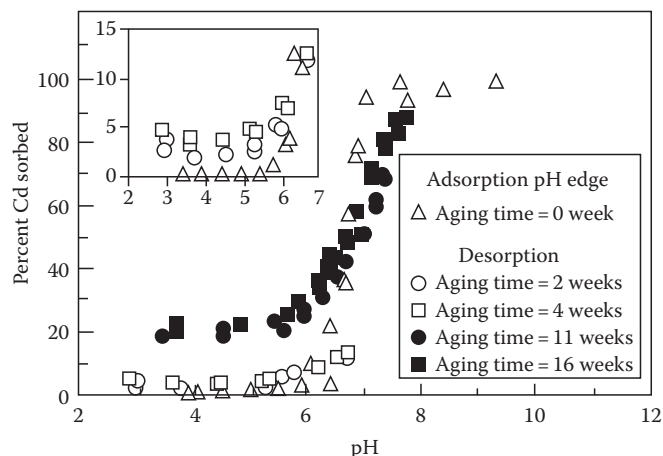
FIGURE 13.18 Fractional sorption–desorption of Cd^{2+} to hydrous Fe oxide (HFO) as a function of pH and HFO–Cd^{2+} aging time; insert shows adsorption–desorption of Cd^{2+} to HFO at 2 and 4 week aging times. (Reprinted from Ainsworth, C.C., J.L. Pilon, P.L. Gassman, and W.G. Van Der Sluys. 1994. Cobalt, cadmium, and lead sorption to hydrous iron oxide: Residence time effect. Soil Sci. Soc. Am. J. 58:1615–1623. With permission from the Soil Science Society of America.)
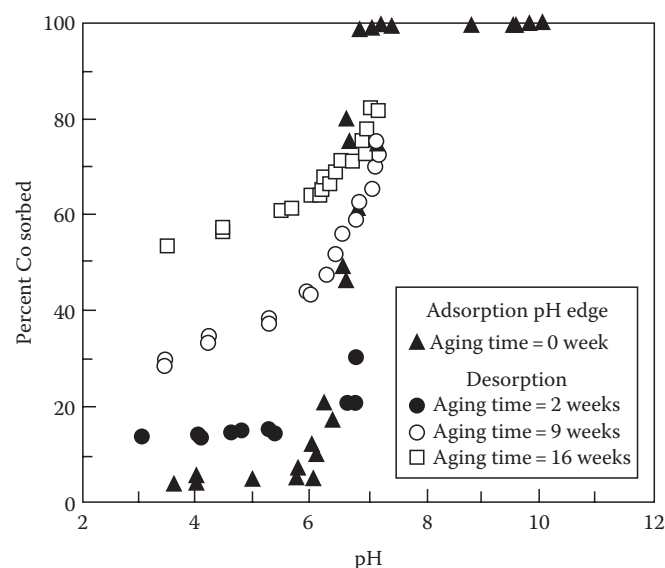
FIGURE 13.19 Fractional adsorption of Co^{2+} to hydrous Fe oxide (HFO) as a function of pH and HFO–Co^{2+} aging time. (Reprinted from Ainsworth, C.C., J.L. Pilon, P.L. Gassman, and W.G. Van Der Sluys. 1994. Cobalt, cadmium, and lead sorption to hydrous iron oxide: Residence time effect. Soil Sci. Soc. Am. J. 58:1615–1623. With permission from the Soil Science Society of America.)

of reversibility with aging for Co^{2+}, Cd^{2+}, and Pb^{2+} was inversely proportional to the ionic radius of the ions, namely, $Co^{2+} < Cd^{2+} < Pb^{2+}$. Ainsworth et al. (1994) attributed the hysteresis to Co and Cd incorporation into a recrystallizing solid (probably goethite) by isomorphic substitution and not to micropore diffusion.

Strawn and Sparks (2000) investigated the role that residence time and organic matter content had on Pb desorption kinetics from a Matapeake soil. Residence time had little effect on the amount of Pb desorbed, but marked hysteresis was observed at all residence times, which was ascribed to diffusion of Pb into the organic matter and strong metal–soil complexes.

Fuller et al. (1993) combined kinetic sorption and desorption experiments with spectroscopic observations (Waychunas et al., 1993) to study As sorption on ferrihydrite. Using XAFS spectroscopy, they found that As was sorbed predominantly as inner-sphere bidentate complexes, regardless of whether the As was adsorbed after the mineralization of the ferrihydrite or it was present during precipitation. No As surface precipitates were observed. Slow As sorption and desorption were explained as slow diffusion of the As to or from interior surface complexation sites that exist within disordered aggregates of crystallites. The kinetic reactions could be described using a Fickian diffusion model.

Slow metal sorption/desorption has also been ascribed to conversion of the metal sorbate from a high- to a low-energy state (Kuo and Mikkelsen, 1980; Padmanabham, 1983; Schultz et al., 1987; Backes et al., 1995). Lehmann and Harter (1984) measured the kinetics of chelate-promoted Cu^{2+} release from a soil to assess the strength of the bond formed. Sorption/desorption was biphasic, which was attributed to high- and low-energy bonding sites.

With increased residence time from 30 min to 24 h, Lehmann and Harter(1984) speculated that there was a transition of Cu from low to higher energy sites (as evaluated by release kinetics). Incubations for up to 4 days showed a continued uptake of Cu and a decrease in the fraction released within the first 3 min, which was referred to as the low-energy sorbed fraction.

Recent studies particularly using surface spectroscopic and microscopic techniques such as XAS and transmission electron microscopy (TEM) have shown that the formation of polynuclear surface species (e.g., surface precipitates) on natural materials such as clay minerals, metal oxides, and in soils is an important sorption and sequestration mechanism (Charlet and Manceau, 1993; Fendorf et al., 1994; Junta and Hochella, 1994; O'Day et al., 1994a, 1994b; Wersin et al., 1994; Scheidegger et al., 1996, 1997, 1998; Towle et al., 1997; Xia et al., 1997; Elzinga and Sparks, 1999; Roberts et al., 1999; Thompson et al., 1999a, 1999b; Ford and Sparks, 2000; Nachtegaal et al., 2005; McNear et al., 2007).

The surface precipitates of Co, Cr(III), Cu, and Ni on metal oxides and clay minerals have been observed at metal surface loadings far below a theoretical monolayer coverage and in a pH range well below the pH where the formation of metal hydroxide precipitates would be expected according to the thermodynamic solubility product. In Al-bearing soil components, the precipitates are mixed metal–Al hydroxide phases of the layered double hydroxide type. Recent studies using microfocused x-ray fluorescence (micro-XRF) and x-ray absorption spectroscopy (micro-XAS) have shown that Zn–Al layered double hydroxides (LDH) (Nachtegaal et al., 2005) and Ni–Al hydroxide LDH (McNear et al., 2007) precipitates occur in field-contaminated soils.

In general, the formation of surface precipitates has been considered a slow phenomenon. However, Scheidegger et al. (1998) have demonstrated that the rate of surface precipitate formation can be quite rapid. The appearance of surface precipitates during sorption of Ni to pyrophyllite at pH 7.5 occurred over time scales less than 1 h. Similar results were observed for Ni sorption to kaolinite. However, the kinetics of Ni sorption onto gibbsite, and subsequent surface precipitate formation, were slower than for pyrophyllite (Figures 13.16 and 13.20). Time-resolved XAFS studies demonstrated continued growth of the surface precipitate during Ni uptake as the structure and Ni:Al stoichiometry of the sorption complex approached that of a takovite-like phase (Scheidegger et al., 1998).

13.5.1.2 Organic Contaminants

Numerous studies on the kinetics of organic chemical sorption/desorption reactions on soils and soil components have shown that sorption/desorption is characterized by a rapid, reversible stage followed by a much slower, nonreversible stage (Karickhoff et al., 1979; DiToro and Horzempa, 1982; Karickhoff and Morris, 1985; Kan et al., 1997; Xing and Pignatello, 2005), or biphasic kinetics. The rapid phase has been ascribed to retention of the organic chemical in a labile form that is easily desorbed. However, the much slower reaction phase involves the entrapment of the chemical in a nonlabile form that is difficult to desorb. The labile form of the
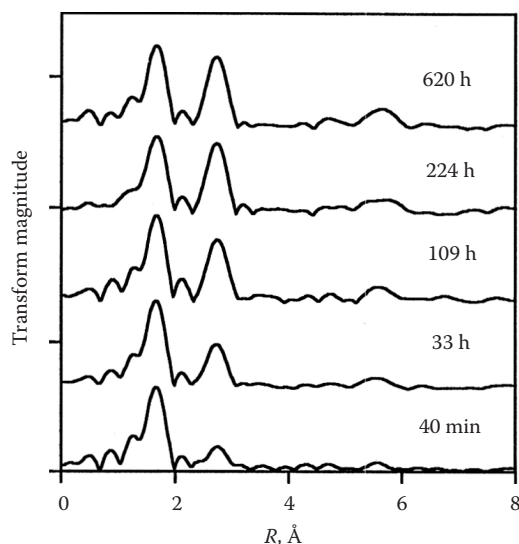
FIGURE 13.20 Radial structure functions for Ni sorption to gibbsite for reaction times up to 620h demonstrating the appearance and growth of second shell contributions due to surface precipitation and growth of a hydrotalcite-like phase. (Reprinted from Scheidegger, A.M., D.G. Strawn, G.M. Lamble, and D.L. Sparks. 1998. The kinetics of mixed Ni–Al hydroxide formation on clay and aluminum oxide minerals: A time-resolved XAFS study. Geochim. Cosmochim. Acta 62:2233–2245. With permission from Elsevier.)

chemical is available for microbial attack, while the nonlabile portion is resistant to biodegradation.

This slower sorption/desorption reaction has been ascribed to intraparticle and interparticle diffusion of the chemical into soil organic matter (SOM) and inorganic soil components (Wu and Gschwend, 1986; Steinberg et al., 1987; Ball and Roberts, 1991).

Theories proposed by a number of researchers (Weber and Huang, 1996; Xing and Pignatello, 1997; Pignatello, 2000) have explained the slow diffusion of organic chemicals into SOM by considering SOM as a combination of "rubbery" and "glassy" polymers. The rubber-like phases are characterized by an expanded, flexible, and highly solvated structure with pores of subnanometer dimensions (holes) (Xing and Pignatello, 1997; Pignatello, 1998, 2000). Sorption in the rubbery phases results in linear, noncompetitive, and reversible behavior. The glassy phases have pores that are of subnanometer size, and sorption in this phase is characterized by nonlinearity and is competitive (Xing and Pignatello, 1997). These theories, relating slow diffusion into organic matter to diffusion in polymers, are somewhat validated in some studies measuring energies of activation (E_a) for organic chemical sorption. Cornelissen et al. (1997) studied the temperature dependence of slow adsorption and desorption kinetics of some chlorobenzenes, polychlorinated biphenyls (PCBs), and polycyclic aromatic hydrocarbons (PAHs) in laboratory- and field-contaminated sediments and obtained E_a values of 60–70kJ mol^{-1}, which are in the range for diffusion in polymers. These values are much higher than those for pore diffusion (20–40kJ mol^{-1}) suggesting that intraorganic matter diffusion may be a more important mechanism for slow organic chemical sorption than interparticle pore diffusion.

However, Chang et al. (1997) who studied sorption of toluene (TOL), *n*-hexane, and acetone on pressed humic acid disks found an insignificant amount of irreversibly bound residue with activation energies in the range of 42.3–65.8kJ mol^{-1}, suggesting a physical sorption process with little diffusion.

An example of the biphasic kinetics that is observed for many organic CRs in soils/sediments is shown in Figure 13.21. In this study, 55% of the labile PCB was desorbed from sediments in
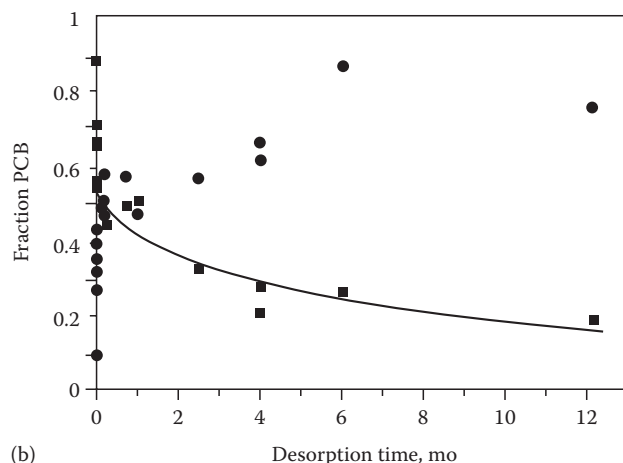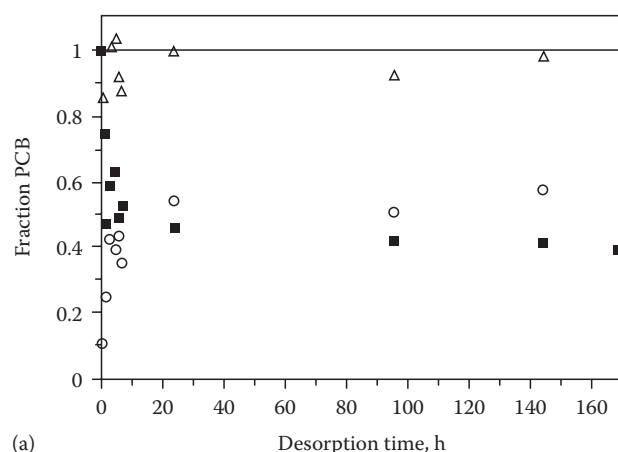
FIGURE 13.21 (a) Short-term polychlorinated biphenyl (PCB) desorption in hours (h) from Hudson River sediment contaminated with 25 mg kg^{-1} PCB. Distribution of the PCB between the sediment (v) and XAD-4 resin (o) as well as the overall mass balance (D) is shown. The resin acts as a sink to retain the PCB that is desorbed (Carroll et al., 1994). (b) Long-term PCB desorption in months (mo) from Hudson River sediment contaminated with 25 mg kg^{-1} PCB. Distribution of the PCB between the sediment (■) and XAD-4 resin (●) is shown. The line represents a nonlinear regression of the data by a two-site model. (Reprinted from Carroll, K.M., M.R. Harkness, A.A. Bracco, R.R. Balcarcel. 1994. Application of a permeant/polymer diffusional model to the desorption of polychlorinated biphenyls from Hudson River sediments. Environ. Sci. Technol. 28:253–258. With permission from American Chemical Society.)

a 24 h period, while little of the remaining 45% nonlabile fraction was desorbed in 170 h (Figure 13.21a). Over another 1 year period, ~50% of the remaining nonlabile fraction desorbed (Figure 13.21b).

Pavlostathis and Mathavan (1992) observed a biphasic desorption process for field soils contaminated with trichloroethylene (TCE), tetrachloroethylene (PCE), TOL, and xylene (XYL). A fast desorption reaction occurred in 24 h followed by a much slower desorption reaction beyond 24 h. In 24 h, 9%–29%, 14%–48%, 9%–40%, and 4%–37% of the TCE, PCE, TOL, and XYL, respectively, were released. However, the apparent irreversibility or hysteresis may be an artifact caused by not reaching a true sorption equilibrium. For example, DiVincenzo and Sparks (1997), studying pentachlorophenol sorption/desorption on a soil, found that if desorption was initiated after an apparent sorption equilibrium (i.e., slow sorption was measured) was reached, hysteresis or irreversibility was significantly reduced.

A number of studies have also shown that with aging the nonlabile portion of the organic chemical in the soil/sediment becomes more resistant to release (McCall and Agin, 1985; Steinberg et al., 1987; Pignatello and Huang, 1991; Pavlostathis and Mathavan, 1992; Scribner et al., 1992; Alexander, 1995; Loehr and Webster, 1996). However, Connaughton et al. (1993) did not observe the nonlabile fraction increasing with age for naphthalene-contaminated soils.

One way to gauge the effect of time on organic contaminant retention in soils is to compare K_d (sorption distribution coefficient) values for freshly aged and aged soil samples. In most studies, K_d values are measured based on a 24 h equilibration between the soil and the organic chemical. When these values are compared to K_d values for field soils previously reacted with the organic chemical (aged samples), the latter have much higher K_d values, indicating that much more of the organic chemical is in a sorbed state. For example, Pignatello and Huang (1991) measured K_d values in freshly aged (K_d) and "aged" soils (K_{app}, apparent sorption distribution coefficient) reacted with atrazine and metolachlor, two widely used herbicides. The aged soils had been treated with the herbicides 15–62 months before sampling. The K_{app} values ranged from 2.3 to 42 times higher than the K_d values (Table 13.4).

Scribner et al. (1992) studying simazine (a widely used triazine herbicide for broadleaf and grass control in crops) desorption and bioavailability in aged soils found that K_{app} values were 15 times higher than K_d values. Scribner et al. (1992) also showed that 48% of the simazine added to the freshly aged soils was biodegradable over a 34-day incubation period while none of the simazine in the aged soil was biodegraded.

One of the implications of these results is that while many transport and degradation models for organic contaminants in soils and waters assume that the sorption process is an equilibrium process, the above studies clearly show that kinetic reactions must be considered when making predictions about the mobility and fate of organic chemicals. Moreover, calculation of K_d values based on a 24 h equilibration period, which are commonly used in fate and risk assessment models, can be

TABLE 13.4 Sorption Distribution Coefficients for Herbicides in Freshly Aged and Aged Soils

Herbicide	Soil	K_d (L kg^{-1})[a]	K_{app} (L kg^{-1})[b]
Metolachlor	Cva	2.96	39
	CVb	1.46	27
	W1	1.28	49
	W2	0.77	33
Atrazine	Cva	2.17	28
	CVb	1.32	29
	W3	1.75	4

Source: Adapted from Pignatello, J.J., and L.Q. Huang. 1991. Sorptive reversibility of atrazine and metolachlor residues in field soil samples. J. Environ. Qual. 20:222–228. With permission of the American Society of Agronomy.

[a] Sorption distribution coefficient (L kg^{-1}) of freshly aged soil based on a 24 h equilibration period.

[b] Apparent sorption distribution coefficient (L kg^{-1}) in contaminated soil (aged soil) determined using a 24 h equilibration period.

inaccurate since 24 h K_d values often overestimate the amount of organic chemical in the solution phase.

The finding that many organic chemicals are quite persistent in the soil environment has both good and bad features. The beneficial aspect is that the organic chemicals are less mobile and may not be readily transported in groundwater supplies. The negative aspect is that their persistence and inaccessibility to microbes may make decontamination more difficult, particularly if in situ remediation techniques such as biodegradation are employed.

13.5.2 Kinetics of Mineral Dissolution

13.5.2.1 Rate-Limiting Steps

Dissolution of minerals involves several steps (Stumm and Wollast, 1990): (1) mass transfer of dissolved reactants from the bulk solution to the mineral surface, (2) adsorption of solutes, (3) interlattice transfer of reaction species, (4) surface CRs, (5) removal of reactants from the surface, and (6) mass transfer of products into the bulk solution. Under field conditions, mineral dissolution is slow and mass transfer of reactants or products in the aqueous phase (Steps 1 and 6) is not rate limiting. Thus, the rate-limiting steps are either transport of reactants and products in the solid phase (Step 3) or surface CRs (Step 4) and removal of reactants from the surface (Step 5).

Transport-controlled dissolution reactions or those controlled by mass transfer or diffusion can be described using a parabolic rate law given later (Stumm and Wollast, 1990):

$$r = \frac{dC}{dt} = kt^{-1/2}, \qquad (13.17)$$

where

r is the reaction rate
C is the concentration in solution
t is time
k is the reaction rate constant

Integrating, C increases with $t^{1/2}$,

$$C = C_0 + 2kt^{1/2}, \qquad (13.18)$$

where C_0 is the initial concentration in solution.

If the surface reactions are slow compared to the transport reactions, dissolution is surface controlled, which is the case for most dissolution reactions of silicates and oxides. In surface-controlled reactions, the concentrations of solutes next to the surface are equal to the bulk solution concentrations and the dissolution kinetics are zero order if steady-state conditions are operational on the surface. Thus, the dissolution rate (r) is

$$r = \frac{dC}{dt} = kA, \qquad (13.19)$$

and r is proportional to the mineral's surface area, A. Thus, for a surface-controlled reaction, the relationship between t and C should be linear. Figure 13.22 compares transport- and surface-controlled dissolution mechanisms.

Intense arguments have ensued over the years concerning the mechanism for mineral dissolution. Those that supported a transport-controlled mechanism believed that a leached layer formed as mineral dissolution proceeded and that subsequent dissolution took place by diffusion through the leached layer (Wollast, 1967; Petrovic et al., 1976). Advocates of this theory found that dissolution was described by the parabolic rate law (Equation 13.18). However, the apparent transport-controlled kinetics may be an artifact caused by dissolution of hyperfine particles formed on the mineral surfaces after grinding that are highly reactive sites or by the use of batch methods that cause reaction products to accumulate causing precipitation of secondary minerals. These experimental artifacts can cause incongruent reactions and pseudoparabolic kinetics. Studies employing x-ray photoelectron spectroscopy (XPS) and nuclear resonance profiling (Schott and Petit, 1987; Casey et al., 1989) have demonstrated
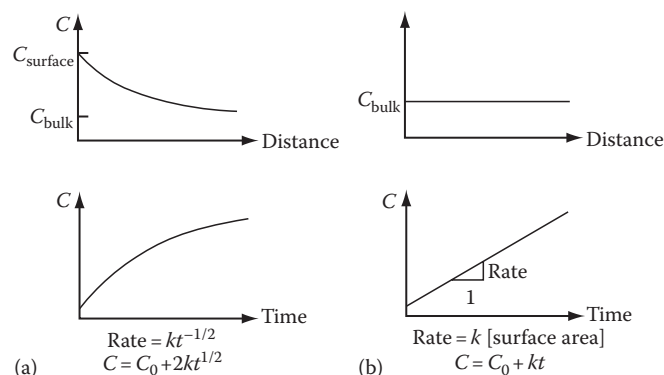
FIGURE 13.22 (a) Transport- versus (b) surface-controlled dissolution. Schematic representation of concentration in solution, C, as a function of distance from the surface of the dissolving mineral. In the lower part of the figure, the change in concentration is given as a function of time. (Reprinted with permission from Stumm, W. 1992. Chemistry of the solid–water interface. Wiley, New York.)
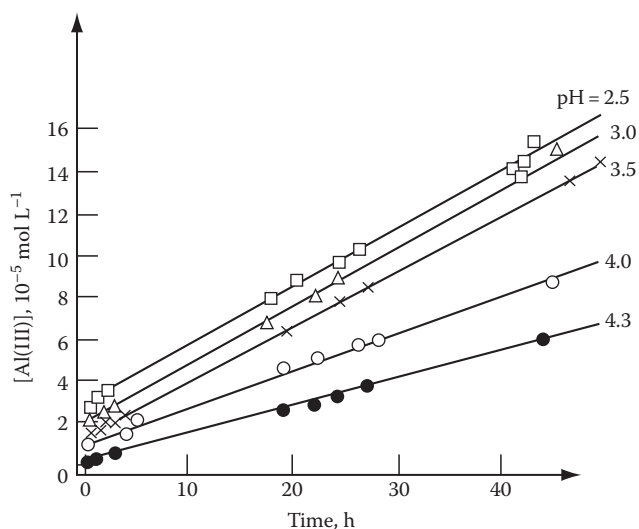
FIGURE 13.23 Linear dissolution kinetics observed for the dissolution of γ-Al$_2$O$_3$. Representative of processes whose rates are controlled by a surface reaction and not by transport. (Reprinted from Furrer, G., and W. Stumm. 1986. The coordination chemistry of weathering. I. Dissolution kinetics of γ-Al$_2$O$_3$ and BeO. Geochim. Cosmochim. Acta 50:1847–1860. Copyright 1986. With permission from Elsevier.)

that although some incongruency may occur in the initial dissolution process, which may be diffusion controlled, the overall reaction is surface controlled. Energies of activation from 60 to 86 kJ mol^{-1} have been observed for dissolution of oxides and silicates, further suggesting surface-controlled dissolution (Lasaga, 1984; Jordan and Rammensee, 1996). An illustration of the surface-controlled dissolution of γ-Al$_2$O$_3$ resulting in a linear release of Al^{3+} with time is shown in Figure 13.23. The dissolution rate (r) can be obtained from the slope of Figure 13.23.

Scanning force microscopy (SFM), which has also been used increasingly as an in situ technique for imaging mineral surfaces immersed in aqueous solution during the course of dissolution (Hellman et al., 1992; Hillner et al., 1992a, 1992b; Johnsson et al., 1992; Bosbach and Rammensee, 1994; Maurice et al., 1995), permits a direct measure of surface-controlled dissolution rates by providing 3D data on changes in microtopography. In situ SFM has the unique ability to detect separate processes, such as dissolution and secondary phase formation, occurring simultaneously on a mineral surface (Maurice, 1998).

13.5.2.2 Surface-Controlled Dissolution Mechanisms

Dissolution of oxide minerals through a surface-controlled reaction by ligand- and proton-promoted processes has been described by Furrer and Stumm (1986), Zinder et al. (1986), and Stumm and Furrer (1987) using a surface coordination approach. The important reactants in these processes are H$_2$O, H$^+$, OH, ligands, and reductants and oxidants. The reaction mechanism occurs in two steps (Stumm and Wollast, 1990):

$$\text{Surface sites} + \text{reactants}\,(\text{H}^+, \text{OH}^-, \text{ligands}) \xrightarrow{\text{fast}} \text{surface species},$$

$$(13.20)$$
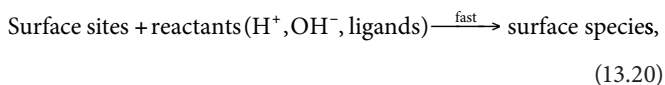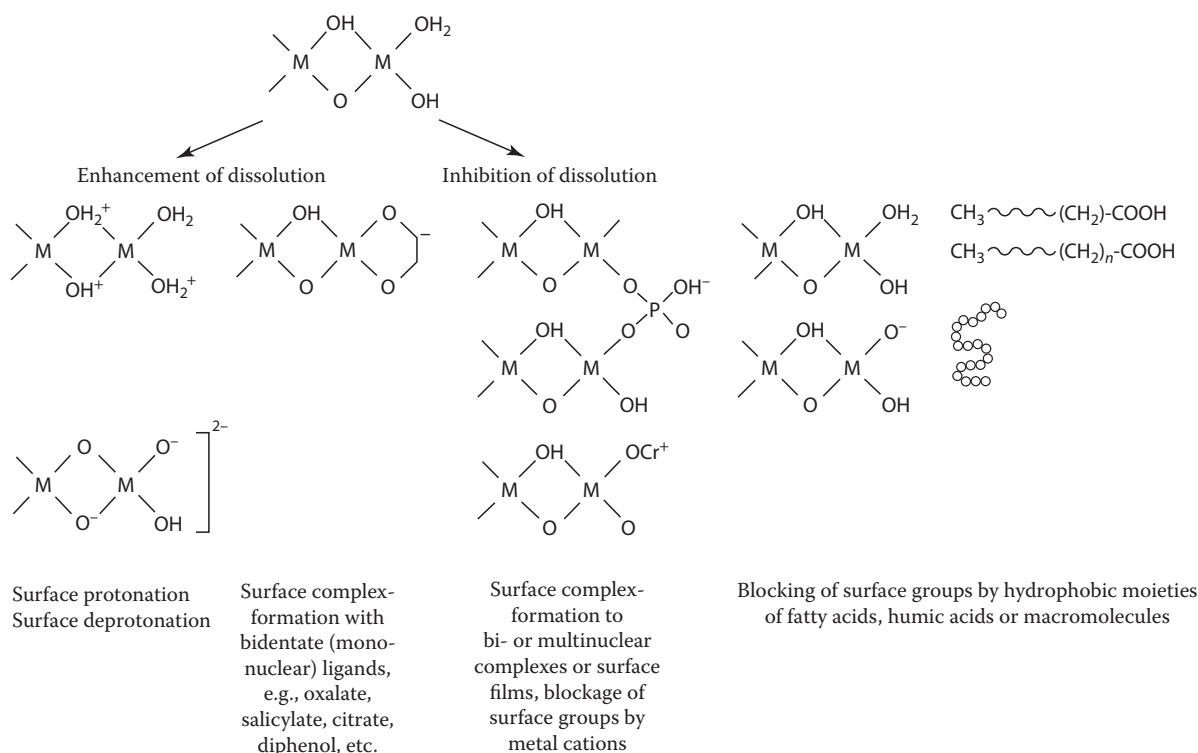
FIGURE 13.24 The dependence of surface reactivity and of kinetic mechanisms on the coordinative environment of the surface groups. (Reprinted from Stumm, W., and R. Wollast. 1990. Coordination chemistry of weathering. Kinetics of the surface-controlled dissolution of oxide minerals. Rev. Geophys. 28:53–69. With permission from American Geophysical Union.)

$$\text{Surface species} \xrightarrow[\text{detachment of metal(M)}]{\text{slow}} \text{M(aq).} \qquad (13.21)$$
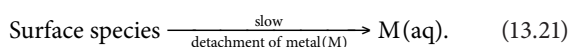
Thus, the attachment of the reactants to the surface sites is fast, and detachment of metal species from the surface into solution is slow and rate limiting.

13.5.2.3 Ligand-Promoted Dissolution

Figure 13.24 shows how the surface chemistry of the mineral affects dissolution. One sees that surface protonation of the surface ligand increases dissolution by polarizing interatomic bonds close to the central surface ions that promote the release of a cation surface group into solution. Hydroxyls that bind to surface groups at higher pHs can ease the release of an anionic surface group into the solution phase.

Ligands that form surface complexes by ligand exchange with a surface hydroxyl add negative charge to the Lewis acid center coordination sphere and lower the Lewis acid acidity. This polarizes the M–O bonds causing detachment of the metal cation into the solution phase. Thus, inner-sphere surface complexation plays an important role in mineral dissolution. Ligands such as oxalate, salicylate, F^-, EDTA, and NTA increase dissolution, but others, such as SO_4^{2-}, CrO_4^{2-}, and benzoate, inhibit dissolution. Phosphate, arsenate, and selenite enhance dissolution at low pH, and dissolution is inhibited at pH > 7 (Bondietti et al., 1993).

The reason for these differences may be that bidentate species that are mononuclear promote dissolution while binuclear bidentate species inhibit dissolution. With binuclear bidentate complexes, more energy may be needed to remove two central atoms from the crystal structure. With phosphate and arsenate, at low pH, mononuclear species are formed while at higher pH (~pH 7), binuclear or trinuclear surface complexes form. Mononuclear bidentate complexes are formed with oxalate while binuclear bidentate complexes form with CrO_4^{2-}. Additionally, the electron donor properties of CrO_4^{2-} and oxalate are also different. With CrO_4^{2-}, a high redox potential is maintained at the oxide surface, which restricts reductive dissolution (Stumm and Wollast, 1990; Stumm, 1992). Dissolution can also be inhibited by cations such as VO^{2+}, Cr(III), and Al(III) that block surface functional groups (Bondietti et al., 1993).

One can express the rate of the ligand-promoted dissolution (R_L) as

$$R_L = k_L' (\equiv ML) = k_L' C_L^s, \qquad (13.22)$$

where

k_L' is the rate constant for ligand-promoted dissolution (t^{-1})
ML is the metal–ligand complex
C_L^s is the surface concentration of the ligand (mol m^{-2})

Equation 13.22 adequately describes ligand-promoted dissolution of γ-Al$_2$O$_3$ (Figure 13.25).
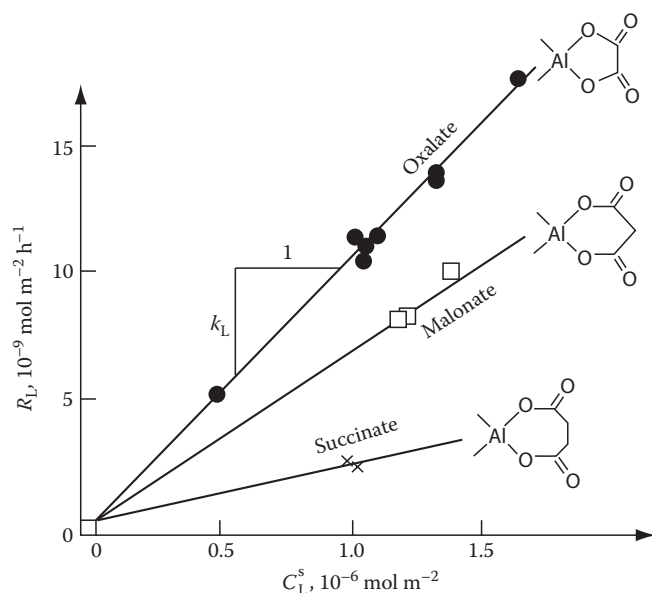
FIGURE 13.25 The rate of ligand catalyzed dissolution of γ-Al$_2$O$_3$ by the aliphatic ligands oxalate, malonate, and succinate, R_L (nmol m^{-2} h^{-1}), can be interpreted as a linear dependence on the surface concentrations of the ligand complexes, C_L^s. In each case, the individual values for C_L^s were determined experimentally. (Reprinted from Furrer, G., and W. Stumm. 1986. The coordination chemistry of weathering. I. Dissolution kinetics of γ-Al$_2$O$_3$ and BeO. Geochim. Cosmochim. Acta 50:1847–1860. With permission from Elsevier.)

13.5.2.4 Proton-Promoted Dissolution

Under acid conditions, protons can promote mineral dissolution by binding to surface oxide ions, causing bonds to weaken which is followed by detachment of metal species into solution. The proton-promoted dissolution rate (R_H) can be expressed as (Stumm and Wollast, 1990)

$$R_H = k_H'(\equiv MOH_2^+)^j = k_H'(C_H^s)^j, \qquad (13.23)$$

where

k_H' is the rate constant for proton-promoted dissolution
MOH_2^+ is the metal–proton complex
C_H^s is the concentration of the surface-adsorbed proton complex (mol^{-2})
j corresponds to the oxidation state of the central metal ion in the oxide structure (i.e., $j = 3$ for Al(III) and Fe(III) in simple cases)

If dissolution occurs by only one mechanism, j is an integer. Figure 13.26 shows an application of Equation 13.23 for the proton-promoted dissolution of γ-Al$_2$O$_3$.

13.5.2.5 Overall Dissolution Mechanisms

The rate of mineral dissolution, which is the sum of the ligand, proton, and deprotonation-promoted (or bonding of OH ligands) dissociation [$R_{OH} = k_{OH}'(C_{OH}^s)^i$] rates along with the pH-independent portion of the dissolution rate (k_{H_2O}'), which is due to hydration, can be expressed as (Stumm and Wollast, 1990)
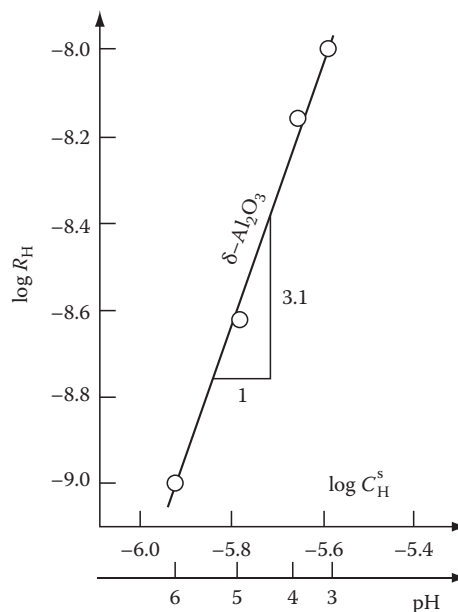
FIGURE 13.26 The dependence of the rate of proton-promoted dissolution of γ-Al$_2$O$_3$, R_H (mol m^{-2} h^{-1}) on the surface concentration of the proton complexes, C_H^s (mol m^{-2}). (Reprinted from Furrer, G., and W. Stumm. 1986. The coordination chemistry of weathering. I. Dissolution kinetics of γ-Al$_2$O$_3$ and BeO. Geochim. Cosmochim. Acta 50:1847–1860. With permission of Elsevier Science, Amsterdam, the Netherlands.)

$$R = k_L'(C_L^s) + k_H'(C_H^s)^j + k_{OH}'(C_{OH}^s)^i + k_{H_2O}'. \qquad (13.24)$$

Equation 13.24 is valid if dissolution occurs in parallel at varying metal centers (Furrer and Stumm, 1986).

13.5.2.6 Dissolution Kinetics of Metal Hydroxide Precipitates

The formation of metal hydroxide surface precipitates appears to be an important way to sequester metals. As the surface precipitates age, metal dissolution is greatly reduced (Figure 13.27). Thus, the metals are less prone to leaching and being taken up by plants and microbes. Peltier et al. (2010) reacted two soils of varying mineralogy and organic matter content with 3 mM Ni at two pHs, 6 and 7.5, and evaluated Ni bioavailability using a biosensor. At pH 6, where surface precipitates did not form, 60% of the Ni was bioavailable. However, at pH 7.5, where precipitates were observed to form from XAFS analyses, Ni bioavailability was markedly reduced to 25%. Similar results were found by Everhart et al. (2006).

The decrease in metal release and bioavailability is linked to the increasing silication of the interlayer of the LDH phases with increased residence time, resulting in a mineral transformation from a LDH phase to a precursor phyllosilicate surface precipitate (Ford et al., 1999; Ford and Sparks, 2000). The mechanism for this transformation is thought to be due to diffusion of Si, originating from weathering of the sorbent, into the interlayer space
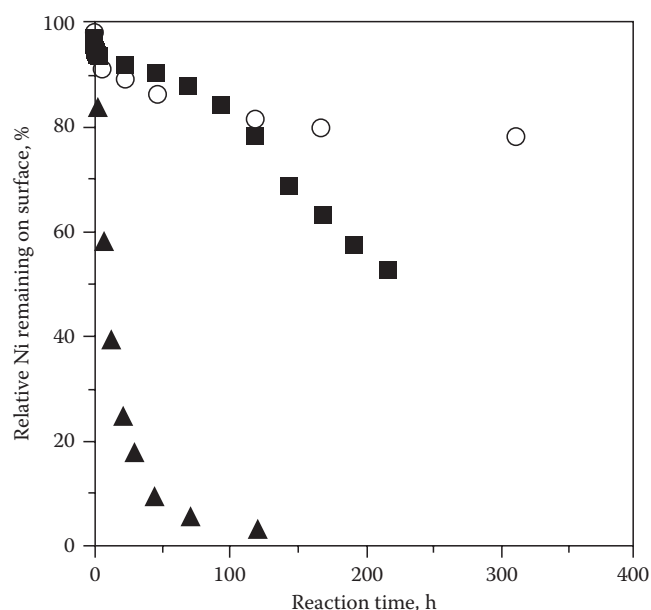
FIGURE 13.27 Kinetics of Ni detachment from surface precipitates at pH = 4. Relative Ni remaining on the surface (%) is shown for the *conventional method* (O) and the *replenishment method* (■) as a function of the reaction time. Ninety-eight percent of the initial Ni was sorbed in the beginning of the detachment experiment. The dissolution of an equivalent amount of crystalline Ni(OH)$_2$ (in mol) at pH = 4 is given for comparison (▲). (Reprinted from Scheidegger, A.M., and D.L. Sparks. 1996a. Kinetics of the formation and the dissolution of nickel surface precipitates on pyrophyllite. Chem. Geol. 132:157–164. With permission of Elsevier Science, Amsterdam, the Netherlands.)

of the LDH, replacing the anions such as NO$_3$. Polymerization and condensation of the interlayer Si slowly transform the LDH into a precursor metal–Al phyllosilicate. The metal stabilization that occurs in surface precipitates on Al-free sorbents (e.g., talc) may be due to Ostwald ripening, resulting in increased crystallization; however, microscopic analyses have shown that Ni–Al LDH phases appear to be amorphous (Livi et al., 2009). Peltier et al. (2006), using acid-solution calorimetry and results from previous calorimetry studies, showed that the enthalpy of formation of LDH phases is more exothermic, indicating great stability, on the order of Cl < NO$_3$ < SO$_4$ < CO$_3$ < Si of interlayer anionic composition, and that LDH phases were much more stable than a Ni(OH)$_2$ phase.

13.5.3 Redox Kinetics

It is well known that Mn(III/IV), Fe(III), Co(III), and Pb(IV) oxides/hydroxides are thermodynamically stable in oxygenated systems at neutral pH. However, under anoxic conditions, reductive dissolution of oxides/hydroxides by reducing agents occurs as shown below for MnOOH and MnO$_2$ (Stone, 1991):

$$Mn(III)OOH(s) + 3H^+ + e^- = Mn^{2+} + 2H_2O, \quad E^0 = +1.50\,V,$$

$$(13.25)$$

$$Mn(IV)O_2(s) + 4H^+ + 2e^- = Mn^{2+} + 2H_2O, \quad E^0 = +1.23\,V.$$

$$(13.26)$$
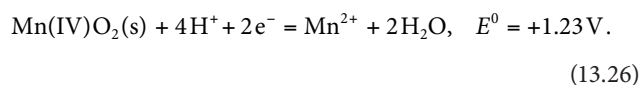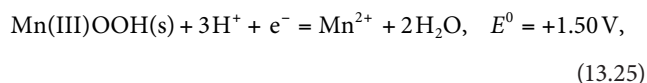
Changes in the oxidation state of the metals associated with the oxides above can greatly affect their solubility and mobility in soil and aqueous environments. The reductants can be either inorganic or organic.

There are a number of natural and xenobiotic organic functional groups that are good reducers of oxides and hydroxides. These include carboxyl, carbonyl, phenolic, and alcoholic functional groups of SOM. Microorganisms in soils and sediments are also examples of organic reductants. Stone (1987a) showed that oxalate and pyruvate, two microbial metabolites, could reduce and dissolve Mn(III/IV) oxide particles. Inorganic reductants include As(III), Cr(III), and Pu(III).

13.5.3.1 Mechanisms for Reductive Dissolution of Metal Oxides/Hydroxides

The reductive dissolution of metal oxides/hydroxides appears to occur in the following sequential steps (Stone, 1986, 1991): (1) diffusion of the reductant molecules to the oxide surface, (2) a surface CR, and (3) release of reaction products and diffusion away from the oxide surface. Steps (1) and (3) are transport steps. The rate-controlling step in reductive dissolution of oxides appears to be surface CR control. Reductive dissolution can be described by both inner- and outer-sphere complex mechanisms that involve (1) precursor complex formation, (2) electron transfer, and (3) breakdown of the successor complex (Figure 13.28). Inner-sphere and outer-sphere precursor complex formations are adsorption reactions that increase the density of reductant molecules at the oxide surface, which promotes electron transfer (Stone, 1991). In the inner-sphere mechanism, the reductant enters the inner coordination sphere by ligand exchange and bonds directly to the metal center prior to electron transfer. With the outer-sphere complex, the inner coordination sphere is left intact and electron transfer is enhanced by an outer-sphere precursor complex (Stone, 1986). Kinetic studies have shown that high rates of reductive dissolution are favored by high rates of precursor complex formation (i.e., large k_1 and low k_{-1} values), high electron transfer rates (i.e., large k_2), and high rates of product release (i.e., high k_3; Figure 13.28).

Specifically, adsorbed cations and anions may reduce reductive dissolution rates by blocking oxide surface sites or by causing release of Mn(II) into solution. Stone and Morgan (1984a) showed that PO$_4^{3-}$ inhibited the reductive dissolution of Mn(III/IV) oxides by hydroquinone. Addition of 10^{-2} M PO$_4^{3-}$ at pH 7.68 caused the dissolution rate to be only 25% of the rate when PO$_4^{3-}$ was not present. Phosphate had a greater effect than Ca^{2+}.

13.5.3.2 Oxidation of Pollutants

As mentioned earlier, Mn oxides can oxidize a number of environmentally important ions that can be toxic to humans and animals. Chromium and plutonium are similar in their chemical

Inner-sphere   Outer-sphere

(a) Precursor $M^{III}(H_2O)_6{}^{3+} + HA \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} M^{III}(A)(H_2O)_5{}^{2+} + H_3O^4$ $M^{III}(H_2O)_6{}^{3+} + HA \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} M^{III}(H_2O)_6{}^{3+}, HA$
    complex
    formation

(b) Electron $M^{III}(A)(H_2O)_5{}^{2+} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} M^{III}(\cdot A)(H_2O)_5{}^{2+}$ $M^{III}(H_2O)_6{}^{3+}, HA \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} M^{II}(H_2O)_6{}^{2+}, A^{\cdot} + H^+$
    transfer

(c) Breakdown $M^{II}(\cdot A)(H_2O)_5{}^{2+} + H_2O \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} M^{II}(H_2O)_6{}^{2+} + A^{\cdot}$ $M^{II}(H_2O)_6{}^{2+}, A^{\cdot} \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} M^{II}(H_2O)_6{}^{2+} + A^{\cdot}$
    of successor
    complex
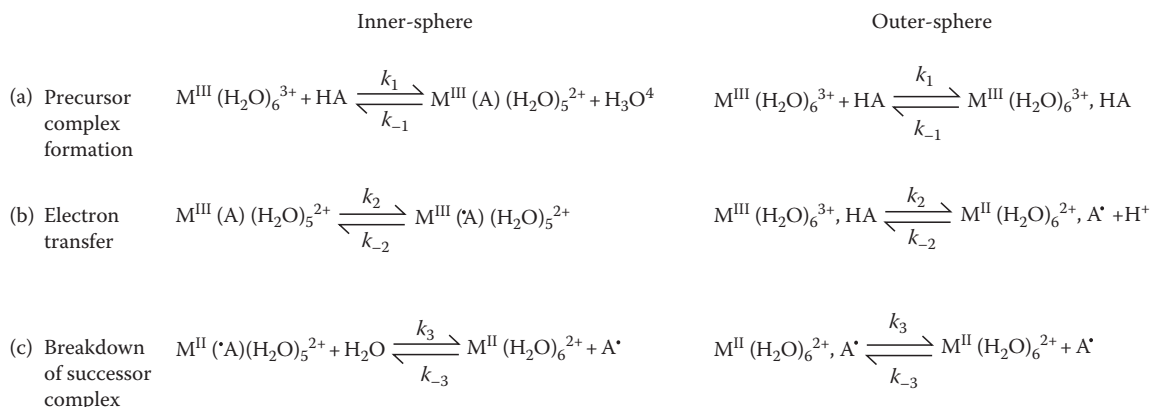
FIGURE 13.28 Reduction of $M(H_2O)_6{}^{3+}$ by phenol (HA) in homogeneous solution. (Reprinted with permission from Stone, A.T. 1986. Adsorption of organic reductants and subsequent electron transfer on metal oxide surfaces, p. 446–461. *In* J.A. Davis and K.F. Hayes (eds.) Geochemical processes at mineral surfaces. American Chemical Society, Washington, DC.)

behavior in aqueous settings (Rai and Serne, 1977; Bartlett and James, 1979). They can exist in multiple oxidation states and as both cationic and anionic species. Chromium(III) is quite stable and innocuous and occurs as Cr^{3+} and its hydrolysis products or as $CrO_2{}^-$. Chromium(III) can be oxidized to Cr(VI) by Mn(III/IV) oxides (Bartlett and James, 1979; Fendorf and Zasoski, 1992). Chromium(VI) is mobile in the soil environment and is a suspected carcinogen. It occurs as the dichromate ($Cr_2O_7{}^{2-}$) or chromate ($HCrO_4{}^-$ and $CrO_2{}^-$) anions (Huang, 1991).

Figure 13.29 shows the oxidation kinetics of Cr(III) to Cr(VI) in a soil. Most of the oxidation occurred during the
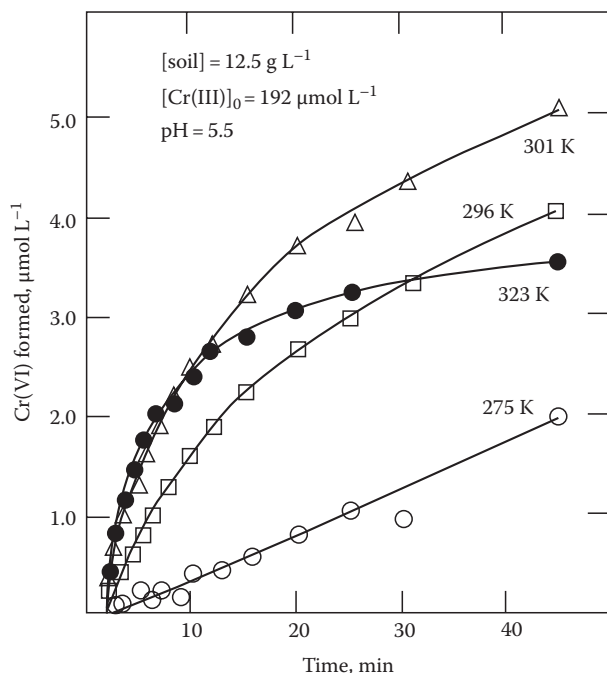
FIGURE 13.29 Effect of temperature on the kinetics of Cr(III) oxidation in moist Hagerstown silt loam soil. (Reprinted with permission from Amacher, M.C., and D.E. Baker. 1982. Redox reactions involving chromium, plutonium, and manganese in soils. DOE/DP/04515 Pennsylvania State University, University Park, PA.)

first hour. At higher temperatures, there was a rapid oxidation rate, followed by a slower rate. Fendorf and Zasoski (1992) found that Cr(III) oxidation on δ-MnO_2 was more rapid at pH 5 than pH 3 with overall production of Cr(VI) being greater at pH 3 at a Cr(III) concentration of $770\,\mu M$. The rate and extent of Cr(III) oxidation are affected by a number of factors including formation of surface precipitates at higher pHs and Cr(III) concentrations that effectively inhibit oxidation (Fendorf et al., 1992).

Plutonium can exist in the III to VI oxidation states as Pu^{3+}, Pu^{4+}, $PuO_2{}^+$, and $PuO_2{}^{2+}$ in strongly acid solutions (Huang, 1991). Plutonium(VI), which can result from oxidation of Pu(III/IV) by Mn(III/IV) oxides (Amacher and Baker, 1982), is very toxic and mobile in soils and waters.

Arsenic (As) can exist in several oxidation states and forms in soils and waters. In waters, As can exist in the +5, +3, 0, and -3 oxidation states. Arsenite, As(III), and arsine (AsH_3, where the oxidation state of As is -3) are much more toxic to humans than arsenate, As(V). Manganese(III/IV) oxides can oxidize As(III) to As(V) as shown later where As(III) as $HAsO_2$ is added to MnO_2 to produce As(V) as H_3AsO_4 (Oscarson et al., 1983):

$$HAsO_2 + MnO_2 = (MnO_2)HAsO_2, \tag{13.27}$$

$$(MnO_2)HAsO_2 + H_2O = H_3AsO_4 + MnO, \tag{13.28}$$

$$H_3AsO_4 = H_3AsO_4{}^- + H^+, \tag{13.29}$$

$$H_2AsO_4{}^- = HAsO_4{}^{2-} + H^+, \tag{13.30}$$

$$(MnO_2)HAsO_2 = 2H^+ = H_3AsO_4 + Mn^{2+}. \tag{13.31}$$
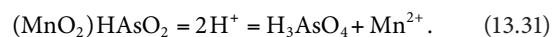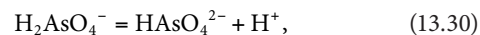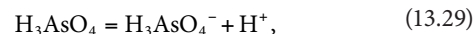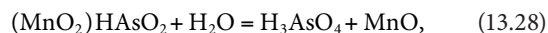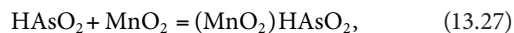
Equation 13.28 involves the formation of an adsorbed layer. Oxygen transfer occurs and $HAsO_2$ is oxidized to H_3AsO_4 (Equation 13.28). At pH ≤ 7, the predominant As(III) species is arsenious acid ($HAsO_2$), but the oxidation product, H_3AsO_4, will
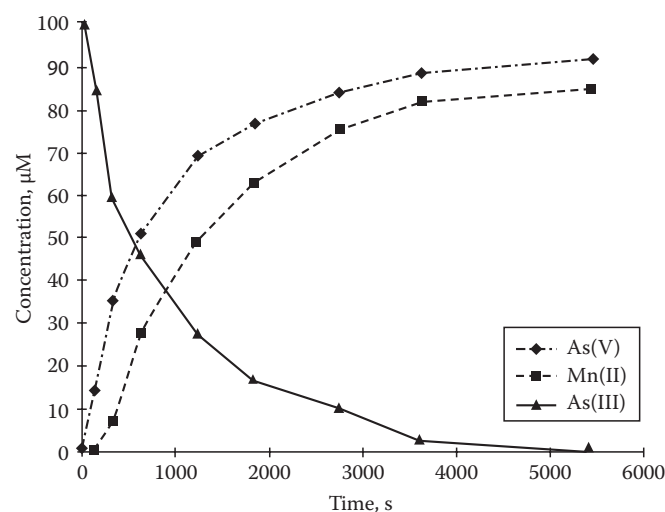
FIGURE 13.30 Experimental behavior of aqueous As(III), As(V), and Mn(II) following 99.6 μM As(III) addition to a 0.21 g L^{-1} of γ-MnO$_2$ particle suspension at pH 4, 25°C, and 0.1 M NaClO$_4$. (Reprinted from Scott, M.J., and J.J. Morgan. 1995. Reactions at oxide surfaces. 1. Oxidation of As(III) by synthetic birnessite. Environ. Sci. Technol. 29:1898–1905. With permission from American Chemical Society.)

dissociate and form the same quantities of HAsO$_4^-$ and HAsO$_4^{2-}$ with little H$_3$AsO$_4$ present at equilibrium (Equations 13.29 and 13.30). Each mole of As(III) oxidized releases about 1.5 mol H$^+$. The H$^+$ produced after H$_3$AsO$_4$ dissociation reacts with the adsorbed HAsO$_2$ on MnO$_2$, forming H$_3$AsO$_4$, and leads to the reduction and dissolution of Mn(IV) (Equation 13.31). Thus, every mole of As(III) that is oxidized to As(V) results in 1 mol of Mn(IV) in the solid phase being reduced to Mn(II) and partially dissolved in solution (Oscarson et al., 1981).

Oscarson et al. (1980) studied the oxidation of As(III) to As(V) in sediments from five lakes in Saskatchewan, Canada. Oxidation of As(III) to As(V) occurred within 48 h. In general, >90% of the added As was sorbed on the sediments within 72 h. Scott and Morgan (1995) studied the oxidation of As(III) by synthetic birnessite. The depletion of As(III) was rapid with 50% of the initial As(III) removed from solution in 10 min, and after 90 min, the As(III) concentration was below the detection limit. Arsenic(V) was released into solution as fast as As(III) was depleted and the total concentration of aqueous As was about constant over the duration of the experiment (Figure 13.30). Scott and Morgan (1995) concluded that the process of electron transfer and release of As(V) were fast compared to the sorption of As(III), the rate-limiting step.

Recently, Parikh et al. (2008), Ginder-Vogel et al. (2009), and Landrot et al. (2010) have used real-time molecular scale ATR–FTIR and QXAS techniques to elucidate the rapid redox kinetics and mechanisms of As(III) oxidation (Parikh et al., 2008; Ginder-Vogel et al., 2009) and Cr(III) oxidation (Landrot et al., 2010) on Mn oxides. Figure 13.31 shows that the oxidation of As(III) on hydrous Mn oxide at pHs of 6–9 is extremely rapid with 50% of the reaction occurring within 1 min. Spectra for times >1 min are very similar with only small differences
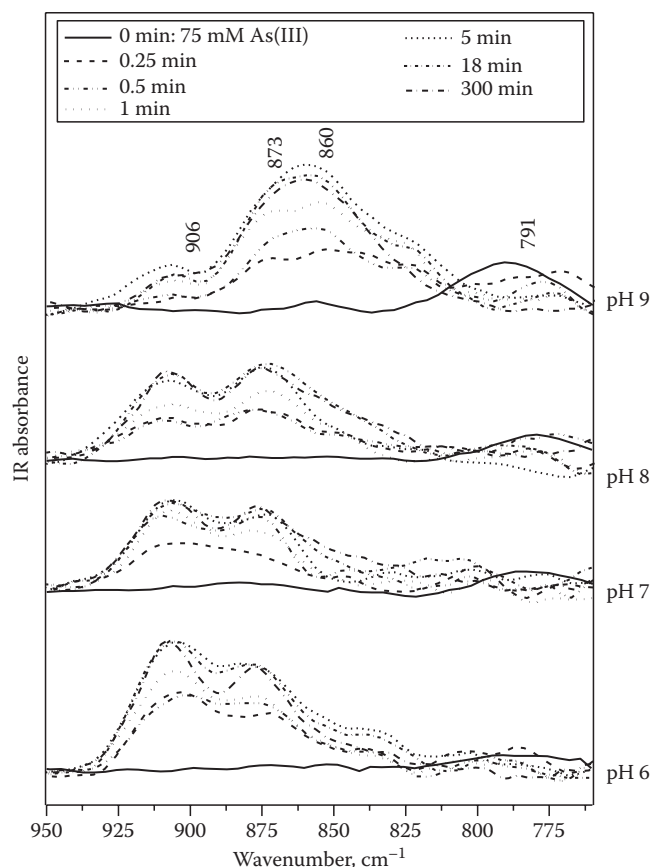
FIGURE 13.31 ATR–FTIR spectra for HMO reacted with 75 mmol kg^{-1} As(III) at pHs 6, 7, 8, and 9. Within 1 min of reaction, As(V) peaks (860, 873, and 906 cm^{-1}) are observed and only minimal changes in spectra are observed in a subsequent 300 min period. (Reprinted from Parikh, S.J., B.J. Lafferty, and D.L. Sparks. 2008. An ATR–FTIR spectroscopic approach for measuring rapid kinetics at the mineral/water interface. J. Colloid Interface Sci. 320:177–185. With permission from Elsevier.)

in peak intensities. The reaction process is almost complete in 5 min. Figure 13.32 shows QXAS and traditional batch data for As(III) oxidation on hydrous Mn oxide. An initial concentration of 1.5 mM of As(III) was used. After 1 s of reaction, the As(V) concentration, indicating oxidation of As(III), reached 0.37 mM and continued to increase rapidly for 45 s to reach a concentration of 1 mM. The As(V) concentration continued to increase for the remainder of the reaction, reaching the initial 1.5 mM concentration after 300 s of reaction time (Ginder-Vogel et al., 2009). The initial apparent As(III) depletion rate constants ($t < 30$ s) measured with QXAS were nearly twice as large as rate constants measured with traditional analytical techniques such as the batch method (Ginder-Vogel et al., 2009). Being able to measure the initial reaction rates with these techniques allows one to measure chemical rate constants that are independent of concentration, and, thus, important mechanistic information can be gleaned (Landrot et al., 2010).

Manganese oxides and hydroxides (e.g., Mn$_3$O$_4$ and MnOOH) may also catalyze the oxidation of other trace metals such as Co^{2+},
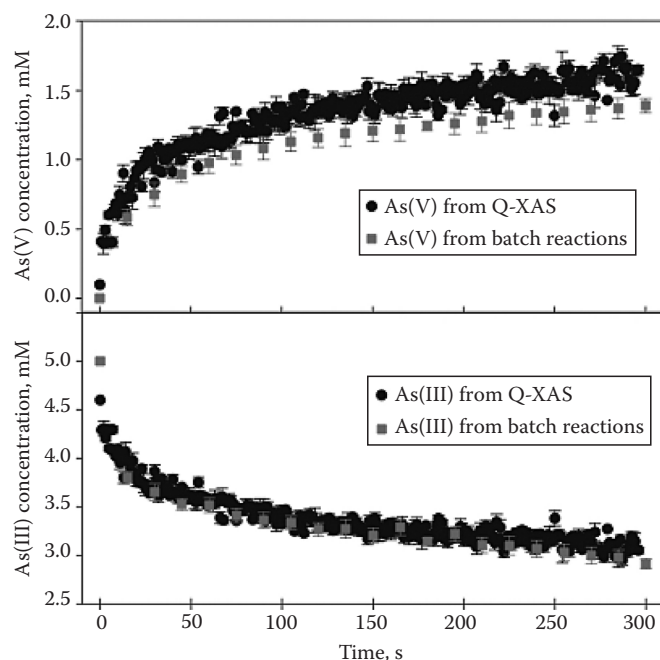
FIGURE 13.32 As(V) and As(III) concentrations determined from traditional batch and QXAS reactions. Error bars represent the SD of three measurements made at each time point. (Reprinted from Ginder-Vogel, M., G. Landrot, J.S. Fischel, and D.L. Sparks. 2009. Quantification of rapid environmental redox processes with quick-scanning X-ray absorption spectroscopy (Q-XAS). Proc. Natl Acad. Sci. U. S. A. 106:16124–16128. With permission National Academy of Sciences, USA.)

Co^{3+}, Cu^{2+}, Ni^{2+}, Ni^{3+}, and Pb^{2+} by disproportionation to Mn^{2+} and MnO_2 (Hem, 1978). The disproportionation results in vacancies in the Mn oxide structure. Since the Mn^{2+} and Mn^{3+} in the oxides have similar physical sizes as Co^{2+}, Co^{3+}, Cu^{2+}, Ni^{2+}, Ni^{3+}, and Pb^{2+}, these metals can occupy the vacancies in the Mn oxide and become part of the structure. With disproportionation or with other redox processes involving the Mn oxides, the solubility of the metals can be affected. For example, if during the disproportionation process Co_3O_4, the oxidized form of the metal forms from Co^{2+}, the reaction can be expressed as (Hem, 1978)

$$2Mn_3O_4(s) + 3Co^{2+} + 4H^+$$
$$\rightarrow MnO_2(s) + Co_3O_4(s) + 5Mn^{2+} + 2H_2O, \qquad (13.32)$$
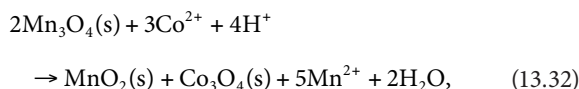
and the equilibrium constant (K^o) is (Hem, 1978)

$$\frac{(Mn^{2+})^5}{(Co^{2+})^3(H^+)^4} = 10^{18.73}. \qquad (13.33)$$

Thus, the oxidation of Co(II) to Co(III) reduces its solubility and mobility in the environment. Using XPS analyses (Murray and Dillard, 1979), this reaction has been shown to occur. More recent evidence for heterogeneous redox reactions of trace metals is discussed in Chapter 19. Scott and Morgan (1996) studied the

TABLE 13.5 Inorganic Redox Reactions with Manganese Dioxides

System	Time to Oxidize 50%	Driving Force at pH 4 Δe^o (V)[a]	Source
δ-MnO_2: As(III) $\rightarrow$ As(V) pH 4, 25°C, 14 m² L⁻¹	10 min	+0.529	Scott and Morgan (1995)
δ-MnO_2: Se(IV) $\rightarrow$ Se(VI) pH 4, 35°C, 14 m² L⁻¹	10 days		
pH 4, 25°C, 28 m² L⁻¹	16 days	+0.092	
pH 4, 25°C, 14 m² L⁻¹	30 days		
β-MnO_2: Cr(III) $\rightarrow$ Cr(VI) pH 4, 25°C, 71 m² L⁻¹	95 days	+0.011	Eary and Rai (1987)

Source: Reprinted from Scott, M.J., and J.J. Morgan. 1995. Reactions at oxide surfaces. 1. Oxidation of As(III) by synthetic birnessite. Environ. Sci. Technol. 29:1898–1905. With permission American Chemical Society.

[a] The activity ratio for each oxidant/reductant pair is taken as unity.

oxidation of Se(IV) by synthetic birnessite. Se(IV) was oxidized to Se(VI) with Se(VI) first appearing in the aqueous suspension after 12 h and was produced at a constant rate over the duration of the experiment (28 days). Scott and Morgan (1996) suggested the following oxidation mechanisms: (1) birnessite directly oxidized Se(IV) through a surface complex mechanism; (2) the rate-limiting step in the production of Se(VI) was the electron transfer step involving a transfer of two electrons from the anion to the metal ion, breaking of two Mn–O bonds, and addition of an O from water to Se(VI); and (3) the reaction products Se(VI) and Mn(II) were released from the surface by different steps.

Scott and Morgan (1996) compared their results with those of Eary and Rai (1987) who studied Cr(III) oxidation by pyrolusite (β-MnO_2) between pH 3.0 and 4.7 and Scott and Morgan (1995) who studied As(III) oxidation by birnessite (δ-MnO_2) at pH values between 4.0 and 8.2 (Table 13.5). The Cr(III) redox transformation on pyrolusite was slowest which Scott and Morgan (1996) attributed to unfavorable adsorption on both a positively charged surface and aqueous species and the small thermodynamic driving force. Also, the transfer of three electrons from Cr(III) to Mn(IV) requires the involvement of more than one Mn(IV) atom per Cr(III) atom.

Manganese oxides also appear to play an important role in ligand-facilitated metal transport. Using soil columns that consisted of fractured saprolite coated with amorphous Fe and Mn oxides, Jardine et al. (1993) studied the transport of Co(II) $EDTA^{2-}$, a mixture of Co(II) $EDTA^{2-}$ and Co(III) $EDTA^-$ and Sr $EDTA^{2-}$. The Mn oxides oxidized Co(II) $EDTA^{2-}$ into Co(III) $EDTA^-$, a very stable complex (log K value of 41.4; Xue and Traina, 1996). The formation of this complex resulted in enhanced transport of Co.

Xue and Traina (1996) found that an aerobic goethite suspension catalyzed oxidation of Co(II) $EDTA^{2-}$ to Co(III) $EDTA^-$ by dissolved O_2. The kinetics were described using a pseudo-first-order rate constant, k_1 of 0.0078 ± 0.0002 h⁻¹ at pH 5 and a goethite concentration of 3.09 g L⁻¹.

A number of investigators have studied the reductive dissolution of Mn oxides by organic pollutants such as hydroquinone (Stone and Morgan, 1984a), substituted phenols (Stone, 1987b), and other organic compounds (Stone and Morgan, 1984b). With substituted phenols, the rate of dissolution was proportional to substituted phenol concentration and the rate increased as pH decreased (Stone, 1987b). Phenols containing alkyl, alkoxy, or other electron-donating substituents were more slowly degraded; *p*-nitrophenol reacted slowly with Mn(III/IV) oxides. The increased rate of reductive dissolution at lower pH may be due to more protonation reactions that enhance the creation of precursor complexes or increases in the protonation level of the surface precursor complexes that increase electron transfer rates (Stone, 1987b). Further discussions on this topic can be found in Chapter 19.

13.6 Conclusions

Research on the kinetics and mechanisms of soil CRs will be a common theme in soil and environmental sciences for decades to come. This research emphasis is in large part due to the need to more accurately understand and predict the long-term fate and transport of contaminants in the subsurface environment. Without such data, economically sound decisions about soil remediation cannot be made and risk assessment models are incomplete and most probably inaccurate.

While some very fine and informative research on rates and mechanisms of soil CRs/processes has been conducted in the past few decades, there are many gaps that need to be filled. The following research is needed: (1) more accurate kinetic models that describe reactions on multireactive, heterogeneous particle surfaces; (2) long-term sorption and particularly, desorption rate studies; (3) a better understanding of residence time effects on plant nutrient, radionuclide, metal, and organic retention/release mechanisms on soils and other natural materials; (4) an increased knowledge of nucleation/precipitation and dissolution reaction rate phenomena at the mineral/water interface and their effect on nutrient/contaminant mobility and bioavailability in the soil environment; (5) more studies on the kinetics and mechanisms of redox processes in soils, particularly the role that soil components such as Mn oxides have on oxidation/reduction of inorganic and organic pollutants; and (6) increased use of real-time molecular scale in situ spectroscopic and microscopic techniques to confirm reaction mechanisms.

References

Aharoni, C. 1984. Kinetics of adsorption: The S-shaped Z(t) plot. Adsorp. Sci. Technol. 1:1–29.

Aharoni, C., and D.L. Sparks. 1991. Kinetics of soil chemical reactions: A theoretical treatment, p. 1–18. *In* D.L. Sparks and D.L. Suarez (eds.) Rates of soil chemical processes. SSSA, Madison, WI.

Aharoni, C., and Y. Suzin. 1982a. Application of the Elovich equation to the kinetics of occlusion: Part 1. Homogenous microporosity. J. Chem. Soc., Faraday Trans. 1 78:2313–2320.

Aharoni, C., and Y. Suzin. 1982b. Application of the Elovich equation to the kinetics of occlusion: Part 3. Heterogenous microporosity. J. Chem. Soc., Faraday Trans. 1 78:2329–2336.

Aharoni, C., and M. Ungarish. 1976. Kinetics of activated chemisorption. Part 1. The non-Elovichian part of the isotherm. J. Chem. Soc., Faraday Trans. 1 72:400–408.

Ainsworth, C.C., J.L. Pilon, P.L. Gassman, and W.G. Van Der Sluys. 1994. Cobalt, cadmium, and lead sorption to hydrous iron oxide: Residence time effect. Soil Sci. Soc. Am. J. 58:1615–1623.

Alexander, M. 1995. How toxic are toxic-chemicals in soil. Environ. Sci. Technol. 29:2713–2717.

Allen, E.R., D.W. Ming, L.R. Hossner, and D.L. Henninger. 1995. Modeling transport kinetics in clinoptilolite–phosphate rock systems. Soil Sci. Soc. Am. J. 59:248–255.

Amacher, M.C. 1991. Methods of obtaining and analyzing kinetic data, p. 19–59. *In* D.L. Sparks and D.L. Suarez (eds.) Rates of soil chemical processes. SSSA, Madison, WI.

Amacher, M.C., and D.E. Baker. 1982. Redox reactions involving chromium, plutonium, and manganese in soils. DOE/DP/04515 Pennsylvania State University, University Park, PA.

Atkinson, R.J., F.J. Hingston, A.M. Posner, and J.P. Quirk. 1970. Elovich equation for the kinetics of isotope exchange reactions at solid–liquid interfaces. Nature 226:148–149.

Backes, E.A., R.G. Mclaren, A.W. Rate, and R.S. Swift. 1995. Kinetics of cadmium and cobalt desorption from iron and manganese oxides. Soil Sci. Soc. Am. J. 59:778–785.

Ball, W.P., and P.V. Roberts. 1991. Long-term sorption of halogenated organic chemicals by aquifer material: 1. Equilibrium. Environ. Sci. Technol. 25:1223–1237.

Bartlett, R., and B. James. 1979. Behavior of chromium in soils. III. Oxidation. J. Environ. Qual. 8:31–35.

Bernasconi, C.F. 1976. Relaxation kinetics. Academic Press, New York.

Bondietti, G., J. Sinniger, and W. Stumm. 1993. The reactivity of Fe(III) (hydr)oxides: Effects of ligands in inhibiting the dissolution. Colloids Surf. A 79:157–167.

Borda, M.J., and D.L. Sparks. 2008. Kinetics and mechanisms of sorption–desorption in soils: A multiscale assessment, p. 97–124. *In* A. Violante, P.M. Huang, and G.M. Gadd (eds.) Biophysico-chemical processes of heavy metals and metalloids in soil environments. John Wiley & Sons, Inc., New York.

Bosbach, D., and W. Rammensee. 1994. In situ investigation of growth and dissolution on the (010) surface of gypsum by scanning force microscopy. Geochim. Cosmochim. Acta 58:843–849.

Bruemmer, G.W., J. Gerth, and K.G. Tiller. 1988. Reaction kinetics of the adsorption and desorption of nickel, zinc and cadmium by goethite: I. Adsorption and diffusion of metals. J. Soil Sci. 39:37–52.

Brusseau, M.L., and P.S.C. Rao. 1989. Sorption non-ideality during organic contaminant transport in porous media. CRC Crit. Rev. Environ. Control 19:33–99.

Bunnett, J.F. 1986. Kinetics in solution, p. 171–250. *In* C.F. Bernasconi (ed.) Investigations of rates and mechanisms of reactions. Wiley, New York.

Bunzl, K., W. Schmidt, and B. Sansoni. 1976. Kinetics of ion exchange in soil organic matter. IV. Adsorption and desorption of Pb^{2+}, Cu^{2+}, Zn^{2+}, and Ca^{2+} by peat. J. Soil. Sci. 27:32–41.

Carroll, K.M., M.R. Harkness, A.A. Bracco, R.R. Balcarcel. 1994. Application of a permeant/polymer diffusional model to the desorption of polychlorinated biphenyls from Hudson River sediments. Environ. Sci. Technol. 28:253–258.

Casey, W.H., H.R. Westrich, G.W. Arnold, and J.F. Banfield. 1989. The surface chemistry of dissolving labradorite feldspar. Geochim. Cosmochim. Acta 53:821–832.

Chang, M.L., S.C. Wu, and C.Y. Chen. 1997. Diffusion of volatile organic compounds in pressed humic acid disks. Environ. Sci. Technol. 31:2307–2312.

Charlet, L., and A. Manceau. 1993. Structure, formation, and reactivity of hydrous oxide particles: Insights from X-ray absorption spectroscopy, p. 117–164. *In* J. Buffle and H.P. van Leeuwen (eds.) Environmental particles. Lewis Publishers, Boca Raton, FL.

Chien, S.H., and W.R. Clayton. 1980. Application of Elovich equation to the kinetics of phosphate release and sorption in soils. Soil Sci. Soc. Am. J. 44:265–268.

Chute, J.H., and J.P. Quirk. 1967. Diffusion of potassium from mica-like materials. Nature 213:1156–1157.

Comans, R.N.J., and D.E. Hockley. 1992. Kinetics of cesium sorption on illite. Geochim. Cosmochim. Acta 56:1157–1164.

Connaughton, D.F., J.R. Stedinger, L.W. Lion, and M.L. Shuler. 1993. Description of time-varying desorption kinetics: Release of naphthalene from contaminated soils. Environ. Sci. Technol. 27:2397–2403.

Cornelissen, G., P.C.M. van Noort, J.R. Parsons, and H.A.J. Govers. 1997. Temperature dependence of slow adsorption and desorption kinetics of organic compounds in sediments. Environ. Sci. Technol. 31:454–460.

Crank, J. 1976. The mathematics of diffusion. 2nd edn. Oxford University Press (Clarendon), London, U.K.

DiToro, D.M., and L.M. Horzempa. 1982. Reversible and resistant components of PCB adsorption–desorption: Isotherms. Environ. Sci. Technol. 16:594–602.

Divincenzo, J.P., and D.L. Sparks. 1997. Slow sorption kinetics of pentachlorophenol on soil: Concentration effects. Environ. Sci. Technol. 31:977–983.

Eary, L.E., and D. Rai. 1987. Kinetics of chromium(III) oxidation to chromium(VI) by reaction with manganese dioxide. Environ. Sci. Technol. 21:1187–1193.

Elzinga, E.J., and D.L. Sparks. 1999. Nickel sorption mechanisms in a pyrophyllite–montmorillonite mixture. J. Colloid Interface Sci. 213:506–512.

Everhart, J.L., D. McNear Jr., E. Peltier, D. van der Lelie, R.L. Chaney, and D.L. Sparks. 2006. Assessing nickel bioavailability in smelter-contaminated soils. Sci. Total Environ. 367:732–744.

Fendorf, S., M.J. Eick, P. Grossl, and D.L. Sparks. 1997. Arsenate and chromate retention mechanisms on goethite. 1. Surface structure. Environ. Sci. Technol. 31:315–320.

Fendorf, S.E., M. Fendorf, D.L. Sparks, and R. Gronsky. 1992. Inhibitory mechanisms of Cr(III) oxidation by δ-MnO_2. J. Colloid Interface Sci. 153:37–54.

Fendorf, S.E., G.M. Lamble, M.G. Stapleton, M.J. Kelley, and D.L. Sparks. 1994. Mechanisms of chromium(III) sorption on silica: 1. Cr(III) surface structure derived by extended X-ray absorption fine structure spectroscopy. Environ. Sci. Technol. 28:284–289.

Fendorf, S.E., and R.J. Zasoski. 1992. Chromium(III) oxidation by δ-MnO_2.1. Characterization Environ. Sci. Technol. 26:79–85.

Fendorf, S.E. et al. 1993. Soil Sci. Soc. Am. J., 57:57–62.

Ford, R.G., A.C. Scheinost, K.G. Scheckel, and D.L. Sparks. 1999. The link between clay mineral weathering and the stabilization of Ni surface precipitates. Environ. Sci. Technol. 33:3140–3144.

Ford, R.G., and D.L. Sparks. 2000. The nature of Zn precipitates formed in the presence of pyrophyllite. Environ. Sci. Technol. 34:2479–2483.

Fuller, C.C., J.A. Davis, and G.A. Waychunas. 1993. Surface chemistry of ferrihydride: Part 2. Kinetics of arsenate adsorption and coprecipitation. Geochim. Cosmochim. Acta 57:2271–2282.

Furrer, G., and W. Stumm. 1986. The coordination chemistry of weathering. I. Dissolution kinetics of γ-Al_2O_3 and BeO. Geochim. Cosmochim. Acta 50:1847–1860.

Gardiner, W.C., Jr. 1969. Rates and mechanisms of chemical reactions. Benjamin, New York.

Ginder-Vogel, M., G. Landrot, J.S. Fischel, and D.L. Sparks. 2009. Quantification of rapid environmental redox processes with quick-scanning X-ray absorption spectroscopy (Q-XAS). Proc. Natl Acad. Sci. U. S. A. 106:16124–16128.

Grossl, P.R., M. Eick, D.L. Sparks, S. Goldberg, and C.C. Ainsworth. 1997. Arsenate and chromate retention mechanisms on goethite. 2. Kinetic evaluation using a pressure-jump relaxation technique. Environ. Sci. Technol. 31:321–326.

Grossl, P.R., and D.L. Sparks. 1995. Evaluation of contaminant ion adsorption/desorption on goethite using pressure-jump relaxation kinetics. Geoderma 67:87–101.

Grossl, P.R., D.L. Sparks, and C.C. Ainsworth. 1994. Rapid kinetics of Cu(II) adsorption/desorption on goethite. Environ. Sci. Technol. 28:1422–1429.

Hachiya, K., M. Sasaki, I. Ikeda, N. Mikami, and T. Yasunaga. 1984. Static and kinetic studies of adsorption–desorption of metal ions on a γ-Al_2O_3 surface. 2. Kinetic studies by means of pressure-jump technique. J. Phys. Chem. 88:27–31.

Hamaker, J.W., and J.M. Thompson. 1972. Adsorption, p. 39–151. *In* C.A.I. Goring and J.W. Hamaker (eds.) Organic chemicals in the environment. Marcel Dekker, New York.

Harmon, T.C., L. Semprini, and P.V. Roberts. 1992. Simulating solute transport using laboratory-based sorption parameters. J. Environ. Eng. 118:666–689.

Havlin, J.L., and D.G. Westfall. 1985. Potassium release kinetics and plant response in calcareous soils. Soil Sci. Soc. Am. J. 49:366–370.

Hayes, K.F., and J.O. Leckie. 1986. Mechanism of lead ion adsorption at the goethite–water interface. ACS Symp. Ser. 323:114–141.

Hellman, R., B. Drake, and K. Kjoller. 1992. Using atomic force microscopy to study the structure, topography and dissolution of albite surfaces, p. 149–152. *In* Y.K. Kharaka and A.S. Maest (eds.) Water–rock interaction VII. A.A. Balkema, Rotterdam, the Netherlands.

Hem, J.D. 1978. Redox processes at the surface of manganese oxide and their effects on aqueous metal ions. Chem. Geol. 21:199–218.

Hillner, P.E., A.J. Gratz, S. Manne, and P.K. Hansma. 1992a. Atomic-scale imaging of calcite growth and dissolution in real-time. Geology 20:359–362.

Hillner, P.E., S. Manne, A.J. Gratz, and P.K. Hansma. 1992b. AFM images of dissolution and growth on a calcite crystal. Ultramicroscopy 44:1387–1393.

Huang, P.M. 1991. Kinetics of redox reactions on manganese oxides and its impact on environmental quality, p. 191–230. *In* D.L. Sparks and D.L. Suarez (eds.) Rates of soil chemical processes. SSSA, Madison, WI.

Jardine, P.M., F.M. Dunnivant, H.M. Selim, and J.F. McCarthy. 1992. Comparison of models for describing the transport of dissolved organic carbon in aquifer columns. Soil Sci. Soc. Am. J. 56:393–401.

Jardine, P.M., G.K. Jacobs, and J.D. Odell. 1993. Unsaturated transport processes in undisturbed heterogeneous porous-media. 2. Cocontaminants. Soil Sci. Soc. Am. J. 57:954–962.

Jardine, P.M., and D.L. Sparks. 1984. Potassium-calcium exchange in a multireactive soil system. I. Kinetics. Soil Sci. Soc. Am. J. 48:39–45.

Johnsson, P.A., M.F. Hochella, Jr., G.A. Parks, A.E. Blum, and G. Sposito. 1992. Direct observation of muscovite basal-plane dissolution and secondary phase formation: An XPS, LEED, and SFM study, p. 159–162. *In* Y.K. Kharaka and A.S. Maest (eds.) Water–rock interaction VII. A.A. Balkema, Rotterdam, the Netherlands.

Jordan, G., and W. Rammensee. 1996. Dissolution rates and activation energy for dissolution of brucite(001): A new method based on the microtopography of crystal surfaces. Geochim. Cosmochim. Acta 60:5055–5062.

Junta, J.L., and M.F. Hochella, Jr. 1994. Manganese(II) oxidation at mineral surfaces: A microscopic and spectroscopic study. Geochim. Cosmochim. Acta 58:4985–4999.

Kan, A.T., G.M. Fu, M.A. Hunter, and M.B. Tomson. 1997. Irreversible adsorption of naphthalene and tetrachlorobiphenyl to Lula and surrogate sediments. Environ. Sci. Technol. 31:2176–2185.

Karickhoff, S.W. 1980. Sorption kinetics of hydrophobic pollutants in natural sediments, p. 193–205. *In* R.A. Baker (ed.) Contaminants and sediments. Ann Arbor Science, Ann Arbor, MI.

Karickhoff, S.W., D.S. Brown, and T.A. Scott. 1979. Sorption of hydrophobic pollutants on natural sediments. Water Res. 13:241–248.

Karickhoff, S.W., and K.R. Morris. 1985. Sorption dynamics of hydrophobic pollutants in sediment suspensions. Environ. Toxicol. Chem. 4:469–479.

Keren, R., P.R. Grossl, and D.L. Sparks. 1994. Equilibrium and kinetics of borate adsorption–desorption on pyrophyllite in aqueous suspensions. Soil Sci. Soc. Am. J. 58: 1116–1122.

Krishnamurti, G.S.R., G. Cieslinski, P.M. Huang, and K.C.J. VanRees. 1997. Kinetics of cadmium release from soils as influenced by organic acids: Implication in cadmium availability. J. Environ. Qual. 26:271–277.

Krishnamurti, G.S.R., and P.M. Huang. 1992. Dynamics of potassium chloride induced manganese release in different soil orders. Soil Sci. Soc. Am. J. 56:11115–11123.

Kuo, S., and E.G. Lotse. 1974. Kinetics of phosphate adsorption and desorption by lake sediments. Soil Sci. Soc. Am. Proc. 38:50–54.

Kuo, S., and D.S. Mikkelsen. 1980. Kinetics of zinc desorption from soils. Plant Soil 56:355–364.

Landrot, G., M. Ginder-Vogel, and D.L. Sparks. 2010. Kinetics of chromium(III) oxidation by manganese(IV) oxides using quick scanning X-ray absorption fine structure spectroscopy (Q-XAFS). Environ. Sci. Technol. 44:143–149.

Lasaga, A.C. 1984. Chemical-kinetics of water–rock interactions. J. Geophys. Res. 89:4009–4025.

Lee, L.S., P.S.C. Rao, M.L. Brusseau, and R.A. Ogwada. 1988. Nonequilibrium sorption of organic contaminants during flow through columns of aquifer materials. Environ. Toxicol. Chem. 7:779–793.

Leenheer, J.A., and J.L. Ahlrichs. 1971. A kinetic and equilibrium study of the adsorption of carbaryl and parathion upon soil organic matter surfaces. Soil Sci. Soc. Am. Proc. 35:700–704.

Lehmann, R.G., and R.D. Harter. 1984. Assessment of copper–soil bond strength by desorption kinetics. Soil Sci. Soc. Am. J. 48:769–772.

Livi, K.J.T., G.S. Senesi, A.C. Scheinost, and D.L. Sparks. 2009. Microscopic examination of nanosized mixed Ni–Al hydroxide surface precipitates on pyrophyllite. Environ. Sci. Technol. 43:1299–1304.

Loehr, R.C., and M.T. Webster. 1996. Behavior of fresh vs aged chemicals in soil. J. Soil Contam. 5:361–383.

Lövgren, L., S. Sjöberg, and P.W. Schindler. 1990. Acid/base reactions and Al(III) complexation at the surface of goethite. Geochim. Cosmochim. Acta 54:1301–1306.

Low, M.J.D. 1960. Kinetics of chemisorption of gases on solids. Chem. Rev. 60:267–312.

Matocha, C.J., K.G. Scheckel, and D.L. Sparks. 2005. Kinetics and mechanisms of soil biogeochemical processes, p. 309–342. *In* M.A. Tabatbai and D.L. Sparks (eds.) Chemical processes in soils. SSSA, Madison, WI.

Maurice, P.A. 1998. Scanning probe microscopy of environmental surfaces, p. 109–154. *In* P.M. Huang, N. Senesi, and J. Buffle (eds.) Structure and surface reactions of soil particles. Vol. 4. John Wiley & Sons, New York.

Maurice, P.A., M.F. Hochella, Jr., G.A. Parks, G. Sposito, and U. Schwertmann. 1995. Evolution of hematite surface microtopography upon dissolution by simple organic acids. Clays Clay Miner. 43:29–38.

McCall, P.J., and G.L. Agin. 1985. Desorption kinetics of picloram as affected by residence time in the soil. Environ. Toxicol. Chem. 4:37–44.

McNear, D.H., R.L. Chaney, and D.L. Sparks. 2007. The effects of soil type and chemical treatment on nickel speciation in refinery enriched soils: A multi-technique investigation. Geochim. Cosmochim. Acta 71:2190.

Mikami, N., M. Sasaki, K. Hachiya, R.D. Ikeda, and T. Yasunaga. 1983a. Kinetics of the adsorption of PO_4 on the γ-Al_2O_3 surface using the pressure-jump technique. J. Phys. Chem. 87:1454–1458.

Mikami, N., M. Sasaki, K. Hachiya, and T. Yasunaga. 1983b. Kinetic study of the adsorption–desorption of the uranyl ion on a γ-Al_2O_3 surface using the pressure-jump technique. J. Phys. Chem. 87:5478–5481.

Mikami, N., M. Sasaki, T. Kikuchi, and T. Yasunaga. 1983c. Kinetics of the adsorption–desorption of chromate on γ-Al_2O_3 surfaces using the pressure-jump technique. J. Phys. Chem. 87:5245–5248.

Miller, C.T., and J. Pedit. 1992. Use of a reactive surface-diffusion model to describe apparent sorption–desorption hysteresis and abiotic degradation of lindane in a subsurface material. Environ. Sci. Technol. 26:1417–1427.

Murray, J.W., and J.G. Dillard. 1979. The oxidation of cobalt(II) adsorbed on manganese dioxide. Geochim. Cosmochim. Acta 43:781–787.

Nachtegaal, M., M.A. Marcus, J.E. Sonke, J. Vangronsveld, K.J.T. Livi, D. Van der Lelie, and D.L. Sparks. 2005. Effects of in situ remediation on the speciation and bioavailability of zinc in a smelter contaminated soil. Geochim. Cosmochim. Acta 69:4649–4664.

Nkedi-Kizza, P., J.W. Biggar, H.M. Selim, M.Th. van Genuchten, P.J. Wierenga, J.M. Davidson, and D.R. Nielsen. 1984. On the equivalence of two conceptual models for describing ion exchange during transport through an aggregated oxisol. Water Resour. Res. 20:1123–1130.

O'Day, P.A., G.E. Brown, and G.A. Parks. 1994a. X-Ray absorption spectroscopy of cobalt(II) multinuclear surface complexes and surface precipitates on kaolinite. J. Colloid Interface Sci. 165:269–289.

O'Day, P.A., G.A. Parks, and G.E. Brown, Jr. 1994b. Molecular structure and binding sites of cobalt(II) surface complexes on kaolinite from X-ray absorption spectroscopy. Clays Clay Miner. 42:337–355.

Onken, A.B., and R.L. Matheson. 1982. Dissolution rate of EDTA-extractable phosphate from soils. Soil Sci. Soc. Am. J. 46:276–279.

Oscarson, D.W., P.M. Huang, C. Defosse, and A. Herbillon. 1981. The oxidative power of Mn(IV) and Fe(III) oxides with respect to As(III) in terrestrial and aquatic environment. Nature 291:50–51.

Oscarson, D.W., P.M. Huang, and U.T. Hammer. 1983. Oxidation and sorption of arsenite by manganese dioxide as influenced by surface coatings of iron and aluminum oxides and calcium carbonate. Water Air Soil Pollut. 20:233–244.

Oscarson, D.W., P.M. Huang, and W.K. Liaw. 1980. The oxidation of arsenite by aquatic sediments. J. Environ. Qual. 9:700–703.

Padmanabham, M. 1983. Adsorption–desorption behavior of copper(II) at the goethite–solution interface. Aust. J. Soil Res. 21:309–320.

Parikh, S.J., and J. Chorover. 2007. Infrared spectroscopy studies of cation effects on lipopolysaccharides in aqueous solution. Colloids Surf. B 55:241–250.

Parikh, S.J., B.J. Lafferty, and D.L. Sparks. 2008. An ATR–FTIR spectroscopic approach for measuring rapid kinetics at the mineral/water interface. J. Colloid Interface Sci. 320:177–185.

Pavlostathis, S.G., and G.N. Mathavan. 1992. Desorption kinetics of selected volatile organic compounds from field contaminated soils. Environ. Sci. Technol. 26:532–538.

Pedit, J.A., and C.T. Miller. 1995. Heterogenous sorption processes in subsurface systems. 2. Diffusion modeling approaches. Environ. Sci. Technol. 29:1766–1772.

Peltier, E., R. Allada, A. Navrotsky, and D.L. Sparks. 2006. Nickel solubility and precipitation in soils: A thermodynamic study. Clays Clay Miner. 54:153–164.

Peltier, E., D. van der Lelie, and D.L. Sparks. 2010. Formation and stability of Ni–Al hydroxide phases in soils. Environ. Sci. Technol. 44:302–308.

Petrovic, R., R.A. Berner, and M.B. Goldhaber. 1976. Rate control in dissolution of alkali feldspars. I. Study of residual feldspar grains by X-ray photoelectron spectroscopy. Geochim. Cosmochim. Acta 40:537–548.

Pignatello, J.J. 1998. Soil organic matter as a nanoporous sorbent of organic pollutants. Adv. Colloid Interface Sci. 76:445–467.

Pignatello, J.J. 2000. The measurement and interpretation of sorption and desorption rates for organic compounds in soil media. Adv. Agron. 69:1–73.

Pignatello, J.J., F.J. Ferrandino, and L.Q. Huang. 1993. Elution of aged and freshly added herbicides from a soil. Environ. Sci. Technol. 27:1563–1571.

Pignatello, J.J., and L.Q. Huang. 1991. Sorptive reversibility of atrazine and metolachlor residues in field soil samples. J. Environ. Qual. 20:222–228.

Polyzopoulos, N.A., V.Z. Keramidas, A. Pavlatou. 1986. On the limitations of the simplified Elovich equation in describing the kinetics of phosphate sorption and release from soils. J. Soil Sci. 37:81–87.

Rai, E., and R.J. Serne. 1977. Plutonium activities in soil solutions and the stability and formation of selected plutonium minerals. J. Environ. Qual. 6:89–95.

Roberts, D.R., A.M. Scheidegger, and D.L. Sparks. 1999. Kinetics of mixed Ni–Al precipitate formation on a soil clay fraction. Environ. Sci. Technol. 33:3749–3754.

Scheidegger, A.M., G.M. Lamble, and D.L. Sparks. 1996. Investigation of Ni sorption on pyrophyllite: An XAFS study. Environ. Sci. Technol. 30:548–554.

Scheidegger, A.M., G.M. Lamble, and D.L. Sparks. 1997. Spectroscopic evidence for the formation of mixed-cation hydroxide phases upon metal sorption on clays and aluminum oxides. J. Colloid Interface Sci. 186:118–128.

Scheidegger, A.M., and D.L. Sparks. 1996a. Kinetics of the formation and the dissolution of nickel surface precipitates on pyrophyllite. Chem. Geol. 132:157–164.

Scheidegger, A.M., and D.L. Sparks. 1996b. A critical assessment of sorption–desorption mechanisms at the soil mineral/ water interface. Soil Sci. 161:813–831.

Scheidegger, A.M., D.G. Strawn, G.M. Lamble, and D.L. Sparks. 1998. The kinetics of mixed Ni–Al hydroxide formation on clay and aluminum oxide minerals: A time-resolved XAFS study. Geochim. Cosmochim. Acta 62:2233–2245.

Schott, J., and J.C. Petit. 1987. New evidence for the mechanisms of dissolution of silicate minerals, p. 293–312. *In* W. Stumm (ed.) Aquatic surface chemistry. Wiley Interscience, New York.

Schultz, M.F., M.M. Benjamin, and J.F. Ferguson. 1987. Adsorption and desorption of metals on ferrihydrite—Reversibility of the reaction and sorption properties of the regenerated solid. Environ. Sci. Technol. 21:863–869.

Schwarzenbach, R.T., P.M. Gschwend, and D.M. Boden. 1993. Environmental organic chemistry. Wiley, New York.

Scott, M.J., and J.J. Morgan. 1995. Reactions at oxide surfaces. 1. Oxidation of As(III) by synthetic birnessite. Environ. Sci. Technol. 29:1898–1905.

Scott, M.J., and J.J. Morgan. 1996. Reactions at oxide surfaces. 2. Oxidation of Se(IV) by synthetic birnessite. Environ. Sci. Technol. 30:1990–1996.

Scribner, S.L., T.R. Benzing, S. Sun, and S.A. Boyd. 1992. Desorption and bioavailability of aged simazine residues in soil from a continuous corn field. J. Environ. Qual. 21:115–120.

Sharpley, A.N. 1983. Effect of soil properties on the kinetics of phosphorus desorption. Soil Sci. Soc. Am. J. 47:462–467.

Skopp, J. 1986. Analysis of time dependent chemical processes in soils. J. Environ. Qual. 15:205–213.

Sparks, D.L. 1989. Kinetics of soil chemical processes. Academic Press, San Diego, CA.

Sparks, D.L. 1991. Chemical kinetics and mass transfer processes in soils and soil constituents, p. 585–637. *In* J. Bear and M.Y. Corapcioglu (eds.) Transport processes in porous media. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Sparks, D.L. 1995. Environmental soil chemistry. Academic Press, San Diego, CA.

Sparks, D.L. 1998. Kinetics of sorption/release reactions on natural particles, p. 413–448. *In* P.M. Huang, N. Senesi, and J. Buffle (eds.) Structure and surface reactions of soil particles. John Wiley & Sons, New York.

Sparks, D.L. 1999a. Kinetics and mechanisms of chemical reactions at the soil mineral/water interface, p. 135–192. *In* D.L. Sparks (ed.) Soil physical chemistry. CRC Press, Boca Raton, FL.

Sparks, D.L. 1999b. Kinetics of soil chemical phenomena: Future directions, p. 81–102. *In* P.M. Huang, D.L. Sparks, and S.A. Boyd (eds.) Future prospects for soil chemistry. SSSA, Madison, WI.

Sparks, D.L. 2002. Environmental soil chemistry. 2nd edn. Academic Press, San Diego, CA.

Sparks, D.L. 2005. Sorption–desorption, kinetics, p. 556–561. *In* D. Hillel et al. (eds.) Encyclopedia of soils in the environment. Elsevier Ltd., Oxford, U.K.

Sparks, D.L., S.E. Fendorf, C.V. Toner, IV, and T.H. Carski. 1996. Kinetic methods and measurements, p. 1275–1307. *In* D.L. Sparks (ed.) Methods of soil analysis: Chemical methods. SSSA, Madison, WI.

Sparks, D.L., S.E. Fendorf, P.C. Zhang, and L. Tang. 1993. Kinetics and mechanisms of environmentally important reactions on soil colloidal surfaces, p. 141–168. *In* D. Petruzzelli and F.G. Helfferich (eds.) Migration and fate of pollutants in soils and subsoils. Springer-Verlag, Berlin, Germany.

Sparks, D.L., and P.M. Jardine. 1984. Comparison of kinetic equations to describe K–Ca exchange in pure and in mixed systems. Soil Sci. 138:115–122.

Sparks, D.L., and D.L. Suarez (eds.). 1991. Rates of soil chemical processes. SSSA, Special Publication No. 27, Madison, WI.

Sparks, D.L., and P.C. Zhang. 1991. Relaxation methods for studying kinetics of soil chemical phenomena, p. 61–94. *In* D.L. Sparks and D.L. Suarez (eds.) Rates of soil chemical processes. SSSA, Madison, WI.

Sposito, G. 1994. Chemical equilibria and kinetics in soils. Wiley, New York.

Steinberg, S.M., J.J. Pignatello, and B.L. Sawhney. 1987. Persistence of 1,2 dibromoethane in soils: Entrapment in intra particle micropores. Environ. Sci. Technol. 21:1201–1208.

Stone, A.T. 1986. Adsorption of organic reductants and subsequent electron transfer on metal oxide surfaces, p. 446–461. *In* J.A. Davis and K.F. Hayes (eds.) Geochemical processes at mineral surfaces. American Chemical Society, Washington, DC.

Stone, A.T. 1987a. Microbial metabolites and the reductive dissolution of manganese oxides: Oxalate and pyruvate. Geochim. Cosmochim. Acta 51:919–925.

Stone, A.T. 1987b. Reductive dissolution of manganese(III/IV) oxides by substituted phenols. Environ. Sci. Technol. 21:979–988.

Stone, A.T. 1991. Oxidation and hydrolysis of ionizable organic pollutants at hydrous metal oxide surfaces, p. 231–254. *In* D.L. Sparks and D.L. Suarez (eds.) Rates of soil chemical processes. SSSA, Madison, WI.

Stone, A.T., and J.J. Morgan. 1984a. Reduction and dissolution of manganese (III) and manganese (IV) oxides by organics. 1. Reaction with hydroquinone. Environ. Sci. Technol. 18:450–456.

Stone, A.T., and J.J. Morgan. 1984b. Reduction and dissolution of manganese (III) and manganese (IV) oxides by organics. 2. Survey of the reactivity of organics. Environ. Sci. Technol. 18:617–624.

Strawn, D.G., and D.L. Sparks. 2000. Effects of soil organic matter on the kinetics and mechanisms of Pb(II) sorption and desorption in soil. Soil Sci. Soc. Am. J. 64:144–156.

Stumm, W. 1992. Chemistry of the solid–water interface. Wiley, New York.

Stumm, W., and G. Furrer. 1987. The dissolution of oxides and aluminum silicates: Examples of surface-coordination-controlled kinetics, p. 197–219. *In* W. Stumm (ed.) Aquatic surface chemistry. Wiley Interscience, New York.

Stumm, W., and R. Wollast. 1990. Coordination chemistry of weathering. Kinetics of the surface-controlled dissolution of oxide minerals. Rev. Geophys. 28:53–69.

Tang, L.Y., and D.L. Sparks. 1993. Cation-exchange kinetics on montmorillonite using pressure-jump relaxation. Soil Sci. Soc. Am. J. 57:42–46.

Thompson, H.A., G.A. Parks, and G.E. Brown. 1999a. Dynamic interactions of dissolution, surface adsorption, and precipitation in an aging cobalt(II)–clay–water system. Geochim. Cosmochim. Acta 63:1767–1779.

Thompson, H.A., G.A. Parks, and G.E. Brown. 1999b. Ambient-temperature synthesis, evolution, and characterization of cobalt–aluminum hydrotalcite-like solids. Clays Clay Miner. 47:425–438.

Toner, C.V., IV, and D.L. Sparks. 1995. Chemical relaxation and double-layer model analysis of Boron adsorption on alumina. Soil Sci. Soc. Am. J. 59:395–404.

Towle, S.N., J.R. Bargar, G.E. Brown, Jr., and G.A. Parks. 1997. Surface precipitation of Co(II) (aq) on Al_2O_3. J. Colloid Interface Sci. 187:62–82.

van Genuchten, M.Th., and R.J. Wagenet. 1989. Two-site/two-region models for pesticide transport and degradation: Theoretical development and analytical solutions. Soil Sci. Soc. Am. J. 53:1303–1310.

Waychunas, G.A., B.A. Rea, C.C. Fuller, and J.A. Davis. 1993. Surface chemistry of ferrihydrite: Part 1. EXAFS studies of the geometry of coprecipitated and adsorbed arsenate. Geochim. Cosmochim. Acta 57:2251–2269.

Weber, W.J., Jr., and J.P. Gould. 1966. Sorption of organic pesticides from aqueous solution. Adv. Chem. Ser. 60:280–305.

Weber, W.J., and W.L. Huang. 1996. A distributed reactivity model for sorption by soils and sediments. 4. Intraparticle heterogeneity and phase-distribution relationships under non-equilibrium conditions. Environ. Sci. Technol. 30:881–888.

Weber, W.J., Jr., and C.T. Miller. 1988. Modeling the sorption of hydrophobic contaminants by aquifer materials. 1. Rates and equilibria. Water Res. 22:457–464.

Wersin, P., M.F. Hochella, Jr., P. Persson, G. Redden, J.O. Leckie, and D.W. Harris. 1994. Interaction between aqueous uranium (VI) and sulfide minerals: Spectroscopic evidence for sorption and reduction. Geochim. Cosmochim. Acta 58:2829–2843.

Wollast, R. 1967. Kinetics of the alteration of K-feldspar in buffered solutions at low temperature. Geochim. Cosmochim. Acta 31:635–648.

Wu, S., and P.M. Gschwend. 1986. Sorption kinetics of hydrophobic organic compounds to natural sediments and soils. Environ. Sci. Technol. 20:717–725.

Xia, K., A. Mehadi, R.W. Taylor, and W.F. Bleam. 1997. X-ray absorption and electron paramagnetic resonance studies of Cu(II) sorbed to silica: Surface-induced precipitation at low surface coverages. J. Colloid Interface Sci. 185:252–257.

Xing, B., and J.J. Pignatello. 1997. Dual-mode sorption of low-polarity compounds in glassy poly (vinyl chloride) and soil organic matter. Environ. Sci. Technol. 31:792–799.

Xing, B., and J.J. Pignatello. 2005. Sorption of organic chemicals. *In* D. Hillel (ed.) Encyclopedia of soils in the environment, p. 537–548. Vol. 3. Elsevier Ltd., Oxford, U.K.

Xue, Y., and S.J. Traina. 1996. Oxidation kinetics of Co(II)-EDTA in aqueous and semi-aqueous goethite suspensions. Environ. Sci. Technol. 30:1975–1981.

Zhang, P.C., and D.L. Sparks. 1989. Kinetics and mechanisms of molybdate adsorption/desorption at the goethite/water interface using pressure-jump relaxation. Soil Sci. Soc. Am. J. 53:1028–1034.

Zhang, P.C., and D.L. Sparks. 1990a. Kinetics and mechanisms of sulfate adsorption desorption on goethite using pressure-jump relaxation. Soil Sci. Soc. Am. J. 54:1266–1273.

Zhang, P.C., and D.L. Sparks. 1990b. Kinetics of selenate and selenite adsorption/desorption at the goethite/water interface. Environ. Sci. Technol. 24:1848–1856.

Zinder, B., G. Furrer, and W. Stumm. 1986. The coordination chemistry of weathering. II. Dissolution of Fe(III) oxides. Geochim. Cosmochim. Acta 50:1861–1869.

14

Oxidation–Reduction Phenomena

Bruce R. James
University of Maryland

Dominic A. Brose
University of Maryland

14.1 Concepts, Principles, and Theories

14.1.1 Nature of the Electron and Proton in Soil Solutions

Electrons are subatomic particles with wave-like properties that have defied exact characterization since their discovery in 1897, despite their central role in chemical reactions (Castellan, 1983). Electrons are considered fundamental subatomic particles, but recent findings suggest that they may possess a substructure of leptoquarks, that electron orbitals do not exist, and that the role of electrons in predicting periodicity of the elements is only approximate (Scerri, 1997). In addition, new frontiers are expanding to explain electron-transfer processes at interfaces (Tributsch and Pohlmann, 1998). Studying and understanding electron-transfer processes in soils, therefore, is challenging and must be based on sources of information from relatively simple aqueous systems and natural waters, as well as on a large base of empirical studies using soils. Electron-transfer processes in microbial cells are central to nutrient and pollutant chemical reactions, especially oxidation state changes governed by intra- and extracellular enzymes (Paul, 2007). Recent developments in environmental microbiology have found that bacteria form biofilms almost ubiquitously in aqueous and colloidal environments, such as natural waters, soils, and the human body (Stoodley et al., 2004). Therefore, understanding the role of microbes in controlling redox processes in soils requires an appreciation of such "matrix enclosed accretions that adhere to biological and nonbiological surfaces" (Stoodley et al., 2004).

To understand and characterize the nature of "electron activity" and oxidation–reduction reactions in biotic and abiotic, heterogeneous, multiphase soils and soil solutions, an appreciation of the characteristics of electrons and closely allied protons is needed (James, 1989; Stumm and Morgan, 1996; Bartlett, 1998; Bartlett and Ross, 2005). An examination of the complementary nature of electrons and protons affirms the importance of hydrogen ion and electron activities as master variables in soils (Sillén, 1967).

The H atom comprises one proton and one electron, and it may be visualized as a spherical puff of cotton candy with a radius of approximately 10 cm and a proton nucleus with a radius of 5 μm; essentially an invisible fleck of unspun sugar in the center. The remaining volume of the atom is occupied by the electron and the density of the spun sugar represents the probability of finding the electron in any one location. It becomes increasingly less dense (less probable) with distance away from the positively charged proton (Castellan, 1983). The radius of the H atom (0.3 Å), therefore, is approximately 20,000 times greater than that of the proton ($\sim 1.5 \times 10^{-5}$ Å). The proton also may be visualized as the size of a 0.1 μm colloidal clay particle, compared to a 2000 μm sand grain in a soil. In contrast to the large proportion of the volume of the H atom occupied by the negatively charged, wave-like electron, the electron has only negligible mass equal to approximately 550 μg mol^{-1}, 1/1836 of the mass of the H atom (10^6 μg mol^{-1}).

The tiny, heavy proton persists in hydrated form in aqueous media as H_3O^+, the hydroxonium or hydronium ion. The H atom

can only form the H^+ ion when its compounds are dissolved in media that solvate protons, so the H^+ cannot exist in solid phases (Cotton and Wilkinson, 1980). The solvation enthalpy of $-1091\,kJ$ mol^{-1} provides the energy for bond rupture. Protons migrate rapidly between water molecules, and an individual H_3O^+ ion has a lifetime of approximately 10^{-13} s. Its concentration and activity can, therefore, be measured and understood in terms of the hydration of a cation in solution. The single ion activity coefficient can be calculated based on ionic strength, temperature, effective diameter, and solvent characteristics; and H^+ activity can be calculated or measured (Westcott, 1978; Bates, 1981).

The extremely large "charge-to-size ratio" of the electron prevents it from persisting in free form in aqueous systems, as does the solvated H^+. The ephemeral "hydrated electron" is a powerful reducing agent with a potential of $-2.7\,V$ relative to the H^+/H_2 reference potential of $0.0\,V$ and has a half-life of $<1\,ms$. It reacts rapidly with second-order rate constants of 10^8–10^{10} M^{-1} s^{-1}, near the diffusion-controlled limit (Sullivan et al., 1976). Hydrated electrons may be transferred to aqueous solution from photoexcited hydrocarbons as $\bullet e_{aq}^-$. Such "free" or "aqueous" electrons react with H_2O and H_3O^+ at femtosecond rates (10^{-15} s) to form the "conjugate acid" of the electron, $H^\bullet$, and the reaction with H_3O^+ is approximately 10^7 times faster than with water (Leffler, 1993). The role of such aqueous electrons in soils remains unstudied, and photochemical studies of soils may reveal their presence and importance.

In soil chemical calculations and theory, one considers the electron as a "species," designated "e^-," with negligible mass and thermodynamically as a ligand, reactant, and product (Sposito, 1981). The electron is not ionic, but it is "negatively charged" as the carrier of negative electricity (Thompson, 1923). Its "activity" is conceptually analogous to that of H^+, but its concentration in "mol L^{-1}" is undefined. All these caveats about the electron require that one understand that electron activity in soils and natural waters should be regarded as related strictly to energy functions. Such functions can be described simply as "the ability to do work," "electrochemical potential," and more colloquially, "electron pressure."

Generically, for a reduction half-reaction: ox + ne$^-$ = red, the electron activity, (e^-), equals $[(red)/K(ox)]^{1/n}$, in which "ox" is the oxidized species, "red" is the reduced species, n is the number of moles of electrons, and K is the equilibrium constant. Therefore, electron activity is a function of the ratio of reduced-to-oxidized species activities and of the equilibrium constant [or Gibbs free energy for the reaction, since $\Delta G_r^o = -RT\ln K$, where R is the gas law constant ($8.313\,J\,K^{-1}\,mol^{-1}$) and T is the absolute temperature measured in Kelvin ($298\,K$ under standard conditions)]. A "favorable reaction" with a large positive value for K (or large negative ΔG_r^o) will therefore be associated with low electron activity. For example, a value of the equilibrium constant, $K = 6 \times 10^8$, corresponds to $\Delta G_r^o = -50\,kJ\,mol^{-1}$ and $K = 57$ corresponds to $-10\,kJ\,mol^{-1}$ (Compton and Sanders, 1996). That is, such reactions will proceed spontaneously at relatively low electron activities or pressures, corresponding to positive electrode voltages. In these examples, since $\Delta G_r^o = -nFE^o$, the voltages

(E^o) associated with the electron transfers are $+5.2$ and $+1.04\,V$, respectively, where n is the number of moles of electrons and F is the Faraday constant [96,487 C (coulombs) mol^{-1} or $J\,V^{-1}\,mol^{-1}$].

The concept of H^+ activity and pH also can be described in terms of thermodynamic work (Δw) defined as the product of an intensive property and extensive variation, such as P (ΔV), where P is pressure and ΔV is change in volume (Stumm and Morgan, 1996). Similarly, electron activity may be defined as E (Δe), where E is potential and Δe is change in charge of the system. Proton and electron activities also may be defined as chemical work, μ (Δn), where μ is chemical potential and Δn is change in moles (Sposito, 1981).

Viewing the sibling concepts of "proton activity" and "electron activity" in soils, they must not be considered twins. Recognition must be given to similarities and differences in the formulation of conceptual and operational definitions for these key variables, and such comparisons are based on the differences in the nature of the proton and electron, as described above. In both cases, however, their thermodynamic activities in a given system are relative to their activities under standard-state conditions. In the case of the e^-, the standard state is at $(H^+) = 1\,M$, a partial pressure of $H_2 = 1\,atm$, and a temperature of $298\,K$, characteristics of the standard hydrogen electrode (SHE) with $0\,V$ under these conditions (Lindsay, 1979; Compton and Sanders, 1996; Stumm and Morgan, 1996). For H^+ activity, the reference state is $1\,M$ or pH 0 (Bates, 1981).

The familiar concept of buffer capacity for pH is defined as the change in acid or base added (capacity factor) to effect a one unit change in pH (intensive variable; Stumm and Morgan, 1996). Similarly, the "capacity factor" in redox is referred to as poise and is defined as the change in added equivalents of reductant or oxidant to bring about a one unit change in pe (–log of electron activity) or an Eh (voltage relative to the SHE) change of $59\,mV$, as discussed in Section 14.1.2. Poise in soils and natural waters has been less intensively studied than pH buffering, but it is a central concept governed by reductant and oxidant activities and microbial processes.

The choice of oxidant or reductant to add in a redox titration will affect the calculated poise (resistance to change in pe as measured by the Pt electrode), and poise will be operationally defined by the couple interacting with the Pt wire. For instance, if the soil is titrated with Cr(VI), the Cr(VI)–Cr(III) couple will be potential determining, so the poise will reflect that pe range. If Fe(II) were added, the Fe(III)–Fe(II) couple would set a lower pe range than would Cr(VI)–Cr(III). Heron et al. (1994) titrated aquifer sediments with the reducing agent, Ti^{3+} EDTA, to determine the poise associated with oxidants, such as Fe(III) and Mn(III,IV) (hydr)oxides. Barcelona and Holm (1991) titrated with Cr^{2+} to quantify poise due to oxidants, and with $Cr_2O_7^{2-}$ to determine poise governed by reductants in the soil, respectively. Bauer et al. (2007) measured electron-transfer capacities and reaction kinetics of peat-dissolved organic matter, an organic matter analog for soil C. They used Zn and H_2S oxidation and complexed Fe(III) reduction to show that dissolved organic matter is a redox poising system between -0.9 and $1.0\,V$.

Redox intensity measurements (Eh or pe) and estimates of poise as the redox capacity of soils and sediments are also distinguished from those only associated with H^+ in that microbial

catalysis controls the rates of many of the oxidation and reduction reactions governing soil redox status (Russell, 1973; Banerjee, 2008). Also, the adaptation and growth of higher plants and algae in soils affects redox intensity and capacity. Examples of such microbial processes governing redox in soils include Fe(III) and SO_4^{2-} reduction in sediments (Coleman et al., 1993), denitrification in soil microenvironments (Højberg et al., 1994), N_2 fixation in flooded soils (e.g., due to pH and pe changes by the *Anabaena–Azolla* symbiosis in rice or that of the *Frankia* sp. actinomycete infection of alder (*Alnus* spp.; Buresh et al., 1980; Madigan and Martinko, 2006), methane oxidation in flooded rice soils (Wang et al., 1997), the reactions of H_2 in anoxic groundwater (Lovley et al., 1994), methanogenesis (Welsch and Yavitt, 2007; Goldhammer and Blodau, 2008; Saggar et al., 2008), and the formation of oxidized root channels (rhizospheres) in soils (Mendelssohn et al., 1995).

The poise and pe of soils also differ from natural waters in that the solids-to-solution ratio is high; and subsequently, redox reactions across colloid–solution interfaces are centrally important. Biofilm–water interfaces may be especially important in the case of microbially governed redox processes. Relatively few studies have been reported on redox reactions across soil colloid–water interfaces, but this area is critically important for predicting oxidation state and potential mobility of nutrients and pollutants in soil environments. Examples include reductive dissolution of Fe(III) and Mn(III,IV) (hydr)oxides and redox reactions on such surfaces or with constituent ions (Bartlett and James, 1979; Stone and Morgan, 1987; Suter et al., 1991; Stumm and Sulzberger, 1992; Postma and Jakobsen, 1996; White and Peterson, 1996).

14.1.2 Thermodynamic Relationships for Electron Activity in Soils

Since electrons are transferred from reductants (e^- donors or reducing agents) to oxidants (e^- acceptors or oxidizing agents) and do not persist free in soil solution, reduction reactions must be coupled to oxidation reactions to describe complete oxidation–reduction processes (redox reactions). By convention, reduction half-reactions can be written and thermodynamic relationships derived from them, despite the fact that they do not occur in isolation. The log K values (where K is the equilibrium constant for the half-reaction) for the coupled reactions may be compared to predict the likelihood of a reaction occurring spontaneously as written.

For example, development of reduction half-reactions to predict whether gaseous H_2S could reduce colloidal FeOOH across a sulfidic–aerobic interface in a wetland soil may be completed in the following manner:

1. Write the oxidized and reduced species on the left and right sides of the equation, respectively, with the appropriate number of electrons per mole of oxidant written on the left side. Balance moles of elements other than O and H:

$$FeOOH + e^- = Fe^{2+}, \tag{14.1}$$

$$SO_4^{2-} + 8e^- = H_2S. \tag{14.2}$$
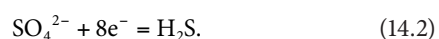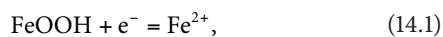
The number of electrons needed for reduction is calculated from the oxidation numbers or states of the elements in the oxidized and reduced species (Vincent, 1985). In this example, the oxidation numbers are III+ and VI+ for Fe and S in FeOOH and SO_4^{2-}, respectively.

2. Add H_2O to the equation (usually to the right side) to balance the moles of O:

$$FeOOH + e^- = Fe^{2+} + 2H_2O, \tag{14.3}$$

$$SO_4^{2-} + 8e^- = H_2S + 4H_2O. \tag{14.4}$$
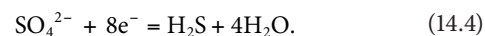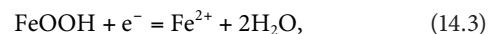
3. Then add H^+ to balance moles of H, usually on the left side of the equation:

$$FeOOH + e^- + 3H^+ = Fe^{2+} + 2H_2O, \tag{14.5}$$

$$SO_4^{2-} + 8e^- + 10H^+ = H_2S + 4H_2O. \tag{14.6}$$
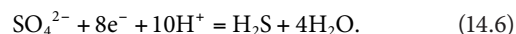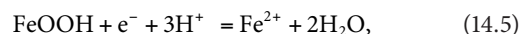
4. Check charge and mass balances for the equations.

These equations can now be used to develop expressions for electron activity based on equilibrium expressions and free energy of formation (ΔG_f^o) data.

The principles for doing this can be understood by starting with a generic reduction half-reaction:

$$Aox + Be^- + CH^+ = Dred + EH_2O \tag{14.7}$$
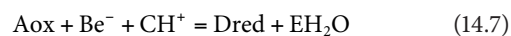
where A, B, C, D, and E are the stoichiometric coefficients. The expression for the equilibrium constant (K) is as follows:

$$K = \frac{\left[(red)^D (H_2O)^E\right]}{\left[(ox)^A (e^-)^B (H^+)^C\right]} \tag{14.8}$$

where
() denotes activity
(H_2O) has a value of 1, by convention

Taking the log of both sides of the equation,

$$\log K = \log\left[\frac{(red)^D}{(ox)^A}\right] + \log\left[\frac{1}{(e^-)^B}\right] + \log\frac{1}{(H^+)^C} \tag{14.9}$$

and

$$\log K = D\log(red) - A\log(ox) + Bpe + CpH \tag{14.10}$$

where
pe and pH are defined as −log electron activity
H^+ activity, respectively.

The "p" notation means power and denotes the exponent for the H^+ activity (pH 4 or 10^{-4} mol H^+ kg^{-1} water). Therefore, $(H^+) = 10^{-pH}$. In contrast, the exponent for e^- activity is not 10^{-pe}, since the electron

activity is not defined in terms of "mol e⁻ L⁻¹." Both pe and pH are analogous, though, if viewed as related to the ability of e⁻ and H⁺ to do thermodynamic work (Section 14.1.1).

Further rearrangement of Equation 14.10 yields a pe–pH relationship of the following form:

$$\left[\left(\frac{1}{B}\right)\log K - \left(\frac{D}{B}\right)\log(\text{red}) + \left(\frac{A}{B}\right)\log(\text{ox})\right] - \left(\frac{C}{B}\right)\text{pH} = \text{pe},$$

(14.11)

which gives a straight line with slope C/B and the intercept in square brackets that is a function of log K for the half-reaction and the activities of the oxidized and reduced species.

For a one-electron transfer (B = 1) coupled with one-proton consumption (C = 1), and when D = A and (red) = (ox),

$$\text{pe} + \text{pH} = \log K. \qquad (14.12)$$

And when pH = 0 (standard state),

$$\text{pe} = \log K. \qquad (14.13)$$

Knowledge of pe and pH is pertinent to describing the equilibrium condition of a soil as defined by the master variables, pe, pH, and their sum (Lindsay, 1979). The concepts of electron activity and hydrogen ion activity are closely coupled and cannot be separated in assessing the oxidation–reduction status of a soil system. The pe + pH parameter has been described metaphorically by Bartlett (1998) as a seesaw with a hungry baby bird on one side (the electron-deficient player) and earthworms on the other (the electron-rich source). The parent bird carries the earthworms to the baby, representing electron transfer and causing a shift in the position of the seesaw, quantified by the pe–pH balance. As worms are transferred, the pe increases and the pH decreases, countering and interacting controls on the redox equilibrium. As the worms are consumed by the baby, the seesaw shifts back again, representing a dynamic, metastable system governed by the flow of energy and the cycling of nutrients, characteristics of ecosystems. The position of the seesaw governed by pe and pH reflects the equilibrium condition of a soil, and the values of pe and pH are intensity measures of redox status. The change in measured pe and pH per mole of electrons plus H⁺ (H atoms in principle) transferred reflects the buffer capacity or poise of the system, $\Delta(\text{pe} + \text{pH})/\Delta$ (mol H atoms transferred).

To relate the concept of log K to the Gibbs free energy change (ΔG_r^o) for a given half-reaction in the soil, the following expressions are pertinent:

$$\Delta G_r^o = -RT \ln K \qquad (14.14)$$

where

ΔG_r^o is Gibbs free energy of reaction under standard conditions (298 K and 1 atm)

R is the universal gas law constant (0.00831 kJ K⁻¹ mol⁻¹)

T is absolute temperature (298 K)

Converting to log₁₀ (ln K = 2.303 log K) yields the following:

$$\frac{\Delta G_r^o}{-5.70} = \log K. \qquad (14.15)$$

log K may therefore be estimated simply from knowledge of free energies of formation of H_2O, the red, and the ox; since those of H⁺ and e⁻ are zero, by convention. To relate log K directly to pe as defined by Equation 14.11 for one-electron transfers, log K values must be divided by B, the mol of e⁻ consumed in the reaction. Therefore, pe at a given pH and other defined conditions is equivalent to log K, and the parameter pe + pH couples the master variables in defining the equilibrium redox state of a soil–water system.

Log K values also are related to thermodynamic electrochemical potentials (Eh) according to the following expressions (Compton and Sanders, 1996):

$$\Delta G_r^o = -nFEh \qquad (14.16)$$

where

n is the number of e⁻ transferred mol⁻¹

F is the Faraday constant (96.5 kJ V⁻¹ equivalent⁻¹ or 96,487 C mol⁻¹ where C is coulombs, derived from 1.60×10^{-19} C/e⁻ × 6.02×10^{23} e⁻/mol)

Since both Equations 14.16 and 14.14 are expressions for ΔG_r^o,

$$-nFEh = -RT \ln K \qquad (14.17)$$

and, when T = 298 K,

$$Eh = \frac{RT\, 2.303 \log K}{nF} = 0.0592 \log K. \qquad (14.18)$$

If n = 1, (red) = (ox), and pe = log K at a given pH (Equation 14.12), then

$$Eh\ (V)\ =\ 0.0592\,\text{pe} \qquad (14.19)$$

or

$$\frac{Eh}{0.0592} = \text{pe}. \qquad (14.20)$$

Therefore, calculating or interpreting Eh values rigorously and linking soil measurements to thermodynamics require knowledge of the pH of the soil–water system.

Applying these principles to the H_2S–FeOOH problem above (Equations 14.5 and 14.6), the following pe–pH expressions can be derived:

For FeOOH–Fe²⁺:

$$\text{pe} = \log K_{Fe} - \log(Fe^{2+}) - 3\text{pH} \qquad (14.21)$$

where activity of FeOOH is assumed to be 1, an assumption that may or may not be valid for redox processes (Stumm, 1993; White and Peterson, 1996).

For SO_4^{2-}–H_2S:

$$pe = \frac{1}{8}\log K_S - \frac{1}{8}\log\left[\frac{P_{H2S}}{SO_4^{2-}}\right] - \frac{5}{4}pH \qquad (14.22)$$

where

K_{Fe} and K_S are the equilibrium constants for the FeOOH and SO_4^{2-} reduction half-reactions

P_{H_2S} is the partial pressure of H_2S

Calculating log K values and substituting into Equations 14.21 and 14.22 yield the following:

$$pe = 13.0 - \log(Fe^{2+}) - 3pH \qquad (14.23)$$

and

$$pe = 5.21 - \left(\frac{1}{8}\right)\log\left[\frac{P_{H_2S}}{SO_4^{2-}}\right] - \left(\frac{5}{4}\right)pH. \qquad (14.24)$$

At defined activities for the ions and partial pressure for H_2S, pe may be calculated, and if pe for Fe(III) reduction > pe for SO_4^{2-} reduction, then FeOOH will be reduced to Fe^{2+} and H_2S oxidized to SO_4^{2-} at equilibrium and at the hypothetical soil interface. The same result will be obtained if after subtracting the log K value for S from that for Fe, a positive answer is obtained.

The Equations 14.23 and 14.24 can be plotted on an *x–y*-graph called a pe–pH diagram (Figure 14.1). In this graph, each reduction half-reaction is plotted for (red)/(ox) = 1 and 10^{-4}. Note that
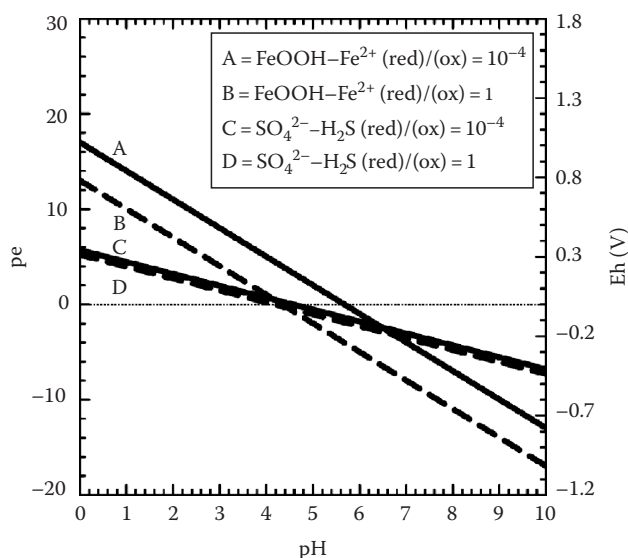
FIGURE 14.1 pe–pH diagram for FeOOH–Fe^{2+} and SO_4^{2-}–H_2S equilibria at two activity ratios.

the positions of the lines are shifted to higher pe values for the lower activities of the reduced species; the effect being greater for the FeOOH–Fe^{2+} couple than for SO_4^{2-}–H_2S, due to the greater oxidation state change per mole for S than for Fe. This effect of changing the activity ratio on pe values is an easy way to perform a sensitivity analysis for these thermodynamic models and to plot a family of curves representative of a range of activities relevant to any soil system or conditions. Similar calculations can be performed for variations in the values of ΔG_f^o, since the conditions (e.g., temperature, pressure, and ionic strength) under which these values are determined may result in a range of values of up to 10 or more kJ mol^{-1}. The ΔG_f^o values tabulated by Lindsay (1979) were reviewed and selected for their accuracy and their relevance to soils and natural waters.

An important use of pe–pH diagrams is for predicting thermodynamically whether or not the oxidized species in one reduction half-reaction will oxidize the reduced species in a second reduction half-reaction (i.e., when the second half-reaction is reversed to become the oxidation half-reaction and is added to the first reduction half-reaction). When the line for a pe–pH couple is above another over a pH range, the calculated Δpe or ΔEh values will be >0 (and corresponding ΔG_r^o values <0), signifying that the reaction is spontaneous as written. In the above example, FeOOH is expected to oxidize H_2S over the pH range 0–6.6 for the activity ratio = 1 but only up to pH 4.6 for the activity ratio = 10^{-4}. Theoretically, above the crossover pH values, SO_4^{2-} would be expected to oxidize Fe^{2+} to FeOOH, while being reduced to H_2S; the opposite of the reaction at pHs < the crossover value.

Calculations of myriad pe–pH relationships are conveniently performed using a spreadsheet template that incorporates balanced half-reactions, appropriate ΔG_f^o values to calculate log K, chosen activity ratios, and Equation 14.11. Using such a spreadsheet, numerous pe values for common reduction half-reactions pertinent to soil chemical and biochemical processes were calculated and are presented in Table 14.1. Table 14.1 also includes numerous pe values derived from Eh data in various published databases, especially for coenzymes and some intermediates governing microbial respiration. This list provides a quick, semi-quantitative means of assessing what the products of selected, coupled reduction and oxidation half-reactions are likely to be.

The U.S. Environmental Protection Agency developed MINTEQ (for "Mineral Thermal Equilibria"), a DOS-based computer program that calculates the distribution of myriad species based on minimizing the Gibbs free energy of a system of multiple chemical equilibria. Different from earlier such programs, such as GEOCHEM, it incorporates algorithms for redox reactions and allows the estimation of pe, pH, and reductant and oxidant activities in conjunction with dissolution and exchange equilibria. The use of MINTEQ also permits an easy way to conduct a sensitivity analysis for estimated pe, pH, or ion activities when just one value for a given parameter is varied (Allison and Brown, 1995). Visual MINTEQ was written and is updated for Microsoft Windows by J.P. Gustafsson, and version 3.0 (as of August 2010) is now available free online at http:/www.lwr.kth.se/English/OurSoftware/vminteq/

TABLE 14.1 Reduction Half-Reactions for Soils

	Log K^a	pe^b	
	(pe at pH 0)	pH 5	pH 7
Nitrogen species			
$\frac{1}{2}N_2O + e^- + H^+ = \frac{1}{2}N_2 + \frac{1}{2}H_2O$	29.8	22.9	20.9
$NO + e^- + H^+ = \frac{1}{2}N_2O + \frac{1}{2}H_2O$	26.8	19.8	17.8
$\frac{1}{2}NO_2^- + e^- + \frac{3}{2}H^+ + \frac{1}{4}N_2O + \frac{3}{4}H_2O$	23.6	15.1	12.1
$\frac{1}{5}NO_3^- + e^- + \frac{6}{5}H^+ = \frac{1}{10}N_2 + \frac{3}{5}H_2O$	21.1	14.3	11.9
$NO_2^- + e^- + 2H^+ = NO + H_2O$	19.8	9.8	5.8
$\frac{1}{4}NO_3^- + e^- + \frac{5}{4}H^+ = \frac{1}{8}N_2O + \frac{5}{8}H_2O$	18.9	12.1	9.6
$\frac{1}{6}NO_2^- + e^- + \frac{4}{3}H^+ = \frac{1}{6}NH_4^+ + \frac{1}{3}H_2O$	15.1	8.4	5.7
$\frac{1}{8}NO_3^- + e^- + \frac{5}{4}H^+ = \frac{1}{8}NH_4^+ + \frac{3}{8}H_2O$	14.9	8.6	6.1
$\frac{1}{2}NO_3^- + e^- + H^+ = \frac{1}{2}NO_2^- + \frac{1}{2}H_2O$	14.1	9.1	7.1
$\frac{1}{6}NO_3^- + e^- + \frac{7}{6}H^+ = \frac{1}{6}NH_2OH + \frac{1}{3}H_2O$	11.3	5.4	3.1
$\frac{1}{6}N_2 + e^- + \frac{4}{3}H^+ = \frac{1}{3}NH_4^+$	4.6	−0.7	−3.3
Oxygen species			
$OH^\bullet + e^- + H^+ = H_2O$	—	—	36.8
$O_3^{\bullet-} + e^- + 2H^+ = H_2O + O_2$	—^{c}	—	30.5^d
$\frac{1}{2}O_3 + e^- + H^+ = \frac{1}{2}O_2 + \frac{1}{2}H_2O$	35.1	28.4	26.4
$OH^\bullet + e^- = OH^-$	33.6	33.6	33.6
$O_2^- + e^- + 2H^+ = H_2O_2$	32.6	22.6	18.6
$\frac{1}{2}H_2O_2 + e^- + H^+ = H_2O$	30.0	23.0	21.0
$HO_2^\bullet + e^- + H^+ = H_2O_2$	—	—	17.9^d
$O_3 + e^- = O_3^{\bullet-}$	—	—	15.1^d
$^1O_2 + e^- = O_2^{\bullet-}$ (singlet)	—	—	14.1^d
$\frac{1}{4}O_2 + e^- + H^+ = \frac{1}{2}H_2O$	20.8	15.6	13.6
$H_2O_2 + e^- + H^+ = OH^\bullet + H_2O$	—	—	7.8
$\frac{1}{2}O_2 + e^- + H^+ = \frac{1}{2}H_2O_2$	11.6	8.2	6.2
$^3O_2 + e^- = O_2^{\bullet-}$ (triplet, ground state)	−9.5	−6.2	−6.2
$^3O_2 + e^- + H^+ = HO_2^\bullet$ (triplet, ground state)	—	—	−7.8
Sulfur species			
$\frac{1}{8}SO_4^{2-} + e^- + \frac{5}{4}H^+ = \frac{1}{8}H_2S + \frac{1}{2}H_2O$	5.2	−1.0	−3.5
$\frac{1}{6}SO_4^{2-} + e^- + \frac{4}{3}H^+ = \frac{1}{6}S + \frac{2}{3}H_2O$	5.3	−1.4	−4.0
$\frac{1}{2}S + e^- + H^+ = \frac{1}{2}H_2S$	4.9	−0.1	−2.1
$\frac{1}{2}SO_4^{2-} + e^- + 2H^+ = \frac{1}{2}SO_2 + H_2O$	2.9	−7.1	−11.1
Iron and manganese compounds			
$\frac{1}{2}Mn_3O_4 + e^- + 4H^+ = \frac{3}{2}Mn^{2+} + 2H_2O$	30.7	16.7	8.7
$\frac{1}{2}Mn_2O_3 + e^- + 3H^+ = Mn^{2+} + \frac{3}{2}H_2O$	25.7	14.7	8.7
$Mn^{3+} + e^- = Mn^{2+}$	25.5	25.5	25.5
$\gamma\text{-MnOOH} + e^- + 3H^+ = Mn^{2+} + 2H_2O$	25.4	14.4	8.4
$0.62MnO_{1.8} + e^- + 2.2H^+ = 0.62Mn^{2+} + 1.1H_2O$	22.1	13.4	8.9
$\frac{1}{2}Fe_3(OH)_8 + e^- + 4H^+ = \frac{3}{2}Fe^{2+} + 4H_2O$	21.9	7.9	−0.1
$\frac{1}{2}MnO_2 + e^- + 2H^+ = \frac{1}{2}Mn^{2+} + H_2O$	20.8	12.8	8.8
$[Mn^{3+}(PO_4)_2]^{3-} + e^- = [Mn^{2+}(PO_4)_2]^{4-}$	20.7	20.7	20.7
$Fe(OH)_2^+ + e^- + 2H^+ = Fe^{2+} + 2H_2O$	20.2	10.2	6.2
$\frac{1}{2}Fe_3O_4 + e^- + 4H^+ = \frac{3}{2}Fe^{2+} + 2H_2O$	17.8	3.9	−4.1
$MnO_2 + e^- + 4H^+ = Mn^{3+} + 2H_2O$	16.5	0.54	−7.5

TABLE 14.1 (continued)   Reduction Half-Reactions for Soils

	Log K[a]	pe[b]	
	(pe at pH 0)	pH 5	pH 7
$Fe(OH)_3 + e^- + 3H^+ = Fe^{2+} + 3H_2O$	15.8	4.8	−1.2
$FeOH^{2+} + e^- + H^+ = Fe^{2+} + H_2O$	15.2	10.2	8.2
$\frac{1}{2}Fe_2O_3 + e^- + 3H^+ = Fe^{2+} + \frac{3}{2}H_2O$	13.4	2.4	−3.6
$FeOOH + e^- + 3H^+ = Fe^{2+} + 2H_2O$	13.0	2.0	−4.0
$Fe^{3+} + e^- = Fe^{2+}$ (phenanthroline)	18.0	—[c]	—
$Fe^{3+} + e^- = Fe^{2+}$ (H_2O only)	13.0	13.0	13.0
$Fe^{3+} + e^- = Fe^{2+}$ (acetate)	—	5.8	—
$Fe^{3+} + e^- = Fe^{2+}$ (malonate), pH 4	—	4.4	—
$Fe^{3+} + e^- = Fe^{2+}$ (salicylate), pH 4	—	4.4	—
$Fe^{3+} + e^- = Fe^{2+}$ (hemoglobin)	—	—	2.4
$Fe^{3+} + e^- = Fe^{2+}$ (EDTA)	—	—	2.0[d]
$Fe^{3+} + e^- = Fe^{2+}$ cyt b_3 (plants)	—	—	0.68
$Fe^{3+} + e^- = Fe^{2+}$ (DTPA)	—	—	0.51
$Fe^{3+} + e^- = Fe^{2+}$ (oxalate)	—	—	0.034
$Fe^{3+} + e^- = Fe^{2+}$ (pyrophosphate)	−2.4	—	—
$Fe^{3+} + e^- = Fe^{2+}$ (peroxidase)	—	—	−4.6
$Fe^{3+} + e^- = Fe^{2+}$ (spinach ferredoxin)	—	—	−7.3
$Fe^{3+} + e^- = Fe^{2+}$ (ferritin)	—	—	−3.2[d]
$\frac{1}{3}KFe_3(SO_4)_2(OH)_6 + e^- + 2H^+ = Fe^{2+} + 2H_2O + \frac{2}{3}SO_4^{2-} + \frac{1}{3}K^+$	8.9	6.9	2.9
$[Fe(CN)_6]^{3-} + e^- = [Fe(CN)_6]^{4-}$	—	—	6.1
Carbon species			
$CH_3CH_2^\bullet + e^- + H^+ = CH_3CH_3$	—	—	32.2[d]
$RO^\bullet + e^- + H^+ = ROH$ (aliphatic)	—	—	27.1[d]
$CO_3^{\bullet-} + e^- + H^+ = CO_3^{2-}$	—	—	25.1[d]
$Tryptophan^\bullet + H^+ + e^- = tryptophan$	—	—	16.9
$Catechol\text{-}O^\bullet + e^- + H^+ = catechol$	—	—	9.0[d]
$Benzosemiquinone + e^- + H^+ = hydroquinone$	—	—	7.8[d]
$\frac{1}{2}Ubiquinone + H^+ + e^- = ubiquinone\text{-}H_2$	—	—	1.7
$\frac{1}{2}o\text{-}quinone + e^- + H^+ = \frac{1}{2}diphenol$	—	—	5.9
$\frac{1}{2}p\text{-}quinone + e^- + H^+ = \frac{1}{2}hydroquinone$	—		4.7
$Cu^{2+} + H^+ + e^- = Cu^+$ (superoxide dismutase)	—	—	7.1
$Mn^{3+} + H^+ + e^- = Mn^{2+}$ (superoxide dismutase)	—	—	5.2
$Fe^{3+} + H^+ + e^- = Fe^{2+}$ (superoxide dismutase)	—	—	4.7
$\frac{1}{2}CH_3OH + e^- + H^+ = \frac{1}{2}CH_4 + \frac{1}{2}H_2O$	9.9	4.9	2.9
$\frac{1}{12}C_6H_{12}O_6 + e^- + H^+ = \frac{1}{4}C_2H_5OH + \frac{1}{4}H_2O$	4.4		−1.9
$\frac{1}{8}CO_2 + e^- + H^+ = \frac{1}{8}CH_4 + \frac{1}{8}H_2O$	2.9	−2.1	−4.1
$\frac{1}{2}CH_2O + e^- + H^+ = \frac{1}{2}CH_3OH$	2.1	−2.9	−4.9
$\frac{1}{2}HCOOH + e^- + H^+ = \frac{1}{2}CH_2O + \frac{1}{2}H_2O$	1.5	−3.5	−5.5
$\frac{1}{4}CO_2 + e^- + H^+ = \frac{1}{24}C_6H_{12}O_6 + \frac{1}{4}H_2O$	−0.21	−5.9	−7.9
$\frac{1}{2}Dehydroascorbate + e^- + H^+ = \frac{1}{2}ascorbic\ acid$	1.0	−3.5	−5.5
$\frac{1}{2}Methylene\ blue\,(ox) + H^+ + e^- = \frac{1}{2}methylene\ blue\,(red)$	—	—	0.2
$FAD + H^+ + e^- = \frac{1}{2}FADH_2$	—	—	−3.0
$\frac{1}{2}Pyruvate + H^+ + e^- = \frac{1}{2}lactate$	—	—	−3.2
$\frac{1}{2}NAD^+ + H^+ + e^- = \frac{1}{2}NADH + \frac{1}{2}H^+$	—	—	−5.4

(continued)

TABLE 14.1 (continued)   Reduction Half-Reactions for Soils

	Log K[a]	pe[b]	
	(pe at pH 0)	pH 5	pH 7
$\frac{1}{2}$ Pyruvate + $\frac{1}{2}$ CO$_2$ + H$^+$ + e$^-$ = $\frac{1}{2}$ malate	—	—	−5.6
Ferredoxin-Fe^{3+} + e$^-$ = ferredoxin-Fe^{2+}	—	—	−7.3
$\frac{1}{4}$ CO$_2$ + e$^-$ + H$^+$ = $\frac{1}{4}$ CH$_2$O + $\frac{1}{4}$ H$_2$O	−1.2	−6.1	−8.1
$\frac{1}{2}$ CO$_2$ + e$^-$ + H$^+$ = $\frac{1}{2}$ HCOOH	−1.9	−6.7	−8.7
$\frac{1}{2}$ Acetate + H$^+$ + e$^-$ = $\frac{1}{2}$ acetaldehyde	—	—	−10.1
$\frac{1}{2}$ Acetate + CO$_2$ + H$^+$ + e$^-$ = $\frac{1}{2}$ pyruvate	—	—	−11.8
CO$_2$ + e$^-$ = CO$_2$$^{\bullet-}$	—	—	−30.5
Pollutant/nutrient group			
Co^{3+} + e$^-$ = Co^{2+}	30.6	30.6	30.6
$\frac{1}{2}$ NiO$_2$ + e$^-$ + 2H$^+$ = $\frac{1}{2}$ Ni^{2+} + H$_2$O	29.8	21.8	17.8
PuO$_2$$^+$ + e$^-$ = PuO$_2$	26.0	22.0	22.0
$\frac{1}{2}$ PbO$_2$ + e$^-$ + 2H$^+$ = $\frac{1}{2}$ Pb^{2+} + H$_2$O	24.8	16.8	12.8
PuO$_2$ + e$^-$ + 4H$^+$ = Pu^{3+} + 2H$_2$O	9.9	−6.1	−14.1
$\frac{1}{3}$ HCrO$_4$$^-$ + e$^-$ + $\frac{4}{3}$ H$^+$ = $\frac{1}{3}$ Cr(OH)$_3$ + $\frac{1}{3}$ H$_2$O	18.9	10.9	8.2
$\frac{1}{2}$ H$_2$AsO$_4$$^-$ + e$^-$ + $\frac{3}{2}$ H$^+$ = $\frac{1}{2}$ H$_3$AsO$_3$ + $\frac{1}{2}$ H$_2$O	11.2	3.7	0.7
Hg^{2+} + e$^-$ = $\frac{1}{2}$ Hg$_2$$^{2+}$	15.4	13.4	13.4
$\frac{1}{2}$ MoO$_4$$^{2-}$ + e$^-$ + 2H$^+$ = $\frac{1}{2}$ MoO$_2$ + H$_2$O	15.0	3.0	−1.0
$\frac{1}{2}$ SeO$_4$$^{2-}$ + e$^-$ + H$^+$ = $\frac{1}{2}$ SeO$_3$$^{2-}$ + $\frac{1}{2}$ H$_2$O	14.9	9.9	7.9
$\frac{1}{4}$ SeO$_3$$^{2-}$ + e$^-$ + $\frac{3}{2}$ H$^+$ = $\frac{1}{4}$ Se + $\frac{3}{4}$ H$_2$O	14.8	6.3	3.3
$\frac{1}{6}$ SeO$_3$$^{2-}$ + e$^-$ + $\frac{4}{3}$ H$^+$ = $\frac{1}{6}$ H$_2$Se + $\frac{1}{2}$ H$_2$O	7.62	1.0	−1.7
$\frac{1}{2}$ VO$_2$$^+$ + e$^-$ + $\frac{1}{2}$ H$_3$O$^+$ = $\frac{1}{2}$ V(OH)$_3$	6.9	2.4	1.4
Cu^{2+} + e$^-$ = Cu$^+$	2.6	2.6	2.6
PuO$_2$ + e$^-$ + 3H$^+$ = PuOH^{2+} + H$_2$O	2.9	−8.1	−14.1
Paraquat + e$^-$ = paraquat$^{\bullet-}$	—	—	−7.6
Analytical couples			
CeO$_2$ + e$^-$ + 4H$^+$ = Ce^{3+} + 2H$_2$O	47.6	31.6	23.6
$\frac{1}{2}$ ClO$^-$ + e$^-$ + H$^+$ = $\frac{1}{2}$ Cl$^-$ + $\frac{1}{2}$ H$_2$O	29.0	24.0	22.0
HClO + e$^-$ + H$^+$ = $\frac{1}{2}$ Cl$_2$ + H$_2$O	27.6	20.6	18.6
$\frac{1}{2}$ Cl$_2$ + e$^-$ = Cl$^-$	23.0	25.0	25.0
$\frac{1}{6}$ IO$_3$$^-$ + e$^-$ + H$^+$ = $\frac{1}{6}$ I$^-$ + $\frac{1}{2}$ H$_2$O	18.6	13.6	11.6
$\frac{1}{2}$ Pt(OH)$_2$ + e$^-$ + H$^+$ = $\frac{1}{2}$ Pt + H$_2$O	16.6	11.6	9.6
$\frac{1}{2}$ I$_2$ + e$^-$ = I$^-$	9.1	11.1	11.1
$\frac{1}{2}$ Hg$_2$ Cl$_2$ + e$^-$ = Hg + Cl$^-$	4.5	3.9	3.9
AgCl + e$^-$ = Ag + Cl$^-$	3.8	3.2	3.2[e]
e$^-$ + H$^+$ = $\frac{1}{2}$ H$_2$	0	−5	−7
$\frac{1}{2}$ PtS + e$^-$ + H$^+$ = $\frac{1}{2}$ Pt + $\frac{1}{2}$ H$_2$S	−5.0	−10.0	−12.0

[a] Calculated for reaction as written according to Equation 14.14. Free energy of formation data were taken from Lindsay (1979) as a primary source, and when not available from that source, from Garrels and Christ (1965) and Loach (1976).

[b] Calculated using tabulated log K values, (red) and (ox) = 10^{-4} mol/L soluble ions and molecules, activities of solid phases = 1, and partial pressures for gases that are pertinent to soils: 10^{-4} atm for trace gases, 0.21 atm for O$_2$, 0.78 for N$_2$, and 0.00032 for CO$_2$. Note: pe × 59.2 = Eh (mV) at the specified pH.

[c] Values not listed by Loach (1976) or Larson (1997).

[d] Values from Larson (1997).

[e] Value for 4 M Cl$^-$ in Ag/AgCl reference electrode.

14.1.3 Kinetic Derivation of Thermodynamic Parameters for Redox

This thermodynamic derivation of log K and its relationship to pe and pH of soils is based on free energy of formation data for oxidants and reductants, and it is not related to reaction mechanisms or rates. An alternate procedure to obtain log K values, in theory, is the kinetic approach suggested by Sparks (1985) for cation exchange and by Harter and Smith (1981) for adsorption processes. This approach has been little used in redox soil chemistry, probably due to difficulties in obtaining rate coefficients for many electron-transfer processes in soils and due to the irreversibility of most redox reactions. The application of such approaches should, however, be appropriate for certain reversible redox reactions in soils, especially in situations where metastability and lack of chemical equilibrium prevail or when accurate free energy of formation data is unavailable for oxidants and reductants.

In principle, this is the method for calculating kinetic parameters: the equilibrium constant (K) for redox equilibria can be estimated from the ratio of forward and reverse rate coefficients for a given reversible reaction in soils:

$$K = \frac{k_f}{k_r} \tag{14.25}$$

where k_f and k_r are the rate constants for forward and reverse reactions, respectively. The standard Gibbs free energy for the reaction then can be calculated by substituting ln K into Equation 14.14. Preparation of Arrhenius plots of the rate coefficients as a function of 1/T permits estimation of activation energies for the forward and reverse reactions, and standard enthalpies for the reactions can be calculated as

$$\Delta H^o = E_f^\star - E_r^\star \tag{14.26}$$

where

ΔH^o is enthalpy for the redox reaction

$E_f^\star$ and $E_r^\star$ are the activation energies for the forward and reverse reactions

With knowledge of ΔG^o and ΔH^o, standard entropy of the reaction can be calculated as

$$\Delta S^o = \frac{\Delta H^o - \Delta G^o}{T}. \tag{14.27}$$

Since the e^- and H^+ are key reactants and products in the thermodynamic sense, a refined knowledge of mechanisms for particular redox reactions is needed because of our current limitations in understanding how similar or different e^- and H^+ reactions are in soils and natural waters. Knowledge of thermodynamic stability for a redox system does not necessarily predict kinetic lability, a concept directly pertinent to reactivity of different types of complexes (Cotton and Wilkinson, 1980). The kinetic lability of redox processes in soils has not been compared systematically with predictions of stability based on thermodynamic data,

but linear free energy relationships (LFERs) are useful plots of observed rate constants versus equilibrium constants or Eh values (Stumm and Morgan, 1996; Bürge and Hug, 1997). Such information could affirm the reliability of using thermodynamic data to predict bioavailability of nutrients and pollutants, and to estimate true reactivity of electron donors and acceptors in nonequilibrium, kinetically governed soil chemical and microbial environments.

Given the constraining limitation of irreversibility of most redox reactions in soils, especially microbial ones, for using experimental data to calculate values for thermodynamic parameters, the use of kinetic knowledge on soil redox processes is less widely used than are pe (and Eh) and pH for assessing equilibrium conditions of soils. Empirical studies in relatively pure, aqueous systems, and complex soil suspensions have, however, provided pertinent information on the kinetics of redox processes and their relation to thermodynamic parameters and reaction pathways (Stumm, 1992, 1993; Stumm and Morgan, 1996; Hug et al., 1997; Typrin, 1998).

14.1.4 Microbiological and Enzymatic Controls for Redox State and Processes in Soils

Bacterial metabolism, the process of producing adenosine triphosphate (ATP) through a series of catabolic reactions and building cell constituents with anabolic reactions, requires energy, electrons, nutrients, and either oxygen or another highly oxidized compound to serve as a terminal electron acceptor (TEA; Sylvia et al., 2005). Bacteria that use O_2 as the electron acceptor are aerobic, those that utilize another oxidized compound in the absence of oxygen are anaerobic, and those capable of both are facultative anaerobes. The use of external TEAs intricately links bacteria to important geochemical processes, such as denitrification, methane production, and metal reduction (Salminen et al., 2006; Banerjee, 2008).

The metabolic classification of bacteria is based on the energy source, electron source (reducing equivalents), and carbon source used in catabolic and anabolic processes. Bacteria can derive energy from natural sunlight or from chemical reactions. Phototrophy is the process of using natural light for energy, as employed by photosynthesizing algae or cyanobacteria, whereas chemotrophy requires a chemical reaction for metabolic growth (Madigan and Martinko, 2006). All nonphotosynthetic cellular organisms utilize this pathway to derive the energy needed for metabolic reactions. Electrons (reducing equivalents) can be derived from either organic or inorganic compounds. Organotrophy is the process of using organic compounds, and lithotrophy is that of using inorganic compounds. Organotrophic bacteria are generally associated with chemotrophy; however, lithotrophic bacteria can be either phototrophs or chemotrophs. Cellular C needed for metabolic processes can be derived from organic compounds, which is referred to as heterotrophy, or from CO_2 (dissolved CO_2, H_2CO_3, HCO_3^-, or CO_3^{2-} in natural waters and soil solution; the predominant

species depending on pH), which is autotrophy. Autotrophy is generally linked with lithotrophy, and heterotrophy is associated with organotrophy (Sylvia et al., 2005).

The oxidation of an organic or inorganic compound and reduction of O_2 or other oxidized species during bacterial metabolism is the coupling of oxidation–reduction half-reactions (redox processes) and is the basis for bioenergetics. Biologically mediated redox reactions in soils are linked to metabolic processes and will follow a thermodynamic sequence based on the energetic possibility of the reaction (Stumm and Morgan, 1996). Aerobic chemoorganoheterotrophs use organic C compounds as a source of energy, electrons, and C; and they use O_2 as the TEA. The half-reactions for the reduction of O_2 and oxidation of an organic compound (generically represented by CH_2O with the oxidation state of C = 0) from Table 14.1 can be coupled to show that the overall $\Delta G_r^o = -130 \, kJ \, mol^{-1}$. An anaerobic chemoorganoheterotroph present in an O_2-depleted environment, such as a wetland soil, can utilize the same C source but would require another TEA. Using $Fe(OH)_3$ instead of O_2 from Table 14.1 would result in an overall $\Delta G_r^o = -102 \, kJ \, mol^{-1}$. This example illustrates how the use of O_2, when available, is thermodynamically more favorable for metabolic growth than other oxidized species.

The reduction of O_2 to H_2O in aerobic respiration involves four, one-electron reduction steps, shown with pe values at pH 7:

$$O_2 + e^- + H^+ = HO_2^\bullet$$

Formation of superoxide radical (pe = −7.8),   (14.28)

$$HO_2^\bullet + e^- + H^+ = H_2O_2$$

Formation of hydrogen peroxide (pe = 17.9),   (14.29)

$$H_2O_2 + e^- + H^+ = HO^\bullet + H_2O$$

Formation of hydroxyl radical (pe = 7.8),   (14.30)

$$HO^\bullet + e^- + H^+ = H_2O$$

Formation of water (pe = 36.8),   (14.31)

$$\frac{1}{4}O_2 + e^- + H + = \frac{1}{2}H_2O$$

Overall reaction (pe = 13.6).   (14.32)
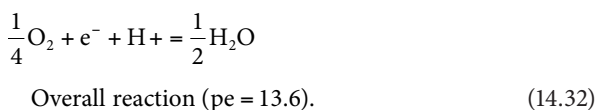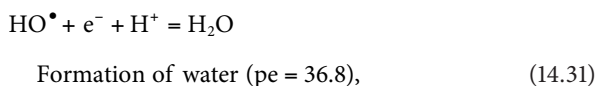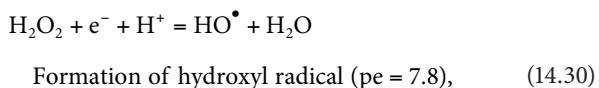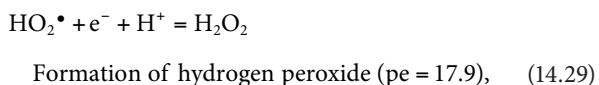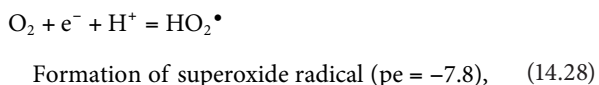
An examination of the pe values for the stepwise reduction of O_2 to H_2O demonstrates that even though the overall reaction has a pe = 13.6 (confirming that it is a good oxidizing agent), superoxide, hydrogen peroxide, and hydroxyl radical intermediates are more powerful oxidants than O_2. Superoxide dismutase or reductase, catalase or peroxidase, and glucose-6-dehydrogenase are essential enzymes used by aerobic cells. They dismutate superoxide and hydrogen peroxide and subsequently reduce hydroxyl radicals more rapidly than the reactive oxygen

species can react with cell constituents, thereby allowing aerobic respiration to proceed. Hydroxyl radicals are rarely formed in microbial cells due to the efficiency of H_2O_2 dismutation to H_2O and O_2, following reduction of $HO_2^\bullet$. Deficiencies of some or all of these enzymes render soil microbes incapable of respiration under fully aerobic conditions but better adapted to microaerophilic or anaerobic conditions (Madigan and Martinko, 2006).

In aerobic respiration, the coupling of C oxidation to O_2 reduction occurs within the bacterial cell either by substrate-level phosphorylation or oxidative phosphorylation, which is carried out by an intermediate compound such as the coenzyme nicotinamide adenine dinucleotide (NAD^+) or nicotinamide adenine dinucleotide phosphate ($NADP^+$) Madigan and Martinko, 2006). Although intermediaries are utilized in oxidative phosphorylation, the overall net change in energy is the difference in reduction potential between the primary electron donor and TEA. The electron-carrying coenzyme NADH or NADPH transfers electrons to a series of four complexes of the mitochondrial respiratory chain, which conserves energy by proton pumping, which further drives the synthesis of ATP (Banerjee, 2008). It is at complex IV, the final complex of electron transfer in oxidative phosphorylation, that oxygen is reduced to water by cytochrome c oxidase in a four-electron, one-step reaction (Equations 14.28 through 14.31). In substrate-level phosphorylation, ATP is directly produced in the enzymatic oxidation of an organic substance but with much lower yields than obtained via oxidative phosphorylation (Madigan and Martinko, 2006).

Calculating the Gibbs free energy of a reaction may indicate that the use of an electron donor and acceptor in a reaction is thermodynamically favorable, but it does not indicate at what rate the reaction would occur. The use of a substrate, although energetically favorable, may not occur at a biologically relevant rate without the catalytic action of an enzyme. Enzymes are proteins configured in secondary, tertiary, or quaternary structures that allow for specific binding properties and often contain non-protein prosthetic groups, cofactors, and coenzymes that participate in the catalytic functions of the enzyme (Madigan and Martinko, 2006). In order to obtain substrates located outside of the cell, enzymes (e.g., hydrolases) can be excreted to break down insoluble polymers such as cellulose, starch, and proteins. These degraded macromolecules are then transported into the cell for use in metabolic reactions (Sylvia et al., 2005). Soil microorganisms using extracellular enzymes are important in the initial depolymerization of cellulose. White rot and brown rot fungi are key microorganisms in the breakdown of cellulose, hemicellulose, and lignin in woody plant tissue in contact with soil.

Either through intracellular or extracellular mechanisms, once substrates are in the cell; oxidative phosphorylation requires a TEA, and in oxygen-depleted environments, bacteria must use alternative oxidized species. Nitrate (NO_3^-), Fe(III) (hydr)oxides, Mn(III,IV) (hydr)oxides, SO_4^{2-}, and CO_2 are oxidized species found in soil, which can be reduced in the terminal step of oxidative phosphorylation, and the reduced forms are excreted from the cell (Sylvia et al., 2005). The reduction of

nitrate coupled to the oxidation of pyrite in anoxic groundwater sediments was recently demonstrated to be a biologically driven redox process (Jorgensen et al., 2009).

In addition to soluble constituents, such as NO_3^- or SO_4^{2-}, as TEAs, the use of less soluble minerals, such as Fe(III) (hydr)oxides or Mn(III,IV) (hydr)oxides, can also be accomplished. The mechanism by which microbial cells interact with the molecular lattice of minerals is yet to be fully understood. It is speculated that structural or chemical defects, or points in the solid where energy has been perturbed, can serve as regions of reactivity or possibly access sites for organisms (Fredrickson and Zachara, 2008).

Nonenzymatic, extracellular strategies for moving electrons from inside the cell to the outside in electron-transfer processes have been observed. One such strategy is the use of a protein on the cell wall surface that extends away from the cell or a molecule that passes between the cell and substrate when a bacterium lacks the c-type cytochrome necessary to reduce a metal internally (Rawlings, 2005; Gralnick and Newman, 2007; Dittrich and Luttge, 2008). When at a distance from a mineral, *Geobacter sulfurreducens* develops pili, or nanowires, which are localized to one side of the cell wall, and extend out to allow electron transfers to the mineral surface of Fe(III) (hydr)oxides (Reguera et al., 2005).

In addition to localizing proteins to the exterior of the outer membrane, another strategy bacteria use in extracellular respiration is to utilize small organic shuttles to transfer electrons to substrates at a distance from the cell. Quinones are ubiquitous structures found in nature and are principal constituents of soil organic matter (SOM). In the human body, coenzyme Q or ubiquinone is present in every cell, and the ubiquinone/ubiquinol redox couple plays a role in shuttling electrons between complex II and III of the mitochondrial electron transport chain during oxidative phosphorylation (Table 14.1; Banerjee, 2008). Ubiquinone also acts as a lipid-soluble antioxidant by accepting one electron and forming a stable radical anion semiquinone.

This same quinone structure is present in soils and natural waters as the primary abiotic electron-donating and accepting moiety of humic and fulvic acids in natural organic matter (Scott et al., 1998; Ratasuk and Nanny, 2007).

Phenolic groups are abundant in natural organic matter and are important in complexation reactions with metals. Phenolic groups can be biologically oxidized by the enzymes phenolase and laccase to produce quinones (Tan, 2003). Converse to this oxidation, a one-electron transfer to quinone forms the same, highly reactive intermediate semiquinone as with ubiquinone and then the second electron transfer forms hydroquinone (Larson, 1997). Hydroquinone is then capable of donating its two electrons in further reduction reactions, making the quinone–hydroquinone species a very dynamic redox constituent in soils. The natural reducing capacities of International Humic Substance Society (IHSS) humic acids were shown to increase several-fold following microbial reduction and were significant even under aerobic conditions (Peretyazhko and Sposito, 2006).

The electron-donating kinetics of dissolved organic matter in aquatic systems was shown to occur within the course of 1 day and capable of shuttling electrons to SO_4^{2-} and Fe(III) (Bauer et al., 2007).

Quinone moieties in natural organic matter can serve as TEAs in microbial respiration, cycling electrons from bacteria to mineral surfaces (Lovley et al., 1996; Bond and Lovley, 2002). Soil bacteria reduced the humic acid analog, anthraquinone-2,6-disulfonate (AQDS), which then shuttled electrons to reduce Fe(III) (Fredrickson et al., 1998; Kappler et al., 2004). The shuttling activity of AQDS was demonstrated further with *G. sulfurreducens* shuttling electrons to ferrihydrite in solution (Straub and Schink, 2003). This use of humic substances as a TEA may allow for more efficient respiration than directly using a mineral. The rate of electron transfer between reduced humic substances to ferrihydrite is approximately 27 times faster than the rate of *G. sulfurreducens* directly transferring electrons to ferrihydrite (Jiang and Kappler, 2008).

The biological reduction of electron shuttles, such as soluble humic acids, is important in soils, but abiotic reduction pathways also significantly contribute to the natural redox cycling of metals and nutrients. Lactic acid, tartaric acid, and citric acid were shown to reduce AQDS, which then shuttled electrons to enhance the reduction of Cr(VI) in solution (Brose and James, 2010). Low-molecular-weight organic acids, such as these, are produced naturally by microbial activity or plant root exudates and can be important reducing agents for quinone moieties in soils and soluble humic acids, thus further contributing to natural redox process in soils. Figure 14.2 illustrates the conceptual model of how citric acid reduces the electron shuttle AQDS to AH_2DS, which ultimately reduces Cr(VI) to Cr(III).

Electrons are in constant flux in soils, being transferred from electron donors to electron acceptors, often mediated by soil microorganisms, which results in metastable redox conditions in soils. Though thermodynamic equilibrium models may predict steady-state conditions, in reality, a dynamic nonequilibrium state is a more accurate representation of natural systems (Stumm and Morgan, 1996). Often, redox reactions in soils are moving

1.5 Citric acid → 1.5 3-Ketoglutarate

3e⁻

1.5 AQDS → 1.5 AH₂DS

3e⁻

Cr(III)(hydr)oxide ← Cr(VI)
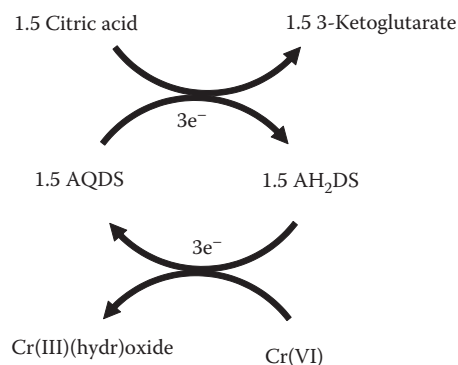
FIGURE 14.2 The stoichiometric transfer of electrons from citric acid to AQDS, reducing it to the hydroquinone form AH_2DS, which ultimately reduces Cr(VI) to Cr(III).

toward equilibrium, but the reactions are constantly being perturbed by the influx and efflux of chemical species and soil biota, thus shifting activities and reestablishing new, partial equilibria. Dioxygen is the most ubiquitous electron acceptor in natural systems, when available, and the electron activity and type of microorganisms present in a soil depend, in part, on its presence.

Because a saturated soil restricts the diffusion of O_2, saturated soils allow for more electron flow to other electron acceptors, such as Fe(III) (hydr)oxides and SO_4^{2-} (Bartlett and James, 1993). When a soil becomes submerged, pe decreases, and the rate of decrease will be a function of easily oxidized SOM, temperature, time submerged, and nature of available electron acceptors (Sparks, 2003), which are also important metabolic requirements of soil microorganisms. Though soils host a variety of microorganisms using a diverse array of electron acceptors for respiration, such as Fe(III) (hydr)oxides in anoxic sediments, SOM is the dominant electron donor and can be fully oxidized to CO_2 (Lovley, 2000), or partially to organic acids. Aerobic microorganisms will thrive until O_2 is depleted, and then facultative and obligate anaerobic microorganisms will be the dominant biological mediators.

Bacteria and their electron-shuttling systems are widely used in remediation strategies for organic and inorganic pollutants. Humus and other quinone analogs were used in the reductive biotransformation of nitroaromatic compounds and to dechlorinate organic pollutants (Field et al., 2000; Becker, 2006). In the anaerobic oxidation of phenol and *p*-cresol, it was shown that the addition of the quinone analog AQDS diverted electrons from methanogenesis to quinone reduction, a more thermodynamically favorable reaction (Cervantes et al., 2000).

The bioremediation of metals, such as Cr(VI) and U(VI) has also proven to be an effective strategy for treating contaminated soils and groundwater (Gu and Chen, 2003; Nyman et al., 2007; Wu et al., 2007). The cell wall of *Arthrobacter oxydans* contains an acid-soluble protein with a positive charge, as shown by electrophoresis, capable of reducing Cr(VI) to an insoluble Cr(III) (hydr)oxide (Asatiani et al., 2004). Bacteria can transform Hg compounds to either less toxic or less bioavailable species. *Desulfovibrio desulfuricans* API strain degrades CH_3Hg^+ by producing H_2S that reacts with CH_3Hg^+ forming the insoluble species $(CH_3Hg)_2S$ (Hobman et al., 2000). The dissimilative reduction of Se(VI), a soluble, toxic form of Se, to the insoluble Se(0) by anaerobic bacteria in anoxic environments has important implications for the continued irrigation of seleniferous soils and industrial applications such as the production of pesticides and semiconductors (Oremland and Stolz, 2000).

As bacterial metabolic activities are continually studied, more ways will be discovered to apply the use of microorganisms to transform pollutants to less toxic or less bioavailable forms. The adaptation of bacteria to a wide variety of electron donors and acceptors provides almost unlimited application of biologically mediated redox reactions. Understanding how bacteria link the oxidation of electron donors to the reduction of electron acceptors will provide insight into further utilizing these mediated reactions to improve ecosystems and natural waters.

14.2 Methods and Procedures

14.2.1 Uses of pe–pH Thermodynamic Information

Values of log K derived from Gibbs free energy of formation data or kinetic evaluations of redox reactions provide tools for predicting if a reduction half-reaction coupled with an oxidation half-reaction will allow the spontaneous transfer of electrons from reductant to oxidant. Since soils are only metastable and highly heterogeneous in nature, such predictive capability is necessary to formulate hypotheses for many processes that may occur in the field, even if they require microbial catalysis or other coupled reactions to occur at ambient temperatures and pressures of soil–water–plant systems.

The reduction half-reactions listed in Table 14.1 are biological and abiotic species of N, O, Mn, Fe, S, C, various pollutants sensitive to redox conditions in soils and several reactions pertinent to the analysis or characterization of redox conditions. Within groups, the half-reactions are arranged in descending order of log K values, calculated as described above. These values are pe at pH = 0 when activities of oxidant and reductant are 1 and may be considered standard, reference pe values for the reactions (formal potentials). The pe values listed at pH 5 and 7 are calculated to represent typical activities of ions and partial pressures of gases in soil environments.

Higher log K or pe values indicate greater "ease of reduction" of an oxidant (left side of equation) to its reduced form than do lower values. This means that for predictive purposes, an oxidant in a particular reduction half-reaction is able to oxidize the reductant in another half-reaction with a lower pe, at a specified pH, as demonstrated quantitatively for the FeOOH–Fe^{2+} and SO_4^{2-}–H_2S couples. Another example shows Mn(III,IV) (hydr) oxides would be expected to oxidize $Cr(OH)_3$ to Cr(VI) at pH 5 since the range of pe values for reduction of Mn (12.8–16.7) is greater than that for Cr(VI) reduction (10.9). This has been demonstrated to occur in most field-moist soils in the pH range of 4–7 containing oxides of Mn(III,IV) (Bartlett and James, 1979; James and Bartlett, 1983). Conversely, these oxides would not be expected to oxidize N_2 to N_2O (pe at pH 5 = 22.9).

Even though the log K or pe for one reduction half-reaction is less than that for a second half-reaction, the reduced form in the first reaction may still be oxidized by the oxidized species in the second half-reaction (the reverse of the above concept). For example, the pe at pH 5 for reduction of CO_2 to $C_6H_{12}O_6$ is −5.9 and that for reduction of O_2 to H_2O is 15.6. Based on this difference, one would predict that reduction of O_2 to H_2O would be coupled to oxidation of $C_6H_{12}O_6$ to CO_2, and coupling oxidation of H_2O to O_2 with reduction of CO_2 to $C_6H_{12}O_6$ would not be thermodynamically probable. In fact, both respiration (the predicted reaction) and photosynthesis (the second, "improbable" reaction) occur together, and the balance of the two is responsible for the existence of the aerobic lifestyle and the persistence of SOM. Photosynthesis (represented simply as $CO_2 \rightarrow C_6H_{12}O_6$) is made possible by a complex series of coupled reactions that

make an overall thermodynamically improbable reaction occurring rapidly in sunlight.

Similarly, a reaction predicted to be thermodynamically probable may not occur at any appreciable rate under natural conditions. The pe for NO_3^- reduction to N_2 at pH 5 (14.3) is greater than that for $HCrO_4^-$ reduction to $Cr(OH)_3$ (10.9), but this NO_3^- oxidation of Cr(III) has not been demonstrated in soils or plants, probably because the reduction of NO_3^- requires enzymatic catalysis to lower the energy of activation at such a high pH.

The order of log K values for reduction half-reactions also has been used to predict the sequence of reduction reactions carried out by respiring soil microorganisms following saturation of a soil (Ponnamperuma, 1972). The descending order of preference (pe) for the electron acceptors at pH 7 (proportional to free energy derived from the reduction) is O_2–H_2O (13.6), NO_3^-–N_2 (11.9), MnO_2 –Mn^{2+} (8.8), $Fe(OH)_3$–Fe^{2+} (–1.2), SO_4^{2-}–H_2S (–3.5), and CO_2–CH_4 (–4.1). Heterotrophic bacteria are using organic compounds as the electron donors and their energy source (chemoorganoheterotrophs) in their respiration to produce CO_2 or organic acids (pe range at pH 7 of –8.7 to –3.1), so most of the organic compounds can be used throughout the reduction sequence following depletion of atmospheric O_2.

14.2.2 Use of pe–pH Diagrams

14.2.2.1 Oxygen Species

While the pe for reduction of O_2 to H_2O ranges from 20.8 at pH 0 to 13.6 at pH 7 (Table 14.1), the intermediates associated with one-electron transfers show a wide fluctuation in their oxidizing power (Equations 14.28 through 14.32), a property of O_2 that is pertinent to understanding the transition in soils from "aerobic" to "anaerobic" conditions. Anaerobic respiratory enzymes are typically produced when P_{O_2} reaches approximately 1% of atmospheric levels (0.0021 atm). The data in Table 14.2 also indicate that the pe for reduction of O_2 is relatively insensitive to the O_2 partial pressure in this range.

Ozone (O_3) and the hydroxyl free radical (HO•) are the most powerful oxidants among the oxygen species (Table 14.1), and the latter may be formed during stepwise, four-electron reduction of O_2 to O_2^-, H_2O_2, and H_2O (Fridovich, 1978; Equations 14.28 through 14.32). The high and low pe values for superoxide reduction to H_2O_2 and for superoxide oxidation to O_2 indicate that both a powerful oxidant and a powerful reductant agent are formed in the first of four, one-electron steps in the reduction of O_2 to H_2O. The enzyme superoxide dismutase scavenges superoxide in living cells using O_2 as the TEA (respiring aerobically), but relatively little is known about its reactivity in biological and chemical processes in soils that may be pertinent to our understanding of the formation of highly reduced components such as SOM and highly oxidized species such as NO_3^- that coexist in soil under non- or quasi-equilibrium conditions.

The wide range of reduction potentials for O_2 and its partially reduced intermediates, coupled with biological processes controlling the partial pressure of this gas in soil solution, create

TABLE 14.2 Sensitivity of Calculated pe at pH 0 to Variation in ΔG_f^o and Activities of Oxidant and Reductant in Selected Half-Reactions Pertinent to Soils

Couple	−log Activity		ΔG_f^o		pe	Δpe^a
	Ox	Red	Ox	Red		
O_2/H_2O	0.68	0	0.00	−237.01	20.61	
	2.68	0	0.00	−237.01	20.11	−0.50
	0.68	0	0.00	−278.81	24.20	3.59
NO_3/N_2	4	0.11	−110.48	0.00	20.30	
	6	0.11	−110.48	0.00	19.90	−0.40
	4	0.11	−152.28	0.00	18.80	−1.50
Mn_3O_4/Mn^{2+}	0	4	−1279.92	−227.39	36.70	
	0	6	−1279.92	−227.39	39.70	3.00
	0	4	−1321.72	−227.39	33.00	−3.70
MnO_2/Mn^{2+}	0	4	−464.40	−227.39	22.80	
	0	6	−464.40	−227.39	23.80	1.00
	0	4	−506.20	−227.39	19.10	−3.70
$MnOOH/Mn^{2+}$	0	4	−556.36	−227.39	29.40	
	0	6	−556.36	−227.39	31.40	2.00
	0	4	−598.16	−227.39	22.10	−7.30
$CO_2/C_6H_{12}O_6$	3.5	4	−394.01	−913.66	−0.91	
	3.5	6	−394.01	−913.66	−0.83	0.08
	3.5	4	−394.01	−955.46	−0.60	−0.31
$Fe(OH)_3/Fe^{2+}$	0	4	−712.27	−91.12	19.80	
	0	6	−712.27	−91.12	21.80	2.00
	0	4	−754.07	−91.12	12.40	−7.40
Fe_2O_3/Fe	0	4	−740.28	−91.12	17.40	
	0	6	−740.28	−91.12	19.40	2.00
	0	4	−782.08	−91.12	13.70	−3.70
SO_4^{2-}/H_2S	4	4	−743.83	−33.52	5.20	
	6	4	−743.83	−33.52	5.50	0.30
	4	4	−702.03	−33.52	6.10	0.90

[a] Change in calculated pe resulting from a change in activity of the ox or red species (column 1 or 2) or resulting from use of a ΔG_f^o value that is 44 kJ mol^{-1} (10 kcal) different from the published value (first row, column 3 or 4).

conditions under which the O_2–H_2O system may rarely attain equilibrium. Therefore, only metastable conditions and a slow approach to chemical equilibrium characterize O_2 behavior, making thermodynamic predictions difficult for aerobic soils (Bartlett, 1981).

14.2.2.2 Nitrogen Species

Most reduction reactions of N species (Table 14.1) are not reversible, are biologically mediated, and therefore are not well defined by thermodynamic pe–pH relationships. The series of half-reactions composing the process of denitrification, though, is instructive in that it identifies the wide range in pe for reduction of each of the intermediates believed to form in the sequence of electron acceptors used by microbes:

$$NO_3^- \xrightarrow[(1)]{} NO_2^- \xrightarrow[(2)]{} NO \xrightarrow[(3)]{} N_2O \xrightarrow[(4)]{} N_2$$
Steps
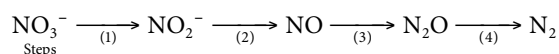
Step (1) of the sequence occurs at pe values less than those for reduction of O_2 to H_2O, while those for steps (2), (3), and (4) are increasingly higher. The overall reduction of NO_3^- to N_2 is almost identical to that for the O_2–H_2O couple. The similar pe range for the O_2 and NO_3^- reduction intermediates indicates that denitrification and aerobic respiration may occur at the same time under certain conditions when organic C is used as the electron donor. They may not be mutually exclusive as predicted from log K values for the overall reactions, O_2 to H_2O and NO_3^- to N_2.

14.2.2.3 Manganese Oxide Species

Manganese exists in soils in the II+, III+, and IV+ oxidation states, and the latter two are most stable as oxides or oxyhydroxides. Trivalent Mn may exist as $Mn(H_2O)_6^{3+}$, especially if stabilized by ligands, such as pyrophosphate or citrate. The pe values (Table 14.1) predict that different oxidation states of Mn in Mn_3O_4, MnOOH, and MnO_2 affect the pe at which Mn^{2+} would be expected to form at pH 5 (pe values of 16.7, 14.4, and 12.8, respectively), but they are all similar at pH 7 (8.7, 8.4, and 8.8). The Mn^{3+}–Mn^{2+} couple indicates that at pH ≈ 5, Mn^{3+} is a powerful oxidant (pe 25.5) similar to superoxide (pe 22.6) if in equilibrium with Mn^{2+}. At pH ≈ 6.5, Mn^{3+} in equilibrium with MnO_2 is a powerful reductant similar to H_2 and again, superoxide. This predicted reducing energetics of Mn^{3+} may be pertinent to anaerobic soils that are exposed to O_2, and in which Mn^{2+} is oxidizing to form Mn(III,IV) (hydr)oxides via Mn^{3+}. In oxidized soils containing MnO_2, flooding and the process of becoming reduced may produce Mn^{3+}, which is a powerful reducing agent. The trivalent Mn species may be ephemeral intermediates in such processes at redox interfaces, such as in the rhizosphere of plant roots or between the vadose zone and groundwater.

Since many Mn(III,IV) (hydr)oxides are nonstoichiometric and no compound with the exact composition of MnO_2 is known (Arndt, 1981), predictions of their redox properties as a function of mineralogy or oxidation state in heterogeneous soils may be hard to formulate. Despite the uncertainty of thermodynamic predictions for the redox behavior of Mn, the chemistry of this element is pertinent to a number of processes governing speciation and oxidation state of trace elements and pollutants found in soils (Bartlett and Ross, 2005). The pe–pH data indicate that oxides of Mn may oxidize Pu(III) to Pu(IV), V(III) to V(V), As(III) to As(V), Se(IV) to Se(VI), N(III) to N(V), and Cr(III) to Cr(VI) since the pe for each of these couples falls below that for Mn oxides (Table 14.1). The oxidations of Pu(III), As(III), Se(IV), N(III), and Cr(III) all have been demonstrated to occur in soils containing Mn(III,IV) (hydr)oxides or by synthetic Mn(III,IV) (hydr)oxides (Bartlett and James, 1979; Amacher and Baker, 1981; Bartlett, 1981; Moore et al., 1990; Blaylock and James, 1994).

The instability of Mn^{3+} and its ability to dismutate, similar to H_2O_2 and superoxide, mean that kinetic constraints and very low steady-state concentrations in soil solution may be particularly important in understanding the redox behavior of Mn in soils undergoing transitions between anaerobic and aerobic conditions. The kinetic lability of these species is poorly understood and new knowledge could contribute significantly to predictions of bioavailability and toxicity of numerous plant nutrients and pollutants in a range of soil types from rice paddies and wetlands to well-drained agricultural and forest soils.

14.2.2.4 Iron Species

While predictions of redox behavior of Fe(II) and Fe(III) species indicate that they fall below most Mn(III,IV) (hydr)oxide species (lower pe values and less free energy released per equivalent upon reduction), intermediate hydrolysis products, such as $Fe(OH)_2^+$, theoretically can oxidize Cr(III) to Cr(VI) at pH values <4 (Table 14.1). In addition, thermodynamically more stable complexation of Fe^{3+} by organic and inorganic ligands such as OH^-, relative to complexation of Fe^{2+}, lowers the pe at which Fe^{3+} is reduced to Fe^{2+}. These Fe(II)–Fe(III)-ligand reduction potentials are similar to those of Fe(III) (hydr)oxides in the pH range 5–7 (Table 14.1). This phenomenon suggests that Fe(II)–Fe(III)-ligand systems create Fe(II) species that are more powerful reductants than hexaquo Fe^{2+} if the Fe(III)-ligand complex is more thermodynamically stable than the Fe(II)-ligand complex. Conversely, if the complex with Fe(II) is more stable than that with Fe(III) (e.g., with phenanthroline), the Fe(II) becomes a less powerful reductant than hexaquo Fe^{2+}. This may explain the ability of a Fe(II,III) system to act as a cofactor in enzymes involved in redox processes, such as peroxidases and superoxide dismutases. These enzymes reduce or dismutate H_2O_2 and superoxide. The application of such concepts to abiotic redox processes in soils remains a key area for future research.

14.2.2.5 Carbon and Sulfur Species

Reduced forms of C and S are normally viewed as reductants in soils, either in abiotic or biological processes. Thermodynamic predictions support this idea for carbohydrates produced in photosynthesis, CH_4 from methanogenesis, and H_2S from reduction of SO_4^{2-} (Table 14.1). The reduction reactions of *o*- and *p*-quinone, suggest that these compounds may be reduced at higher pe values than are CO_2 and SO_4^{2-}. These pe values, however, coincide with the MnO_2/Mn^{3+} couple at pH 7, suggesting that Mn^{3+} may act as a reducing agent for certain organic species in near neutral soils. The coupling of reduction of the organic species with oxidation of Mn may result in formation of free radical species. This is pertinent to understanding the formation and persistence of SOM in high pH soils that may contain reactive forms of Mn(III,IV) (hydr)oxides (Bartlett and Ross, 2005).

Reactions of H_2S and H_2Se are predicted to be similar with respect to SO_4^{2-} and SeO_3^{2-} formation (Table 14.1). While SeO_4^{2-} and SO_4^{2-} are similar chemically, the oxidation of SeO_3^{2-} to SeO_4^{2-} is predicted to occur at higher pe values than that of H_2S to SO_4^{2-} (Table 14.1). Blaylock and James (1994) observed that Mn(III,IV) (hydr)oxides in soils or in pure form will oxidize SeO_3^{2-} to SeO_4^{2-}, as predicted by thermodynamics. They also observed that adding reducing phenolic acids, such as gallic and ascorbic acids, actually enhanced this oxidation reaction of SeO_3^{2-}. They hypothesized that partial reduction of MnO_2

in soils converted the Mn(IV) (hydr)oxide into a Mn(III) (hydr)oxide or ion that is a more powerful oxidant for SeO_3^{2-} than is Mn(IV) oxide. Such a hypothesis is supported by the relative oxidizing power of MnO_2, $MnOOH$, and Mn_3O_4, where the latter two oxides contain Mn(III) (Table 14.1).

14.2.3 Measurement of Soil Redox and Acid–Base Status

The most common method for quantifying electron activity of soils and natural waters is to measure the potential difference between a Pt indicator electrode and a calomel or Ag/AgCl reference electrode, both connected to a voltmeter or a pH meter (Pearsall and Mortimer, 1939; Patrick and DeLaune, 1972; Rowell, 1981; Bricker, 1982; Patrick et al., 1996; Rabenhorst et al., 2009). In this method, the Pt electrode is presumed to be inert and does not react chemically while coming into equilibrium with electroactive species in soil solution and on soil colloids (Compton and Sanders, 1996). Figure 14.3 illustrates how a Pt electrode establishes an equilibrium condition with soil solution. When the electrode contacts the soil solution (and while coupled to a reference electrode), a tiny quantity of electrons jumps either from the soil solution to the electrode (when soil is more reduced than the electrode) or in the opposite direction when the soil solution is more oxidized. In the first case, the Pt electrode becomes more negative relative to the soil, and this lowers the voltage at the electrode–soil solution interface. In the second, the electrode surface and the potential become more positive.

A high input impedance ($\geq 20\,\text{m}\Omega$) is needed to obtain precise and accurate electrode potentials, because if current flows under low impedance, drift will be observed in the voltage readings (Rabenhorst et al., 2009), and redox reactions may take place at the electrode–soil solution interface. Such reactions may form interfering redox couples on the Pt metal (e.g., $Pt(OH)_2$–Pt or PtS–Pt; see Table 14.1). Commercially available pH meters typically have high input impedance, but inexpensive voltmeters that are sometimes used in field studies with multiple Pt electrodes do not.

Generally, Pt electrode measurements are only semiquantitative for assessments of redox status, especially of aerobic soils in the field (Whitfield, 1974; Bartlett, 1981; Grenthe et al., 1992; Hostettler, 1992; Grundl, 1994). Other methods that employ analysis of soil solution analytes indicative of redox status, along with thermodynamic half-reactions, as discussed above, may prove more reliable for calculating pe ranges for aerobic and anaerobic soil systems (Peiffer et al., 1992; Kludze et al., 1994; Lovley et al., 1994; Stumm and Morgan, 1996; Typrin, 1998).

14.2.3.1 Construction and Use of Pt Electrodes

Platinum and suitable reference electrodes are relatively easy and inexpensive to construct (Mueller et al., 1985; Farrell et al., 1991), but measurement technique may significantly alter measured voltages (Bartlett, 1981; Bricker, 1982; Matia et al., 1991). These researchers have described several aspects of electrode use and misuse with respect to the reliability of recorded voltages for natural systems. Comparisons have been made between H_2 and Eh measurements for redox status in a contaminated aquifer (Chapelle et al., 1996). Quantification of H_2 was more reliable than Eh measurements for identifying anoxic redox process, especially when considered with respect to electron acceptor availability. The limits and limitations of Eh measurements have been described for natural systems, especially with respect to how long the electrodes may be left in place in the soil or water (Mansfeldt, 1993; Norrström, 1994) and with respect to interpreting the Eh values obtained (Baas-Becking et al., 1960; Whitfield, 1974; Lindberg and Runnells, 1984; Yu, 1992). Platinum electrode systems also have been incorporated into potential-controlling systems for long-term laboratory studies (Patrick, 1966; Petrie et al., 1998).

14.2.3.2 Inadequacies of Pt Electrode Potentials

Assessing "electron activity" in soils relates strictly to an evaluation of the ability of the electron to be transferred, or do thermodynamic work, and not of its concentration or activity in soil solution, as can be defined for H^+. Because of the nature of the electron and its differences from H^+, a number of caveats must be described and recognized when evaluating Pt electrode potentials.

14.2.3.2.1 Dissolved Oxygen Status

While a stable potential can be obtained for a Pt-reference electrode pair immersed in an oxygenated soil suspension, this potential is unreliable as a measure of dissolved oxygen status (Bricker, 1982; Stumm and Morgan, 1996). The Pt surface may
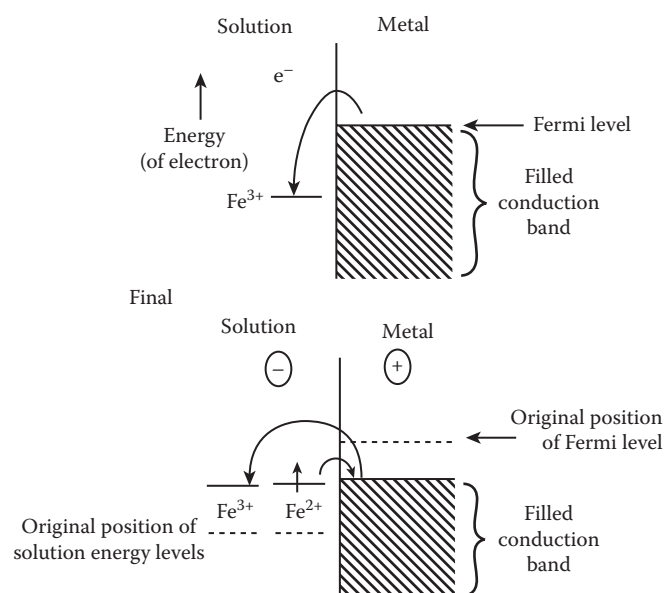
FIGURE 14.3 The energy of electrons in ions in solution and in the Pt electrode. (Taken from Compton, R.G., and G.H.W. Sanders. 1996. Electrode potentials. Oxford Science Publications, Oxford, U.K. p. 4, Fig. 1.2.)

react with O_2 to form $Pt(OH)_2$ that develops a potential with elemental Pt with a pe of 9.6 at pH 7 (Table 14.1). In addition, the measurement may not be that of the O_2–H_2O couple but may be responding to O_2 reduction intermediates, such as H_2O_2 and superoxide (Bricker, 1982). In addition, predicted pe values are relatively insensitive to changes in dissolved O_2 between 0.21 and 0.0021 atm (Table 14.2), the range of O_2 partial pressures in which aerobic respiration occurs (Russell, 1973). For these reasons, Pt electrode potentials cannot be used reliably as a measure of redox status for aerobic soils, but empirical values for pe may be obtained for comparison purposes (Bartlett, 1981, 1998).

While more faith is placed in measurements of soil pH, it also should be considered an empirical measurement because of uncertainty about the form of the H^+ ion in colloidal environments and about the behavior of the glass electrode in such systems. For these reasons, both pe and pH measured with electrodes in soils may be very uncertain for accurate descriptions of the redox status of soil environments containing air-filled pores.

14.2.3.2.2 Irreversibility of Redox Couples

Many of the important redox processes involving C, H, N, O, and S (the "light" elements, relative to the "heavy" metals) are irreversible in the thermodynamic sense, and nonelectroactive gases and molecules may be consumed or formed. As a result, potentials generated by redox couples for these elements are difficult to obtain and interpret using a Pt electrode. In addition, many of these reactions do not reach true chemical equilibrium, and activities measured in soil solution may be kinetically constrained (Liu and Narasimhan, 1989). Since soil redox status is often set by "microbial potentials," as discussed earlier, consuming or producing compounds or ions containing one or more of these elements may render Pt electrode measurements inaccurate.

14.2.3.2.3 Mixed Potentials

Measured redox potentials in soils normally are governed by more than one redox couple ("mixed potentials"), and these couples usually are not at true chemical equilibrium. Therefore, thermodynamic interpretations of measured Pt electrode potentials in soil suspensions need to consider several couples, and analyses of soil solution concentrations of redox active species may be indicative of the couples controlling the potential at the Pt surface (Typrin, 1998). For example, a Fe^{2+}–Fe^{3+} couple at activities greater than 10^{-5} M generates sufficient anodic and cathodic currents to obtain a measurable voltage for the system, and this may coexist with the relatively nonelectroactive couple, O_2–H_2O in heterogeneous soils.

The establishment of a measurable balance between the cathodic and anodic current at the Pt electrode at the point of zero applied voltage requires activities of $>10^{-5}$ M, a condition that may not exist for many redox active species of interest in soils. Activities $<10^{-6}$ M may be common, especially for certain plant nutrients and soil pollutants. For this reason, application of measured Pt electrode potentials to predicting soil composition may not be possible.

14.2.3.2.4 Coupling of pH and pe

Based on the complementary nature of pH and pe (Stumm and Morgan, 1996) summing pe + pH to describe log K for soils is possible theoretically, as described above. Due to the fact that the Pt electrode responds to pH (almost in Nernstian fashion) as well as to electron activity, pH should always be measured and reported with pe. The negative slopes of the pe–pH relationships (Figure 14.1) of many of the reduction half-reactions also indicate that the energy change associated with a particular reduction reaction decreases with increasing pH. Therefore, an Eh measurement cannot be used to predict the presence of a particular redox couple unless pH is known. Since higher pe values at lower pHs correspond to larger releases of free energy, reduction reactions are expected to be more likely to occur at lower pH. That is, such systems are ones in which "reduction is favored." In contrast, loss of electrons from reductants of a particular couple is favored at higher pH, or the system is more "prone to oxidation."

14.2.4 Proposed Alternative Strategies for More Accurate Measurement of Soil Redox Status

14.2.4.1 Using Electrochemical Relations in Reverse

Given the uncertainty associated with measured Pt electrode potentials in soils to quantify Eh or pe, actual measurements of reductant and oxidant activities, along with a reliable pH measurement, may be a better approach (Stumm and Morgan, 1996; Typrin, 1998). The activities are substituted into appropriate half-reactions relating pe and pH, and pe is thereby obtained by calculation.

As shown in Table 14.2, the pe predicted by such a technique will be inaccurate to different degrees for different half-reactions. For example, an error of two log units for H_2S or O_2 partial pressures will only produce errors of 0.3–0.5 pe units. In contrast, similar errors in measurement of Mn^{2+} or Fe^{2+} will result in pe errors of 1.0–3.0 pe units. Since Mn and Fe are relatively easy to measure accurately by atomic absorption or colorimetric methods and mineralogy of associated oxides can be made with infrared or x-ray techniques, assessing pe for Mn and Fe oxide-dominated systems could be reliable. Detailed research is needed to prove this hypothesis. In contrast, dissolved gases are harder to measure accurately, and qualitative estimates may be sufficient to obtain accurate evaluations of pe.

If calculated pe values obtained with this "reverse electrochemical technique" are equal for two different reduction half-reactions, then chemical equilibrium may be assumed to exist. If the pe values are unequal, then disequilibrium and a metastable, kinetically limited soil system probably exists. This latter condition is common in soils due to spatial heterogeneity of soil solution, oxide mineralogy, organic matter reactivity, and microbial controls on key reactions. New thinking and hands-on research are needed to provide new ideas for evaluation of the "electron activity" for soils by this combination of analytical and thermodynamic approaches. A key to its accuracy is knowing what

redox couples are contributing to the electron activity, assumed to exist as a quantifiable parameter at chemical equilibrium.

14.2.4.2 Redox Ranges for Empirical pe Values

These limitations to assessing soil pe based on Pt electrode and reverse electrochemical methods indicate that our sense of accuracy for soil redox status must be modified. If we surrender in our efforts to conceptualize and operationally define soil redox status, we will have lost a challenging scientific crusade. This means that while new ideas are being developed, we accept a lack of knowledge of pe values more accurate than ranges bracketed by whole numbers.

Liu and Narasimhan (1989) have described "redox zones" in which a range in Eh or pe defines an electron activity condition. The oxygen–nitrogen range is defined by Eh values of +250 to +100 mV, the iron range is +100 to 0.0 mV, the sulfate range is 0.0 to −200 mV, and that for methane–hydrogen is defined at < −200 mV. Sposito (1989) proposed "oxic" soils as those with pe > 7, "suboxic" ones in the range of pe between +2 and +7. and "anoxic" soils with pe < +2; all at pH 7. These ranges correspond roughly to redox control by oxygen–nitrogen, manganese–iron, and sulfur couples.

Berner (1981) proposed categories for redox named "oxic, postoxic, sulfidic, and methanic" controlled by transformations of oxygen/nitrogen, iron, sulfur, and methane–hydrogen, respectively. James (1989) has proposed these names be assigned to ranges in EMpe (empirical pe; Bartlett, 1981, 1998) of +7 to +13, +2 to +7, −2 to +2, and −6 to −2 (at pH 7). The appropriateness of these categories and names will require further evaluation of new operational definitions for the concept of "redox status" in soils.

Bartlett and James (1995) proposed a new system for categorizing soil redox status using chemical field tests such as tetramethylbenzidine for Mn status, Cr oxidation–reduction reactions, Fe speciation, sulfide levels, and pH. They proposed the following categories: superoxic, manoxic, suboxic, redoxic, anoxic, and sulfidic, which relate to electron lability in heterogeneous soil systems under field conditions. Bartlett (1998) has refined this system of redox classification and applied it to interfacial processes in soils pertinent to wetlands, hydric soils, and other soil–water systems containing contrasting zones of redox status.

14.2.4.3 Use of "Indicator of Reduction in Soils" (IRIS) Tubes

A novel method for assessing whether or not a soil may become reducing for Fe(III) (hydr)oxides to Fe(II) is to coat 21 mm, polyvinyl chloride (PVC) tubes with paints made from Fe(III) (hydr) oxide suspensions of known mineralogy. The coated tubes are placed in soil columns in the laboratory or in the field, and after a period of incubation, the extent of removal of Fe(III) from the tubes by reduction to Fe(II) is quantified by image processing (Rabenhorst et al., 2008 and references therein). Such in situ, empirical methods may prove useful in delineating wetlands, hydric soils, and the extent of reducing conditions for Fe(III) under various soils conditions.

14.3 Applications of Redox Methods and Concepts to Ecological, Engineered, and Agricultural Soil Systems

The principles and theories of oxidation–reduction processes in soils, and methods for defining them operationally, have many applications that relate to root–soil interactions in the rhizosphere, to pollutant oxidation state in and remediation of contaminated soils, to nutrient management and soil fertility in agricultural and wild ecosystems, to the role of soils in global C balances affecting the earth's climate, to the study of wetlands and hydric soils, and to understanding groundwater chemical composition at the interface with surface waters. A brief discussion of each of these applications follows, coupled to specific examples.

14.3.1 Rhizosphere Processes

Root exudates (rhizodeposition), especially those containing C compounds, are key constituents that influence microbial metabolism and chemical processes in the rhizosphere (Jones et al., 2009). These exudates may be the result of cell sloughing, establishment and maintenance of microsymbionts, gaseous loss (e.g., O_2 and CO_2), and mucilage production. Carbon compounds typically act as electron donors (and in the case of quinone functional groups as electron acceptors) for heterotrophic microbes, and HCO_3^- acts as the C source for autotrophs.

The diffusion of O_2 from the atmosphere through aerenchyma tissue formed in stems and roots in response to the development of anaerobic soil conditions favors the formation of oxidized rhizospheres in plants adapted to grow in anaerobic soils. Examples are rice paddies, freshwater and tidal wetlands, and other poorly drained soils (Marschner, 1995). In plants poorly adapted to growth in anaerobic soils, redox changes related to denitrification, reductive dissolution of Mn(III,IV) (hydr)oxides and Fe(III) (hydr)oxides, sulfate reduction, and methanogenesis can significantly affect plant growth, the maintenance of a symbiotic relationship with N_2-fixing bacteria and mycorrhizal fungi (Marschner, 1995).

The formation of the products of organic matter decomposition and fermentation of glucose to lactate or ethanol can affect rhizosphere chemistry and microbial activity. Formation of ethylene, volatile fatty acids, and phenolic compounds activity affects vascular tissue conductivity and root growth (Marschner, 1995). Ethylene production in root tissues in response to water logging is responsible for the formation of aerenchyma tissue in plants tolerant of anaerobic soil conditions.

The rhizosphere is a soil environment that is hard to study in situ but one that is critically important in plant nutrition, soil fertility, plant pathology, soil biochemistry, microbial ecology, and pedology. The August, 2009 issue of *Plant and Soil* comprises 21 articles dedicated to reviewing current knowledge of rhizosphere processes, many of which are redox based (Dessaux et al., 2009).

14.3.2 Soil Remediation and Pollutant Speciation

In recent years, much emphasis has been placed on the remediation of soils contaminated with organic or inorganic industrial waste products (Clapp et al., 2001). Examples of these contaminants are synthetic organic chemicals, such as dense nonaqueous phase liquids (DNAPLs); benzene, toluene, ethylbenzene, and xylenes (BTEX); and pesticides (herbicides, insecticides, and fungicides) (Schwarzenbach et al., 1993). The oxidation or reduction of functional groups, aliphatic chains, and aromatic rings results in changes in solubility and toxicity (Schwarzenbach et al., 1993; Clapp et al., 2001).

The oxidation state of contaminant metals and metalloids can be transformed through microbiological or chemical redox processes. Such redox-based remediation changes their speciation, solubility, bioavailability, and toxicity (Sauvé and Parker, 2005). In such cases, the total, unspeciated concentration of the element in question remains the same; only the oxidation state is changed. In these cases, the development of new methods for speciation and fractionation (Sauvé and Parker, 2005) based on oxidation state of the element has contributed to novel methods for remediation that do not involve removal of the element from the soil and that allow novel, in situ methods of remediation (James, 1996).

An example of an ongoing remediation-by-reduction are soils contaminated with chromate [Cr(VI)] from industrial plating operations and chromite ore processing residue (James, 1996). In situ or _ex situ_ conversion of toxic, soluble Cr(VI) to much less toxic and insoluble Cr(III) species has been effective in laboratory, pilot, and field-scale studies. Methods studied have included elemental Fe in permeable reactive barriers (PRBs; Rivero-Huguet and Marshall, 2009), steel wool (James, 1994), ascorbic acid (Xu et al., 2005), animal manures (Bartlett and Kimble, 1975), polysulfides (Graham et al., 2005), and $FeSO_4$ (Geelhoed et al., 2003). Central to the acceptance of these methods by government regulatory agencies has been the refinement of an analytical method for soils that selectively dissolves Cr(VI) from the solid matrix without reducing it and without method-induced oxidation of Cr(III) in the soil (Vitale et al., 1997).

14.3.3 Soil Fertility and Nutrient Cycling

Redox transformations are central to the biological controls of the N cycle, including N_2 fixation (N_2 to NH_4^+), nitrification (NH_4^+ to NO_3^-), denitrification (NO_3^- to N_2O and N_2). These processes govern the speciation of N in all ecosystems, whether in wild biomes, engineered soils, or domesticated agricultural ecosystems. Redox processes also govern speciation, bioavailability, and solubility of Mn, Fe, C and S, cofactors in many of the enzymes that catalyzed these N redox changes (Sylvia et al., 2005).

In soils and natural waters, high levels of P contribute to the growth of cyanobacteria and are therefore a concern related to eutrophication of freshwaters (Howarth, 1988). The oxidation state under ambient biological temperatures and pressures remains V+, but reduction of Fe(III) (hydr)oxides and the oxidation of Fe(II) control, in complex ways, the presence of an important sorbent for orthophosphate in soils ($H_2PO_4^-$ and HPO_4^{2-}, the dominant inorganic forms of P between pH 4 and 9; Young and Ross, 2001; Sims and Pierzynski, 2005).

14.3.4 Soil Organic C Dynamics and Climate Change

The oxidation state of C varies from IV– to IV+ in natural environments, and enzymatic processes govern transformations among them. Examples of the classes of C compounds representing the range of oxidation states are alkanes (IV– in methane, CH_4), alcohols (II– in methanol, CH_3OH), aldehydes (0 in formaldehyde or general representations of SOM, CH_2O), carboxylic acids (II+ in formic acid, HCOOH, i.e., hydrated CO), and carbon dioxide (IV+ in CO_2; and when hydrated as H_2CO_3, HCO_3^-, and CO_3^{2-}, depending on pH). The speciation of C in soils is governed by myriad anabolic and catabolic processes in cells (Clapp et al., 2005; Table 14.3).

The knowledge of C speciation and the reactions governing it are central to many contemporary environmental, engineering, and agricultural concerns. Greenhouse gas and particulate releases into the atmosphere (CO_2, CH_4, black C and other C-rich particulates) has encouraged research on C sequestration related to balances between methanogenesis and methanotrophy (Sylvia et al., 2005). Concern about climate change due to greenhouse gas releases is linked to the kinetics of formation and degradation of SOM and to effects of tillage practices and plant and animal waste decomposition dynamics (Ussiri and Lal, 2009). Balances between sources of SOM (via photosynthesis and oxidation polymerization processes) and its decomposition to simple organic acids and CO_2 are keys to how soils may sequester organic C.

Clapp et al. (2005) provide an encyclopedic treatise on SOM, its structure, reactions, and roles in soils as a base for understanding organic matter and C dynamics. Lal (1999) provides a brief description, with examples, of the role of soils in the greenhouse effect. He includes relevant information on greenhouse emissions, with particular reference to rice paddies and organic soils. Falkowski et al. (2008) provide excellent perspective on the microbial redox processes that govern nonequilibrium biological systems, using what they call "nanobiological machines." They also describe how redox processes integrate sediments, water, and the atmosphere through interconnected, biologically mediated cycles for H, C, N, O, S, and Fe.

14.3.5 Wetland Delineation and Function

The delineation and redox-based functioning of freshwater and tidal wetlands (and their constituent hydric soils) are current challenges as a result of increasing interest in preventing loss of wetlands due to human development and sea level rise linked to global climate change and crustal subsidence (Hurt and Vasilas, 2006).

TABLE 14.3 Reduction Half-Reactions for Environmentally Relevant Species

Species		Equations	E^o (V)	Log K	ΔG_r^o (kJ/eq)
A	γ-MnOOH	γ-MnOOH + e^- + 3H^+ → Mn^{2+} + 2H_2O	1.50	25.4	−145.0
B	CrO_4^{2-}	1/3CrO_4^{2-} + e^- + 5/3H^+ → 1/3$Cr(OH)_3$ + 1/3H_2O	1.24	21.0	−119.9
C	MnO_2	1/2MnO_2 + e^- + 2H^+ → 1/2Mn^{2+} + H_2O	1.23	20.8	−118.8
D	$HCrO_4^-$	1/3$HCrO_4^-$ + e^- + 4/3H^+ → 1/3$Cr(OH)_3$ + 1/3H_2O	1.11	18.9	−107.9
E	AQDS	1/2AQDS + e^- + H^+ → 1/2AH_2DS	0.22	3.9	−22.3
F	Lactate	½Pyruvate + e^- + H^+ → 1/2lactate	0.23	3.9	−22.3
G	CH_2O	1/4CO_2 + e^- + H^+ → 1/4CH_2O + 1/4H_2O	−0.07	−1.2	6.9
Examples					
γ-MnOOH reduction by AH_2DS (A and E)				21.5	−122.8
$HCrO_4^-$ reduction by CH_2O (D and G)				20.1	−114.8
$HCrO_4^-$ reduction by lactate (D and F)				15.0	−85.7
$HCrO_4^-$ reduction by AH_2DS (D and E)				15.0	−85.7
$Cr(OH)_3$ oxidation by γ-MnOOH (A and D)				6.5	−37.1
AQDS reduction by CH_2O (E and G)				5.1	−29.1
AQDS reduction by lactate (E and F)				0.0	0.0

Sources: Data from Bartlett, R.J., and B.R. James. 2000. Redox phenomena, p. 169–193. *In* M.E. Sumner (ed.) Handbook of soil science, 1st Ed. CRC Press, Washington, DC; Milazzo, G., V.K. Sharma, and S. Caroli. 1978. Tables of standard electrode potentials. Wiley, New York; Loach, P.A. 1976. Oxidation–reduction potentials, absorbency bands, and molar absorbency of compounds used in biochemical studies, p. 122–130. *In* G. Fasman (ed.) Handbook of biochemistry and molecular biology. Chemical Rubber Co., Cleveland, OH.

These reactions can be combined to form redox reactions, many of which are energetically favorable as indicated by negative G_r values in the given examples.

The definition of hydric soils (Hurt and Vasilas, 2006) and wetlands are linked to redox conditions and dynamics associated with the position and fluctuations of the water table. A so-called technical standard for defining wetlands is based on Eh and pH measurements associated with the oxidation and reduction of Fe(II,III) (Rabenhorst and Castenson, 2005; Hurt and Vasilas, 2006; NTCHS, 2007). Recent research on IRIS tubes and the concept of "biological zero" have contributed new methods and conceptual thinking on this technical standard and how to define a wetland (Rabenhorst, 2005; Rabenhorst et al., 2008).

The chemistry of tidal marshes that are diurnally inundated with SO_4^{2-}-rich seawater leads to the accumulation of numerous iron sulfide minerals, including iron monosulfides (e.g., FeS) and pyrite (FeS_2) (Rabenhorst and James, 1992). In addition, the dynamics of organic C accumulation, methanaogenesis and methane oxidation are important applications of redox concepts because of the importance of CH_4 as a greenhouse gas (Liverman, 2007).

14.3.6 Riparian Soil-Vegetation Systems and Groundwater Nitrate Concentrations

Redox transformations of NO_3^- in groundwater have been major societal concerns due to the potential for eutrophication of natural waters, especially of by non-N_2 fixing, green algae in saline waters of estuaries, such as the Chesapeake Bay of the Coastal Plain of the United States (James et al., 1991). The reduction of N in NO_3^- (oxidation state V+) to N_2O (oxidation state I+) and N_2 (oxidation state 0) via denitrification in riparian ecosystems has been extensively studied (Lowrance et al., 1984; Peterjohn and Correll, 1984; Allan et al., 2008). The role of trees, understory vegetation, and planted grasses in riparian ecosystems has shown that black locust trees (*Pseudoacacia* spp.) act as a source of NO_3^- in groundwater that occurs following N_2 fixation, NH_4^+ release into the soil, and nitrification. When black locust trees were cut and replaced with planted tall fescue, NO_3^- concentrations in groundwater beneath the riparian zones decreased, whereas when non-leguminous trees (red oak, loblolly pine, and sassafras) were cut, NO_3^- concentrations increased. These results demonstrated the important roles played by vegetation–microbe–soil–groundwater interactions affecting N speciation (James et al., 1991).

14.4 Earlier Reviews and Prescient Work on Oxidation–Reduction Processes in Soils

There are numerous general reviews and examples of redox measurements, processes, applications, and data in soils and natural waters. Interested readers are referred to the following sources: Lindsay (1979), Buxton et al. (1988), Neta et al. (1988), Wardman (1989), Sawyer (1991), Bartlett and James (1993), Schwarzenbach et al. (1993), Helz et al. (1994), Blough and Zepp (1995), Compton and Sanders (1996), Stumm and Morgan (1996), Bartlett (1998).

Bartlett and Ross (2005) provide novel and heuristic ideas surrounding a central role played by Mn in soil redox processes, including formation of humic compounds, free radical reactions, nutrient cycling, rhizosphere chemistry, photochemistry, and wetlands. Other chapters in this book tie in soil acidity, chemical kinetics, and other topics related to oxidation–reduction.

Extensive redox-based research conducted by the late William Patrick and coworkers at the Wetland Biogeochemistry Institute at Louisiana State University (Baton Rouge, LA) has been central to much of current knowledge related to Eh and pH as master variable controlling the chemistry and biochemistry of tidal marshes and freshwater wetlands.

Remarkable, almost prescient, insight is provided in several references on electron activity and its measurement by the Pt electrode in the first four decades of the twentieth century following the discovery of the electron in 1897 (Gillespie, 1920; Willis, 1932; Pearsall and Mortimer, 1939). These researchers linked thermodynamic theory, biological processes in soils, and applications of redox chemistry to societal needs of their times—an accomplishment that can serve as a guide for future studies on the elusive electron in soil environments. Research and teaching related to oxidation–reduction processes in myriad soil environments remains a vital and dynamic field of study with many applications for future soil chemists and environmental scientists interested in addressing complex biological and chemical systems.

References

Allan, C.J., P. Vidon, and R. Lowrance. 2008. Frontiers in riparian zone research in the 21st century. Hydrol. Process. 22:3221–3222.

Allison, J.D., and D.S. Brown. 1995. MINTEQA2/PRODEFA2—A geochemical speciation model and interactive processor, p. 241–252. *In* R.H. Loeppert et al. (eds.) Chemical equilibria and reaction models. Soil Science Society of America Special Publication. SSSA, Madison, WI.

Amacher, M.C., and D.E. Baker. 1981. Redox reactions involving chromium, plutonium, and manganese in soils. Final Rep. Div. Energy Tech. US Dep. Energy DOE/DP/0415-1. Washington, DC.

Arndt, D. 1981. Manganese compounds as oxidizing agents in organic chemistry. Open Court Publishing Co., La Salle, IL.

Asatiani, N.V., M.K. Abuladze, T.M. Kartvelishvili, N.G. Bakradze, N.A. Sapojnikova, N.Y. Tsibakhashvili, L.V. Tabatadze, L.V. Lejava, L.L. Asanishvili, and H.Y. Holman. 2004. Effect of chromium(VI) action on *Arthrobacter oxydans*. Curr. Microbiol. 49:321–326.

Baas-Becking, L.G.M., I.R. Kaplan, and D. Moore. 1960. Limits of the natural environment in terms of pH and oxidation–reduction potentials. J. Geol. 68:243–284.

Banerjee, R. 2008. Redox metabolism and life, p. 1–10. *In* R. Banerjee et al. (eds.) Redox biochemistry. John Wiley & Sons Inc., Hoboken, NJ.

Barcelona, M.J., and T.R. Holm. 1991. Oxidation–reduction capacities of aquifer solids. Environ. Sci. Technol. 25:1565–1572.

Bartlett, R.J. 1981. Oxidation–reduction status of aerobic soils, p. 77–102. *In* D. Baker (ed.) Chemistry in soil environments. SSSA, Madison, WI.

Bartlett, R.J. 1998. Characterizing soil redox behavior, p. 371–397. *In* D.L. Sparks (ed.) Soil physical chemistry, 2nd Ed. CRC Press, Boca Raton, FL.

Bartlett, R., and B. James. 1979. Behavior of chromium in soils. III. Oxidation. J. Environ. Qual. 8:31–35.

Bartlett, R.J., and B.R. James. 1993. Redox chemistry of soils. Adv. Agron. 50:151–208.

Bartlett, R.J., and B.R. James. 1995. System for categorizing soil redox status by chemical field testing. Geoderma 68:211–218.

Bartlett, R.J., and B.R. James. 2000. Redox phenomena, p. 169–193. *In* M.E. Sumner (ed.) Handbook of soil science, 1st Ed. CRC Press, Washington, DC.

Bartlett, R.J., and J.M. Kimble. 1975. Behavior of chromium in soils. II. Hexavalent forms. J. Environ. Qual. 5:383–386.

Bartlett, R.J., and D.S. Ross. 2005. Chemistry of redox processes in soils, p. 461–487. *In* M.A. Tabatabai and D.L. Sparks (eds.) Chemical processes in soils. Soil Sci. Soc. Am. Book Ser. 8. SSSA, Madison, WI.

Bates, R.G. 1981. The modern meaning of pH. Crit. Rev. Anal. Chem. 10:247–278.

Bauer, M., T. Heitmann, D.L. Macalady, and C. Blodau. 2007. Electron transfer capacities and reaction kinetics of peat dissolved organic matter. Environ. Sci. Technol. 41:139–145.

Becker, J.G. 2006. A modeling study and implications of competition between *Dehalococcoides ethenogenes* and other tetrachloroethene-respiring bacteria. Environ. Sci. Technol. 40:4473–4480.

Berner, R.A. 1981. A new geochemical classification of sedimentary environments. J. Sediment. Petrol. 51:359–365.

Blaylock, M.J., and B.R. James. 1994. Redox transformations and plant uptake of selenium resulting from root–soil interactions. Plant Soil 158:1–12.

Blough, N.V., and R.G. Zepp. 1995. Reactive oxygen species in natural waters, p. 280–332. *In* C.S. Foote et al. (eds.) Active oxygen in chemistry. Chapman & Hall, New York.

Bond, D.R., and D.R. Lovley. 2002. Reduction of Fe(III) oxide by methanogens in the presence and absence of extracellular quinones. Environ. Microbiol. 4:115–124.

Bricker, O.P. 1982. Redox measurement: Its measurement and importance in water systems. Water Analysis, Vol. 1. Academic Press, Orlando, FL.

Brose, D.A. and B.R. James, 2010. Oxidation–reduction transformations of chromium in aerobic soils: Role of electron-shuttling quinones. *Environ. Sci. Technol*. 44:9438–9444.

Buresh, R.J., M.E. Casselman, and W.H. Patrick, Jr. 1980. Nitrogen fixation in flooded soil systems: A review. Adv. Agron. 33:149–192.

Bürge, I., and S. Hug. 1997. Kinetics and pH dependence of chromium(VI) reduction by iron(II). Environ. Sci. Technol. 31:1426–1432.

Buxton, G.V., C.L. Greenstock, W.P. Helman, and A.B. Ross. 1988. Critical review of rate constants for reactions of hydrated electrons, hydrogen atoms, and hydroxyl radicals ($^{\bullet}OH/^{\bullet}OH^-$) in aqueous solution. J. Phys. Chem. Ref. Data 17:513–886.

Castellan, G.W. 1983. Physical chemistry, 3rd Ed. Addison-Wesley Publishing Co., Reading, MA.

Cervantes, F.J., S. van der Velde, G. Lettinga, and J.A. Field. 2000. Quinones as terminal electron acceptors for anaerobic microbial oxidation of phenolic compounds. Biodegradation 11:313–321.

Chapelle, F.H., S.K. Haack, P. Adriaens, M.A. Henry, and P.M. Bradley. 1996. Comparison of Eh and H_2 measurements for delineating redox processes in a contaminated aquifer. Environ. Sci. Technol. 30:3565–3569.

Clapp, C.E. et al. (eds.). 2001. Humic substances and chemical contaminants. SSSA, Madison, WI.

Clapp, C.E., M.H.B. Hayes, A.J. Simpson, and W.L. Kingery. 2005. Chemistry of soil organic matter, p. 1–150. *In* M.A. Tabatabai and D.L. Sparks (eds.) Chemical processes in soils. SSSA Book Series 8. SSSA, Madison, WI.

Coleman, M.L., D.B. Hedrick, D.R. Lovley, D.C. White, and K. Pyle. 1993. Reduction of Fe(III) in sediments by sulphate-reducing bacteria. Nature 361:436–438.

Compton, R.G., and G.H.W. Sanders. 1996. Electrode potentials. Oxford Science Publications, Oxford, U.K.

Cotton, F.A., and G. Wilkinson. 1980. Advanced inorganic chemistry. John Wiley & Sons, New York.

Dessaux, Y., P. Hinsinger, and P. Lemanceau. 2009. Rhizosphere: So many achievements and even more challenges. Plant Soil 321:1.

Dittrich, M., and A. Luttge. 2008. Microorganisms, mineral surfaces, and aquatic environments: Learning from the past for future progress. Geobiology 6:201–213.

Falkowski, P.G., T. Fenchel, and E.F. DeLong. 2008. The microbial engines that drive earth's biogeochemical cycles. Science 320:1034–1039.

Farrell, R.E., G.D.W. Swerhone, and C. van Kessel. 1991. Construction and evaluation of a reference electrode assembly for use in monitoring in situ soil redox potentials. Comm. Soil Sci. Plant Anal. 22:1059–1068.

Field, J.A., F.J. Cervantes, F.P. van der Zee, and G. Lettinga. 2000. Role of quinones in the biodegradation of priority pollutants: A review. Water Sci. Technol. 42:215–222.

Fredrickson, J.K., and J.M. Zachara. 2008. Electron transfer at the microbe-mineral interface: A grand challenge in biogeochemistry. Geobiology 6:245–253.

Fredrickson, J.K., J.M. Zachara, D.W. Kennedy, H. Dong, T.C. Onstott, N.W. Hinman, and S. Li. 1998. Biogenic iron mineralization accompanying the dissimilatory reduction of hydrous ferric oxide by a groundwater bacterium. Geochim. Cosmochim. Acta 62:3239–3257.

Fridovich, I. 1978. The biology of oxygen radicals. Science 201:875–880.

Garrels, R.M., and C.L. Christ. 1965. Solutions, minerals, and equilibria. Freeman, Cooper, & Co., San Francisco, CA.

Geelhoed, J.S., J.C.L. Meeussen, M.J. Roe, S. Hillier, R.P. Thomas, J.G. Farmer, and E. Paterson. 2003. Chromium remediation or release? Effect of iron(II) sulfate addition on chromium(VI) leaching from columns of chromite ore processing residue. Environ. Sci. Technol. 37:3206–3213.

Gillespie, L.J. 1920. Reduction potentials of bacterial cultures and of water-logged soils. Soil Sci. 9:199–216.

Goldhammer, T., and C. Blodau. 2008. Desiccation and product accumulation constrain heterotrophic anaerobic respiration in peats of an ombrotrophic temperate bog. Soil Biol. Biochem. 40:2007–2015.

Graham, M.C., J.G. Farmer, P. Anderson, E. Paterson, S. Hillier, D.G. Lumsdon, and R.J.F. Bewley. 2005. Calcium polysulfide remediation of hexavalent chromium contamination from chromite ore processing residue. Sci. Total Environ. 364:32–44.

Gralnick, J.A., and D.K. Newman. 2007. Extracellular respiration. Mol. Microbiol. 65:1–11.

Grenthe, I., W. Stumm, M. Laaksuharju, A.C. Nilsson, and P. Wikberg. 1992. Redox potentials and redox reactions in deep groundwater systems. Chem. Geol. 98:131–150.

Grundl, T. 1994. A review of the current understanding of redox capacity in natural, disequilibrium systems. Chemosphere 28:613–626.

Gu, B.H., and J. Chen. 2003. Enhanced microbial reduction of Cr(VI) and U(VI) by different natural organic matter fractions. Geochim. Cosmochim. Acta Suppl. 67:3575–3582.

Harter, R.D., and G. Smith. 1981. Langmuir equation and alternate methods for studying "adsorption" reactions in soils, p. 167–182. *In* D. Baker (ed.) Chemistry in soil environments. SSSA, Madison, WI.

Helz, G.R., R.G. Zepp, and D.G. Crosby (eds.). 1994. Aquatic and surface photochemistry. Lewis Publishers, Boca Raton, FL.

Heron, G., T.H. Christensen, and J.C. Tjell. 1994. Oxidation capacity of aquifer sediments. Environ. Sci. Technol. 28:153–158.

Hobman, J.L., J.R. Wilson, and N.L. Brown. 2000. Microbial mercury reduction. *In* D.R. Lovley (ed.) Environmental microbe–metal interactions. ASM Press, Washington, DC.

Højberg, O., N.P. Revsbech, and J.M. Tiedje. 1994. Denitrification in soil aggregates analyzed with microsensors for nitrous oxide and oxygen. Soil Sci. Soc. Am. J. 58:1691–1698.

Hostettler, J.D. 1992. The physical basis of Eh related to Eh measurements in natural waters. Preprint extended abstract. American Chemical Society Meetings, San Francisco, CA.

Howarth, R.W. 1988. Nutrient limitation of net primary production in marine ecosystems. Annu. Rev. Ecol. 19:89–110.

Hug, S.J., B.R. James, and H.U. Laubscher. 1997. Iron(III) catalyzed photochemical reduction of chromium(VI) by oxalate and citrate in aqueous solutions. Environ. Sci. Technol. 31:160–170.

Hurt, G.W., and L.M. Vasilas. 2006. Field indicators of hydric soils in the United States: A guide for identifying and delineating hydric soils, Version 6.0. USDA–NRCS, Washington, DC.

James, B.R. 1989. Electron activity in soils: A key master variable, p. 201. Agronomy abstract. ASA, Madison, WI.

James, B.R. 1994. Hexavalent chromium solubility and reduction in alkaline soils enriched with chromite ore processing residue. J. Environ. Qual. 23:227–233.

James, B.R. 1996. The challenge of remediating chromium-contaminated soil. Environ. Sci. Technol. 30:248–251.

James, B.R., B.B. Bagley, and P.H. Gallagher. 1991. Riparian zone vegetation effects on concentrations of nitrate in shallow groundwater, p. 605–611. *In* J.A. Mihursky and A. Chaney (eds.) New perspectives in the Chesapeake system: A research and management partnership. Proc. Conf. Chesapeake Res. Consort. Publ. 137. December 4–6, 1990. Baltimore, MD.

James, B.R., and R.J. Bartlett. 1983. Behavior of chromium in soils. VI. Interactions between oxidation–reduction and organic complexation. J. Environ. Qual. 12:173–176.

Jiang, J., and A. Kappler. 2008. Kinetics of microbial and chemical reduction of humic substances: Implications for electron shuttling. Environ. Sci. Technol. 42:3563–3569.

Jones, D.L, C. Nguyen, and R.D. Finlay. 2009. Carbon flow in the rhizosphere: Carbon trading at the soil-root interface. Plant Soil 321:4–33.

Jorgensen, C.J., O.S. Jacobsen, B. Elberling, and J. Aamand. 2009. Microbial oxidation of pyrite coupled to nitrate reduction in anoxic groundwater sediment. Environ. Sci. Technol. 43:4851–4857.

Kappler, A., M. Benz, B. Schink, and A. Brune. 2004. Electron shuttling via humic acids in microbial iron(III) reduction in a freshwater sediment. FEMS Microbiol. Ecol. 47:85–92.

Kludze, H.K., R.D. DeLaune, and W.H. Patrick, Jr. 1994. A colorimetric method for assaying dissolved oxygen loss from container-grown rice roots. Soil Sci. Soc. Am. J. 86:483–487.

Lal, R. 1999. Soil processes and greenhouse effect, p. 199–212. *In* R.L. Lal et al. (eds.) Methods for assessment of soil degradation: Advances in soil science. CRC Press, Boca Raton, FL.

Larson, R.A. 1997. Naturally occurring antioxidants. Lewis Publishers, Boca Raton, FL.

Leffler, J.E. 1993. An introduction to free radicals. John Wiley & Sons, New York.

Lindberg, R.D., and D.D. Runnells. 1984. Ground water redox reactions: An analysis of equilibrium state applied to Eh measurements and geochemical modeling. Science 225:925–927.

Lindsay, W.L. 1979. Chemical equilibria in soils. Wiley-Interscience, New York.

Liu, C.W., and T.N. Narasimhan. 1989. Redox-controlled multiple-species reactive chemical transport. 1. Model development. Water Resour. Res. 25:869–882.

Liverman, D. 2007. From uncertain to unequivocal—The IPCC Working Group I Report: Climate change 2007—The physical science basis. Environment 49:28–32.

Loach, P.A. 1976. Oxidation–reduction potentials, absorbency bands, and molar absorbency of compounds used in biochemical studies, p. 122–130. *In* G. Fasman (ed.) Handbook of biochemistry and molecular biology. Chemical Rubber Co., Cleveland, OH.

Lovley, D.R. 2000. Fe(III) and Mn(IV) reduction, p. 3–30. *In* D.R. Lovley (ed.) Environmental microbe–metal interactions. ASM Press, Washington, DC.

Lovley, D.R., F.H. Chappelle, and J.C. Woodward. 1994. Use of dissolved H_2 concentrations to determine distribution of microbially catalyzed redox reactions in anoxic groundwater. Environ. Sci. Technol. 28:1205–1210.

Lovley, D.R., J.D. Coates, E.L. Blunt-Harris, E.J.P. Phillips, and J.C. Woodward. 1996. Humic substances as electron acceptors for microbial respiration. Nature 382:445–448.

Lowrance, R.R., R.L. Todd, and L.E. Asmussen. 1984. Nutrient cycling in an agricultural watershed I. Phreatic movement. J. Environ. Qual. 13:22–27.

Madigan, M.T., and J.M. Martinko. 2006. Brock biology of microorganisms, 11th Ed. Pearson Prentice Hall, Upper Saddle River, NJ.

Mansfeldt, T. 1993. Redoxpotentialmessungen mit dauerhaft installierten Platinelektroden unter reduzierenden Bedingungen. J. Plant Nutr. Soil Sci. 156:287–292.

Marschner, H. 1995. Mineral nutrition of higher plants, 2nd Ed. Academic Press, London, U.K.

Matia, L., G. Rauret, and R. Rubio. 1991. Redox potential measurement in natural waters. Fresen. J. Anal. Chem. 339:455–462.

Mendelssohn, I.A., B.A. Kleiss, and J.S. Wakeley. 1995. Factors controlling the formation of oxidized root channels: A review. Wetlands 15:37–46.

Milazzo, G., V.K. Sharma, and S. Caroli. 1978. Tables of standard electrode potentials. Wiley, New York.

Mueller, S.C., L.H. Stolzy, and G.W. Fick. 1985. Constructing and screening platinum microelectrodes for measuring soil redox potential. Soil Sci. 139:558–560.

Moore, J.N., J.R. Walker, and T.H. Hayes. 1990. Reaction scheme for the oxidation of As(III) to As(V) by birnessite. Clay. Clay Miner. 38:549–555.

Neta, P., R.E. Huie, and A.B. Ross. 1988. Rate constants for reactions of inorganic radicals in aqueous solution. J. Phys. Chem. Ref. Data 17:1027–1284.

Norrström, A.C. 1994. Field-measured redox potentials in soils at the groundwater-surface-water interface. Eur. J. Soil Sci. 45:31–36.

NTCHS. 2007. The hydric soil technical standard. Deliberations of the National Tech. Comm. Hydric Soils. Available online at http://soils.usds.gov/use/hydric/ntchs/tech_notes/index.html (October 15, 2009).

Nyman, J., M. Gentile, and C. Criddle. 2007. Sulfate requirement for the growth of U(VI)-reducing bacteria in an ethanol-fed enrichment. Bioremediation J. 11:21–32.

Oremland, R.S., and J. Stolz. 2000. Dissimilatory reduction of selenate and arsenate in nature, p. 199–224. *In* D.R. Lovley (ed.) Environmental microbe–metal interactions. ASM Press, Washington, DC.

Patrick Jr., W.H. 1966. Apparatus for controlling the oxidation–reduction potential of waterlogged soils. Nature 212:1278–1279.

Patrick Jr., W.H., and R.D. DeLaune. 1972. Characterization of the oxidized and reduced zones in flooded soils. Soil Sci. Soc. Am. Proc. 36:573–576.

Patrick, W.H., R.P. Gambrell, and S.P. Faulkner. 1996. Redox measurements of soils, p. 1255–1273. *In* D.L. Sparks (ed.) Methods of soil analysis. Part 3. SSSA, Madison, WI.

Paul, E. (ed.) 2007. Soil microbiology, ecology, and biochemistry, 3rd Ed. Elsevier, Amsterdam, the Netherlands.

Pearsall, W.H., and C.H. Mortimer. 1939. Oxidation–reduction potentials in waterlogged soils, natural waters, and muds. J. Ecol. 27:483–501.

Peiffer, S., O. Klemm, K. Pecher, and R. Hollerung. 1992. Redox measurements in aqueous solutions: A theoretical approach to data interpretation, based on electrode kinetics. J. Contam. Hydrol. 10:1–18.

Peretyazhko, T., and G. Sposito. 2006. Reducing capacity of terrestrial humic acids. Geoderma 137:140–146.

Peterjohn, W.T., and D.L. Correll. 1984. Nutrient dynamics in an agricultural watershed: Observations on the role of a riparian forest. Ecology 65:1466–1475.

Petrie, R.A., P.R. Grossl, and R.C. Sims. 1998. Controlled environment potentiostat to study solid-aqueous systems. Soil Sci. Soc. Am. J. 62:379–382.

Ponnamperuma, F.N. 1972. The chemistry of submerged soils. Adv. Agron. 24:29–96.

Postma, D., and R. Jakobsen. 1996. Redox zonation: Equilibrium constraints on the Fe(III)/SO_4^{2-} reduction interface. Geochim. Cosmochim. Acta 60:3169–3175.

Rabenhorst, M.C. 2005. Biologic zero: A soil temperature concept. Wetlands 25:616–621.

Rabenhorst, M.C., R.R. Bourgault, and B.R. James. 2008. Iron oxyhydroxide reduction in simulated wetland soils: Effects of mineralogical composition of IRIS paints. Soil Sci. Soc. Am. J. 72:1838–1842.

Rabenhorst, M.C., and K.L. Castenson. 2005. Temperature effects on iron reduction in a hydric soil. Soil Sci. 170:734–742.

Rabenhorst, M.C., W.D. Hively, and B.R. James. 2009. Measurements of soil redox potential. Soil Sci. Soc. Am. J. 73:668–674.

Rabenhorst, M.C., and B.R. James. 1992. Iron sulfidization in tidal marsh soils, p. 203–217. *In* H.C.W. Skinner and R.W. Fitzpatrick (eds.) Biomineralization processes of iron and manganese. Catena Supplement 21. Catena Verlag, Destedt, Germany.

Ratasuk, N., and M.A. Nanny. 2007. Characterization and quantification of reversible redox sites in humic substances. Environ. Sci. Technol. 41:7844–7850.

Rawlings, D.E. 2005. Characteristics and adaptablity of iron- and sulfur-oxidizing microorganisms used for the recovery of metals from minerals and their concentrates. Microb. Cell Fact. 4:13.

Reguera, G., K.D. McCarthy, T. Mehta, J.S. Nicoll, M.T. Tuominen, and D.R. Lovley. 2005. Extracellular electron transfer via microbial nanowires. Nature 435:1098–1101.

Rowell, D.L. 1981. Oxidation and reduction, p. 401–461. *In* D.J. Greenland and M.H.B. Hayes (eds.) The chemistry of soil processes. John Wiley & Sons, Cleveland, OH.

Rivero-Huguet, M., and W.D. Marshall. 2009. Influence of various organic molecules on the reduction of hexavalent chromium mediated by zero valent iron. Chemosphere 76:1240–1248.

Russell, E.W. 1973. Soil conditions and plant growth, 10th Ed. Longman, London, U.K.

Saggar, S., K.R. Tate, D.L. Giltrap, and J. Singh. 2008. Soil-atmosphere exchange of nitrous oxide and methane in New Zealand terrestrial ecosystems and their mitigation options: A review. Plant Soil 309:25–42.

Salminen, J.M., P.J. Hanninen, J. Leveinen, P.T.J. Lintinen, and K.S. Jorgensen. 2006. Occurrence and rates of terminal electron-accepting processes and recharge processes in petroleum hydrocarbon-contaminated subsurface. J. Environ. Qual. 35:2273–2282.

Sauvé, S., and D.R. Parker. 2005. Chemical speciation of trace elements in soil solution, p. 655–688. *In* M.S. Tabatabai and D.L. Sparks (eds.) Chemical processes in soils. SSSA Book Series 8. SSSA, Madison, WI.

Sawyer, D.T. 1991. Oxygen chemistry. Oxford University Press, New York.

Scerri, E.R. 1997. The periodic table and the electron. Am. Sci. 85:546–553.

Schwarzenbach, R.P., P.M. Gschwend, and D.M. Imboden. 1993. Environmental organic chemistry. John Wiley & Sons, New York.

Scott, D.T., D.M. McKnight, E.L. Blunt-Harris, S.E. Kolesar, and D.R. Lovley. 1998. Quinone moieties act as electron acceptors in the reduction of humic substances by humics-reducing microorganisms. Environ. Sci. Technol. 32:2984–2989.

Sillén, L.G. 1967. Master variables and activity scales, p. 45–56. *In* W. Stumm (ed.) Equilibrium concepts in natural water systems. Advances in Chemistry Series 67. Am. Chem. Soc., Washington, DC.

Sims, J.T., and G.M. Pierzynski. 2005. Chemistry of phosphorus in soils, p. 151–192. *In* M.A. Tabatabai and D.L. Sparks (eds.) Chemical processes in soils. SSSA Book Series 8. SSSA, Madison, WI.

Sparks, D.L. 1985. Kinetics of ionic reactions in clay minerals and soils. Adv. Agron. 38:231–265.

Sparks, D.L. 2003. Environmental soil chemistry, 2nd Ed. Academic Press, Amsterdam, the Netherlands.

Sposito, G. 1981. The thermodynamics of soil solutions. Oxford University Press, New York.

Sposito, G. 1989. The chemistry of soils. Oxford University Press, New York.

Stone, A.T., and J.J. Morgan. 1987. Reductive dissolution of metal oxides, p. 221–254. *In* W. Stumm (ed.) Aquatic surface chemistry. John Wiley & Sons, New York.

Stoodley, L.H., J.W. Costerton, and P. Stoodley. 2004. Bacterial biofilms: From the natural environment to infectious diseases. Nat. Rev. Microbiol. 2:95–108.

Straub, K.L., and B. Schink. 2003. Evaluation of electron-shuttling compounds in microbial ferric iron reduction. FEMS Microbiol. Lett. 220:229–233.

Stumm, W. 1992. Chemistry of the solid–water interface. Wiley-Interscience, New York.

Stumm, W. 1993. Aquatic colloids as chemical reactants: Surface structure and reactivity. Colloids Surf. A 73:1–18.

Stumm, W., and J.J. Morgan. 1996. Aquatic chemistry, 3rd Ed. Wiley-Interscience, New York.

Stumm, W., and B. Sulzberger. 1992. The cycling of iron in natural environments: Considerations based on laboratory studies of heterogeneous redox processes. Geochim. Cosmochim. Acta 56:3233–3258.

Sullivan, J.C., S. Gordan, D. Cohen, W. Mulac, and K.H. Schmidt. 1976. Pulse radiolysis studies of uranium(VI), neptunium(VI), neptunium(V), and plutonium(VI) in aqueous perchlorate media. J. Phys. Chem. 8:1684–1686.

Suter, D., S. Banwart, and W. Stumm. 1991. Dissolution of hydrous iron(III) oxides by reductive mechanism. Langmuir 7:809–813.

Sylvia, D.M., P.G. Hartel, J.J. Fuhrmann, and D.A. Zuberer. 2005. Principles and application of soil microbiology, 2nd Ed. Pearson Prentice Hall, Upper Saddle River, NJ.

Tan, K.H. 2003. Humic matter in soil and the environment: Principles and controversies. Marcel Dekker, Inc., New York.

Thompson, J.J. 1923. The electron in chemistry. Franklin Institute Press, Philadelphia, PA.

Tributsch, H., and L. Pohlmann. 1998. Electron transfer: Classical approaches and new frontiers. Science 279:1891–1895.

Typrin, L.R. 1998. Using a thermodynamically-based approach to assess the redox status and predict the valence state of chromium in simple, aqueous systems and chromium-enriched soils. MS thesis. University of Maryland. College Park, MD.

Ussiri, D.A.N., and R. Lal. 2009. Long-term tillage effects on soil carbon storage and carbon dioxide emissions in continuous corn cropping system from an alfisol in Ohio. Soil Tillage Res. 104:39–47.

Vincent, A., 1985. Oxidation and reduction in inorganic and analytical chemistry. John Wiley & Sons, Chichester, U.K.

Vitale, R. J., G.R. Mussoline, J.C. Petura, and B.R. James. 1997. Cr(VI) soil analytical method: A reliable analytical method for extracting and quantifying Cr(VI) in soils. J. Soil Contam. 6:581–593.

Wang, Z.P., D. Zeng, and W.H. Patrick, Jr. 1997. Characteristics of methane oxidation in a flooded rice profile. Nutr. Cycl. Agroecosyst. 49:97–103.

Wardman, P. 1989. Reduction potentials of one-electron couples involving free radicals in aqueous solution. J. Phys. Chem. Ref. Data 18:1637–1756.

Welsch, M., and J.B. Yavitt. 2007. Microbial CO_2 production, CH_4 dynamics and nitrogen in a wetland soil (New York State, USA) associated with three plant species (Typha, Lythrum, and Phalaris). Eur. J. Soil Sci. 58:1493–1505.

Westcott, C.C., 1978. pH measurements. Academic Press, New York.

White, A.F., and M.L. Peterson. 1996. Reduction of aqueous transition metal species on the surfaces of Fe(II)-containing oxides. Geochim. Cosmochim. Acta 60:3799–3814.

Whitfield, M. 1974. Thermodynamic limitations on the use of the platinum electrode in Eh measurements. Limnol. Oceanogr. 19:857–865.

Willis, L.G. 1932. Oxidation–reduction potentials and the hydrogen ion concentration of a soil. J. Agric. Res. 45:571–575.

Wu, W.-M., J. Carley, J. Luo, M. Ginder-Vogel, E. Cardanans, M.B. Leigh, C. Hwang,. S.D. Kelly, C. Ruan, L. Wu, T. Gentry, K. Lowe, T. Mehlhorn, S.L. Carroll, M.W. Fields, B. Gu, D. Watson, K.M. Kemner, T.L. Marsh, J.M. Tiedje, J. Zhou, S. Fendorf, P. Kitanidis, P.M. Jardine, and C. Criddle 2007. In situ bioreduction of uranium(VI) to submicromolar levels and reoxidation by dissolved oxygen. Environ. Sci. Technol. 41:5716–5723.

Xu, X., H. Li, J. Gu, and X. Li. 2005. Kinetics of the reduction of chromium(VI) by vitamin C. Environ. Toxicol. Chem. 24:1310–1314.

Young, E.O., and D.S. Ross. 2001. Phosphate release from seasonally flooded soils: A laboratory microcosm study. J. Environ. Qual. 30:91–101.

Yu, T.R. 1992. Electrochemical techniques for characterizing soil chemical properties. Adv. Agron. 48:205–250.

15

Soil Colloidal Behavior

Sabine Goldberg
United States Department of Agriculture

Inmaculada Lebron
Centre for Ecology & Hydrology

John C. Seaman
University of Georgia

Donald L. Suarez
United States Department of Agriculture

15.1 Nature of Soil Colloids

15.1.1 Significance of Colloidal Phenomena

The importance of colloids in soil science has been appreciated for many years. However, recent understanding that organic and inorganic contaminants are often transported via colloidal particles has increased interest in colloid science. Essentially, all chemicals and individual species are to some extent reactive with soils, including species such as chloride ions, which undergo repulsion from negatively charged surfaces. With few exceptions, soil chemistry is primarily the chemistry of colloids and surfaces. The primary importance of colloids in soil science stems from their surface reactivity and charge characteristics. The overwhelming majority of surface area and electrostatic charge in a soil resides in the less than 1 μm size fraction with particles with radii between 20 and 1000 nm constituting the major part of the soil surface area (Borkovec et al., 1993). A significant fraction of reactive soil colloidal material falls within the <100 nm size range and thus is relevant to the growing interest in the properties and behavior of nanoparticles. Furthermore, soil is often the ultimate repository for anthropogenic nanomaterials of environmental concern (Hochella, 2008; Theng and Yuan, 2008; Waychunas and Zhang, 2008). The unique aspects of "nanoscience" as a discipline separate from colloid science reflect deviations in material properties in the nanoparticle size range, especially for materials <10 nm, and in many cases the lack of a natural bulk analog in the larger size fractions, for example, ferrihydrite (Hochella, 2008; Waychunas and Zhang, 2008).

Characterizations of size, shape, surface area, surface charge density, and changes in surface charge are required for understanding the processes of adsorption, flocculation, dispersion, and transport in soils and the resultant changes in soil hydraulic properties, as well as chemical migration. Since the major part of the surface area is in the colloidal fraction of the soil, almost all surface-controlled processes including adsorption reactions, nucleation, precipitation, and dissolution involve colloids. Colloids are reactive not only because of their total surface area but also because of enhanced reactivity related to rough surfaces and highly energetic sites, as well as the effects of electrostatic charge. Colloid charge is associated with substitution of lower charge cations for those of higher charge in the mineral lattice (which results in a net permanent charge) as well as surface charge associated with broken bonds. The charge associated with broken bonds is characterized as variable charge in as much as the solution influences the surface speciation (Chapter 16). In addition to these chemical processes, colloids are mobile in soils and thus not only affect the chemical transport of otherwise immobile chemicals but also exert a strong influence on soil hydraulic properties.

15.1.2 Types of Soil Colloids

Colloidal particles are defined as having an equivalent spherical radius smaller than 1 μm (van Olphen, 1977). A homogeneous dispersion of colloidal particles in a liquid is called a colloidal dispersion. If the particles are large and settle rapidly, the dispersion is called a suspension. A colloidal dispersion is defined as a system where particles of colloidal dimensions are dispersed in a continuous phase of a different composition (van Olphen, 1977).

15.1.2.1 Oxides

Oxides, including hydroxides and oxyhydroxides, are ubiquitous constituents of soils, occurring as both discrete particles and as coatings on other soil surfaces. Oxide minerals that are commonly found in the soil clay fraction are discussed in Chapter 22.

Hydroxylation of oxide minerals can either be structural and/or occur by chemisorption of water in an aqueous medium (Schwertmann and Taylor, 1977). Edge hydroxyl groups on oxides and clay minerals represent the most abundant and reactive surface functional groups in soils (Sposito, 1984). Any one type of oxide mineral contains various groups of surface hydroxyls that are distinguishable by crystal plane location and/or extent of coordination to the cations of the bulk structure. However, as a simplification, it is often assumed that each oxide mineral has a single set of homogeneous reactive functional groups. Surface hydroxyl groups on most oxide minerals are amphoteric, exhibiting positive charge at low pH and negative charge at high pH. For this reason, oxide minerals are often referred to as variable charge soil minerals. Table 15.1 provides densities of surface hydroxyl groups for some common oxide minerals in soils.

Boehmite and gibbsite are the only crystalline Al oxides common in soils. Aluminum oxides are the products of intense weathering of aluminosilicate minerals and are most abundant in tropical soils. Gibbsite can also be found in volcanic ash soils of humid regions (Brown et al., 1978). Noncrystalline Al oxides, which have similar structure and chemical characteristics but smaller particle size than crystalline varieties, often dominate the chemical reactions with anions in soils (Hsu, 1977). Aluminum oxides play an important role in ion adsorption, stabilization of soil aggregates, and flocculation of soil particles.

Iron oxides are found in most soils and provide soil horizons with their red, yellow, and brown colors (Brown et al., 1978). Most iron oxides are the weathering products of iron-containing silicates. Goethite, which is the most common Fe oxide in temperate, subtropical, and tropical soils, is usually thermodynamically the most stable (Schwertmann and Taylor, 1977). Soil goethites are usually fine grained and contain appreciable substituted Al. Lepidocrocite is a minor constituent of waterlogged temperate soils undergoing alternating oxidizing and reducing conditions, whereas hematite is a common soil mineral that can

be inherited from parent materials or formed pedogenically in warm climatic regions (Brown et al., 1978). The two magnetic Fe minerals, magnetite and maghemite, occur in soils; the former is inherited from parent rock, while the latter is formed pedogenically in highly weathered soils (Brown et al., 1978). Ferrihydrites are poorly crystalline, have indefinite composition, and occur as very small particles with high surface area (Schwertmann and Taylor, 1977). Ilmenite is an uncommon mineral usually inherited from igneous or metamorphic parent rocks (Brown et al., 1978). Iron oxides play an important role in ion adsorption and in aggregation and cementation of soil particles.

Manganese oxides occur widely in soils as minor constituents, mainly as dark coatings on particle surfaces. Manganese oxides are chemically complex, existing as a continuous range of compositions between MnO and MnO_2 (Brown et al., 1978). Birnessite, vernadite, lithiophorite, and hollandite are the most common crystalline manganese minerals in soils (McKenzie, 1989). Birnessite occurs in both acid and alkaline soils, while lithiophorite occurs mainly in neutral to acid soils (Brown et al., 1978). These oxides supply Mn for plant nutrition. Manganese oxides exhibit a strong adsorption capacity for metal cations, especially copper, due to their pH-dependent charge, small particle size, and large surface area (McKenzie, 1989).

Rutile and anatase are the common titanium oxides occurring in soils. Rutile is a high-temperature form occurring in igneous and metamorphic rocks (Hutton, 1977). Anatase is a low-temperature form occurring as an alteration product of titanium containing minerals such as ilmenite and is much less abundant than rutile (Brown et al., 1978). Titanium oxides are present in both the coarse and fine fractions of soils and are very insoluble (Hutton, 1977).

Quartz is not only the most abundant silicon oxide but also the most abundant mineral in most soils. Most quartz is found predominantly in the sand, silt, and coarse clay fractions of soils (Wilding et al., 1977). Silicon oxides are generally considered inert having a small surface area and little surface charge.

15.1.2.2 Clay Minerals

The clay fraction of most soils is dominated by various layer silicate clay minerals. Layer silicate clay minerals are classified as 1:1 where each layer consists of one tetrahedral silica sheet and one octahedral alumina sheet, 2:1 where each layer consists of one octahedral sheet sandwiched between two tetrahedral sheets, or 2:1:1 where a metal hydroxide sheet is sandwiched between the 2:1 layers. Layer silicate minerals common in soils are discussed in Chapter 21. A discussion of silicate structures is provided by Schulze (2002).

Layer silicate clay minerals are characterized by isomorphic substitution of lower valence cations in either or both the tetrahedral and the octahedral sheets. This excess of negative charge is balanced by other cations, either inside the crystal or on the external surfaces (McBride, 1994). Layer charge is an electrostatic charge balanced outside of the structural unit and determines the strength and type of bonding occurring between the basal planes. Charge arising from isomorphic substitution is

TABLE 15.1 Densities of Surface Hydroxyl Groups on Oxide Minerals

Solid	Site Density Range (Sites nm^{-2})
Gibbsite	2–12
Goethite	2.6–16.8
Hematite	5–22
Ferrihydrite	1.1–10.1
MnO_2	6.2
TiO_2	2–12
Amorphous SiO_2	4.5–12

Source: Adapted from Davis, J.A., and D.B. Kent. 1990. Surface complexation modeling in aqueous geochemistry. Rev. Mineral. Geochem. 23:177–260.

TABLE 15.2 Charge Characteristics and Cation Exchange Capacities of Clay Minerals

Solid	Charge per Unit Half Cell		Cation Exchange Capacity (cmol$_c$ kg^{-1})
	Tetrahedral	Octahedral	
Kaolinite	0	0	1–10
Smectite			80–120
Montmorillonite	0	−0.33	
Beidellite	−0.5	0	
Vermiculite	−0.85	+0.23	120–150
Mica			20–40
Muscovite	−0.89	−0.05	
Chlorite			10–40

Sources: Bohn, H.L., B.L. McNeal, and G.A. O'Connor. 1985. Soil chemistry. John Wiley & Sons, New York; McBride, M.B. 1994. Environmental chemistry of soils. Oxford University Press, New York.

called permanent charge because it is independent of solution pH. Layer silicate clay minerals also possess variable charge located at the broken edges of the particles. At the edges of the octahedral sheet, hydroxyl ions attached to Al cations are called aluminol groups. Similar to the hydroxyl groups on oxide minerals, aluminol groups are amphoteric. At the edges of the tetrahedral sheet, hydroxyl groups attached to Si cations are called silanol groups. Silanol groups do not undergo protonation, but dissociate and become negatively charged at high pH. Adsorbed cations on clay minerals balance both pH dependent and permanent charges. Table 15.2 provides charge characteristics and cation exchange capacities (CEC) for some common clay minerals in soils.

Kaolinite is one of the most widespread clay minerals in soils, being most abundant in soils of warm moist climates (Dixon, 1977); while halloysite is formed through acid weathering and in soils of volcanic origin. The halloysite structure is the same as the kaolinite structure but contains a sheet of water molecules between the layers. Both minerals have low colloidal activity, low surface area, and low CEC and anion exchange capacity (AEC) (Dixon, 1977). The CEC and AEC of the 1:1 minerals are predominantly pH dependent (McBride, 1994).

Micas are abundant in soils, occurring as primary minerals inherited from soil parent materials. Micas strongly retain interlayer potassium ions, rendering them nonexchangeable and reducing the CEC of these minerals. Through weathering, micas release K and provide an important natural source of this plant nutrient (Fanning and Keramidas, 1977). Illite is a secondary mineral that is less crystalline, contains less K, and contains more water than muscovite mica (McBride, 1994). Micas and illites are nonswelling minerals.

Smectites constitute an important group of 2:1 clay minerals. Members that are important in soils include montmorillonite, beidellite, and nontronite. Smectites are most significant in moderately weathered soils and have high colloidal activity and high surface area. Smectites are responsible for a large part of the CEC and the majority of the shrink/swell properties of smectitic soils (Borchardt, 1977).

Vermiculite is an important clay mineral in soils that is formed as an alteration product of muscovite and biotite micas (Douglas, 1977). Vermiculite is widely distributed and has a wide particle size range. Vermiculites contain hydrated magnesium cations that can be readily exchanged by K and ammonium ions, resulting in collapse of the clay layers and fixation of these nutrient ions. Vermiculites have high CEC and high surface area but exhibit limited swelling.

Chlorites are 2:1:1 layer silicates that occur extensively in soils. The hydroxide interlayer sheet is usually dominated either by brucite [Mg(OH)$_2$] or by gibbsite [Al(OH)$_3$]. This interlayer sheet restricts swelling, decreases effective surface area, and decreases effective CEC (McBride, 1994). Chlorites are nonswelling silicates.

15.1.2.3 Organic Matter

Soil organic matter (SOM) refers to the mixture of products resulting from microbial and chemical transformations of organic residues and is discussed in Chapter 11. An important component of SOM is called humus, a complex and microbially resistant mixture of amorphous and colloidal substances. These substances are the result of modifications of original tissues or synthesis by soil microorganisms. Humic substances are subdivided into humic acid, fulvic acid, and humin using a separation scheme based on solubility in strong acid and base (McBride, 1994). The structure and composition of humus are complex and incompletely known. The structure contains a variety of reactive functional groups including carboxyl R–COOH, phenol C$_6$H$_5$OH, alcohol R–CH$_2$OH, enol R–CH=CH–OH, ketone R–CO–R′, quinone O=C$_6$H$_4$=O, ether R–CH$_2$–O–CH$_2$–R′, and amino R–NH$_2$ (Stevenson, 1982). Humus is amorphous and highly colloidal; its surface area, ion adsorption, and CEC are greater than those of layer silicate clay minerals (McBride, 1994). The presence of humus usually promotes aggregation of soil particles.

15.1.3 Properties of Soil Colloids

15.1.3.1 Particle Size and Shape

Colloids in natural systems are characterized by a continuous particle size distribution (PSD; polydispersivity) of extreme complexity and diversity. Organisms, organic macromolecules, minerals, clays, oxides, and combinations of any of them constitute the colloidal fraction in soils. The distribution of shapes, densities, surface chemical properties, and chemical composition vary widely with size. Some fractions of the size spectrum may be living, and all particulates are subject to diverse physical, chemical, and biological processes that can alter size distribution, shape, or chemical composition (Kavanaugh and Leckie, 1980).

Colloids are dynamic particles, subject to constant alteration; the distribution of particle sizes in natural systems is the result of a number of processes, which either bring the particles together (coagulation) or disrupt existing aggregates (dispersion) (Filella and Buffle, 1993; Buffle and Leppard, 1995a, 1995b). Particle size is an important parameter in the characterization

of colloids. Sequential gravimetric sedimentation has been the classical method for measuring PSDs in soils. However, this technique has proven to be unreliable for particle sizes in the colloidal range (1–1000 nm). The reason for the lack of reliability is the combination of Brownian motion and convection currents, which each exert a significant influence on settling at diameters below ~1 μm in water.

Awareness of the environmental importance of colloids, for example, remediation schemes using engineered nanoparticles, and studies on the ecotoxicology of the products created by the emerging nanotechnology industry have accelerated the development of analytical techniques for nanoscale research (Wilkinson and Lead, 2007). These analytical techniques provide quantification, analyses, and characterization of the size, shape, and distribution of colloids in polydisperse systems within the environment (Handy et al., 2008). Some of the techniques more commonly used in soil science for colloid characterization and determination of PSD are reviewed in the following section. These include centrifugation, particle size analysis using the Coulter principle, field flow fractionation (FFF), atomic force microscopy (AFM), electron microscopy (EM), and acoustic spectroscopy. A recent publication by the International Union of Pure and Applied Chemistry (IUPAC) Wilkinson and Lead (2007) provides more detailed information on the various techniques.

It is important to consider that each particle size measurement technique has different accuracy and precision. In other words, detection limits and detection windows, corresponding to different size ranges, are technique dependent (Table 15.3). Not all

techniques are able to accurately measure the full scale of size ranges for colloids in polydisperse samples. Furthermore, most of the colloidal-sizing techniques do not measure size directly, but rather determine a physicochemical property from which the size is calculated (Lead and Wilkinson, 2007). For example, scanning and transmission electron microscopic techniques determine the physical dimensions of the projected area of the particles (Lebron et al., 1999), light scattering and flow-FFF generally determine the diffusion coefficients (Lead et al., 2000; Hassellov et al., 2007), and sedimentation-FFF (Sd-FFF) and other centrifugation-based techniques measure the buoyant mass (Hassellov et al., 2007). Although particle size can be estimated from projected areas, diffusion coefficients, and buoyant mass, the calculations are based on a number of assumptions, which if not met, will reduce the quality of the results (i.e., sphericity, homogeneous charge distribution, absence of coulombic interactions among particles). Therefore, it is not uncommon to obtain different PSDs for the same sample when using different techniques (Lead and Wilkinson, 2007). Even more direct techniques like scanning electron microscopy (SEM) and environmental scanning electron microscopy (ESEM) have their limitations. Doucet et al. (2004) observed differing colloidal morphologies in preparations obtained from the same sample with SEM and AFM. They attributed these differences to the sample preparation required for each technique. The difficulty in obtaining similar PSDs using different techniques indicates the limitation of the individual techniques. Therefore, it is good practice to use the results of several characterization techniques simultaneously (Lead and Wilkinson, 2007; Hassellov et al., 2008).

15.1.3.1.1 Measurement Methodology

Minimization of sample handling and processing is recommended when analyzing colloidal systems. Currently, no in situ technique allows direct measurement of the PSD for soils. Furthermore, measurement of colloid particles cannot be compared to standard samples to evaluate the quality of measurements made because there are no accepted standards or reference materials for natural colloids (Lead and Wilkinson, 2007).

15.1.3.1.2 Electron Microscopy

EM is one of the few techniques capable of measuring the size of particles across the entire colloidal range (Table 15.3). Both transmission electron microscopy (TEM) and SEM require deposition of aqueous suspensions, sample evacuation, and in the case of SEM, sample coating with a conducting material, typically graphite or gold, to reduce charging from the beam. Hence sample preparation can, and most likely will, alter the original particle morphology and size distribution (Buffle and Leppard, 1995a, 1995b; Chanudet and Filella, 2006). However, new sample preparation techniques are constantly being developed to counteract artifacts generated by dehydration. For example, freeze-drying techniques or the use of hydrophilic resins and multimethod TEM sample generation techniques can be used to stabilize the three-dimensional structure of colloids (Mavrocordatos et al., 2007).

TABLE 15.3 Operational Range of Colloid Particle Size Characterization for a Variety of Analytical Techniques and the Inferred Colloidal Dimension Quantified by Each Technique

Analytical Technique	Quantified Parameter	Approximate Analysis Size Range (nm)
Filtration	Equivalent pore size diameter	100–>1000
Ultrafiltration	Equivalent molar mass	1–100
Centrifugation	Equivalent spherical volume diameter	10–>1000
Dialysis	Equivalent molar mass	0.5–100
ESEM	Projected area	40–>1000
SEM	Projected area	10–>1000
TEM	Projected area	1–>1000
AFM	Three dimensions	0.5–>1000
Fl-FFF	Hydrodynamic diameter	1–1000
Sd-FFF	Equivalent spherical volume diameter	50–1000
DLS	Hydrodynamic diameter	3–>1000
Acoustic spectroscopy	Equivalent particle diameter	5–>1000

Source: Adapted from Hassellov, M., J.W. Readman, J.F. Ranville, and K. Tiede. 2008. Nanoparticle analysis and characterization methodologies in environmental risk assessment of engineered nanoparticles. Ecotoxicology 17:344–361.

ESEM can, in theory, be used to quantify colloids under ambient conditions, as can AFM that is generally used only to determine the dimensions and characteristics of individual colloids. While minimal sample manipulation is an advantage, the disadvantage is a reduction in the resolution of the techniques; both ESEM and AFM produce much better resolution at lower relative humidity. Another serious problem in AFM imaging of liquids is the alteration of the AFM-derived signal due to the uptake of nanoparticles onto the AFM cantilever (Lead et al., 2005). Applications and new sample preparation techniques for environmental colloids using AFM are discussed by Balnois et al. (2007). Force-volume mode AFM has been used to evaluate the heterogeneous distribution of charge on clay surfaces (Taboada-Serrano et al., 2005).

Microscopy, despite being a very powerful technique, is not widely used for routine particle size analysis. The reasons for its limited use are the high cost of the equipment and the small sample volumes that can be scanned at any given magnification. Image analysis software facilitates the quantification of the different particle metrics in the micrographs. However, obtaining sufficient particles to allow representative and robust statistics requires the scanning of many micrographs of the specimen. Nevertheless, the determination of morphology and particle shape factor still remains the strength of microscopic techniques. Automated instrumental analysis routines that take advantage of enhanced beam stability combined with image analysis software, and greater computing capacity have the potential to address such limitations (Seaman, 2000; Laskin and Cowin, 2001). Additionally, SEM, ESEM, and TEM coupled with energy dispersive x-ray (EDX) spectroscopy can provide valuable chemical information about individual particles. It must be recognized that the resulting x-ray signal in SEM may be generated from a sample region larger than the particle of interest (Goldstein et al., 1992; Seaman, 2000). Furthermore, TEM can also, by measuring the x-ray spectra emitted by the specimen, resolve crystal spacing. Selective area electron diffraction can be used to identify colloidal minerals and to determine their degree of colloid crystallinity.

15.1.3.1.3 Centrifugation

A particle falling through an infinite fluid will eventually travel at a terminal constant velocity determined by the size of the particle and the resistance offered by the fluid. The terminal velocity in a centrifugal field is not constant, but rather a function of distance from the axis of rotation. Measurement of this radius is necessary in order to calculate the particle size. The relationship between the movement of the particle and the movement of fluid around that particle may be reduced to the Stokes equation. For fluid moving past a particle of diameter (d_p), the ratio of the inertial transfer is described by the dimensionless parameter, the Reynolds number (R_E):

$$R_E = \frac{\rho_0 u d_p}{\eta} \qquad (15.1)$$

where u is the velocity. If $R_E \leq 0.2$, the fluid conditions are described as streamlined or laminar, and the drag on the particle is due mainly to viscous force within the fluid. Particles with high densities or large particle diameters may be moving with velocities that exceed $R_E = 0.2$, and in this situation, they are likely to enter the region of turbulent flow, where velocities are more difficult to calculate. In a centrifugal field, Stokes' equation has the form:

$$u = \frac{(\rho - \rho_0)\omega^2 d_p^2}{18\eta} = \frac{\ln(r/s_0)}{t} \qquad (15.2)$$

where
- ω is the rotational velocity (rad s^{-1})
- t is the time (s) required for a particle of diameter d_p to move from its starting point radius (s_0) to the analytical radius (r)

Application of the Stokes equation requires certain assumptions that are not always achieved. All of these assumptions are critical to the measurement of the size of the sedimenting particles. The first assumption is that the particles are spherical, smooth, and rigid. Since this assumption is almost never valid, the diameter calculated is an equivalent or Stokes diameter (d_{St}). It is assumed that the particle terminal velocity is reached instantly, although calculations show that a finite but small time is actually required before this condition is reached. The particle is assumed to be moving without interference or interaction from other particles in the system. This assumption is only true at high dilutions (<1%) that ensure considerable separation between particles. Also, it is assumed that inertial effects are not present and that the fluid exhibits only Newtonian flow properties. Since water is generally the dilution medium in soils and colloidal particles are <1 μm, these assumptions are usually valid. A more detailed analysis of the methodology as well as a description of different centrifugation methods is provided by Bunville (1984), Groves (1984), Koehler et al. (1987), Holsworth et al. (1987), and Coll and Oppenheimer (1987).

15.1.3.1.4 Coulter Effect

The increase in the resistance across a small aperture produced by a nonconducting particle in a conducting medium is called the Coulter effect. The magnitude of this increase in resistance (ΔR) for a spherical particle of diameter (d_p) suspended in an aperture of diameter (D_a) is

$$\Delta R = \frac{8P_f d_p}{3\pi D_a^4}\left[1 + \frac{4}{5}\left(\frac{d_p}{D_a}\right)^2 + \frac{24}{35}\left(\frac{d_p}{D_a}\right) + \cdots\right] \qquad (15.3)$$

where P_f is the resistivity of the conducting medium. This resistance pulse (ΔR) results in a voltage pulse ($i\Delta R$) for a sphere

of diameter d_p, where i is the current across the aperture. The resulting voltage pulses are counted and scaled using a multichannel analyzer.

Instruments utilizing the resistive pulse technique require calibration using standard particles with known diameter to assign a particle size to each of the thresholds. This procedure takes into account the dimensions and electrical characteristics of the aperture and the conducting medium.

The advantage of the resistive pulse technique is that no other properties of the particle, such as refractive index or specific gravity, are required for the interpretation of the data in terms of a PSD. Developments in instrumentation for particle size analysis using the resistive pulse technique allow the analysis of particle sizes <1 µm but the range is limited (typically 0.4–<1 µm) (Bunville, 1984).

15.1.3.1.5 Dynamic Light Scattering or Photon Correlation Spectroscopy

Dynamic light scattering (DLS), also called photon correlation spectroscopy, or quasielastic light-scattering measures the fluctuation in scattered intensity of a laser beam over small-time intervals when it passes through a small volume of particles under Brownian motion. These fluctuations are dependent on the diffusion coefficient of the particles.

When the particles have a regular shape other than spherical, the depolarized component can be used to study the particle rotational diffusion coefficient. The rotational and translational diffusion coefficients obtained in conjunction with theoretical, hydrodynamic relationships contain information about the particle dimensions. For a nonspherical particle larger than the incident wavelength, light is scattered from different parts of the same particle producing interferences, which are dependent on the angle of the scattered intensity and characteristic of a particular particle shape (Pecora, 1983).

Limitations of DLS are due to the assumptions made in calculating the particle radius from the diffusion coefficients. These assumptions include sphericity, nonpenetrable spheres, and noncolumbic forces existing among the particles. A more major limitation, when applied to colloidal systems, is the strong particle size dependence of the scattered light intensity. Larger particles have a much larger influence than smaller particles, biasing size quantification toward larger particle sizes. Consequently, previous sample fractionation is advised, since small fractions of dust or other micrometer-sized particles will overshadow the signal of the particles in the colloidal range (Hassellov et al., 2008). Despite limitations, DLS is quick and easy to use, is available in most laboratories, and is very useful for monitoring changes in colloid aggregation. It has been applied successfully to measure particle sizes of colloids in natural systems (Rees, 1990; Ryan and Gschwend; 1990; Lebron et al., 1993; Ledin et al., 1993, 1994; Finsy, 1994; Perret et al., 1994; Newman et al., 1994; Filella et al., 1997).

15.1.3.1.6 Field-Flow Fractionation

FFF is a group of separation techniques capable of fractionating and characterizing the PSD of colloids in the range 0.01–1 µm.

Like chromatography, FFF is an elution methodology in which constituents are differentially retained, and thus separated in a flow channel (Beckett et al., 1997). With this technique, a colloidal sample is introduced into a stream of liquid and subjected to a field (such as gravitational, centrifugal, third cross-flow, thermal gradient, electrical, or magnetic) acting perpendicular to the stream direction (Beckett and Hart, 1993). According to theory, the rate at which particles are displaced downstream, measured as emergence times, can be related exactly to particle properties such as mass, size, and density. However, since different kinds of particles move at different velocities in this system, broad particle populations are sorted into graded size (or mass) distributions along the length of the flow channel. Observation of the shape of the emerging distribution, combined with theory, yields PSD curves called fractograms.

If applied as indicated above, FFF provides highly detailed size distribution curves and is a very flexible technique that can be adapted to different particle types in almost any suspending medium. The more commonly used FFFs in environmental applications are Sd-FFF and flow-FFF (Fl-FFF). Sd-FFF uses a centrifugal field to aid the separation of the colloids, while Fl-FFF uses a cross flow. For detailed information about FFF, Giddings (1993) provides a detailed explanation of this family of techniques and Hassellov et al. (2007) highlight the latest accomplishments of FFF for aquatic colloids and macromolecules.

Detection limits for colloid chemical analyses have been limited using traditional analytical techniques because the amount of colloids collected for analysis is generally limited. Presently, with the availability of low detection limit chemical analysis instrumentation such as graphite furnace atomic absorption spectrometer (GFAAS) and inductively coupled plasma-mass spectrometry (ICP-MS), concurrent particle size determination and chemical analyses can be conducted. Instruments can be assembled either off-line (disconnected) or online (connected) to achieve a more complete description of the colloid nature. Blo et al. (1995) used GFAAS as an off-line detector for Sd-FFF, while Contado et al. (1997) coupled online GFAAS to Sd-FFF to characterize suspended particulate matter from rivers. When analyzing colloidal fractions, techniques like inductively coupled optical emission spectrometry (ICP-OES) or ICP-MS that allow the simultaneous collection of multiple element signals are preferable. Sd-FFF was coupled with ICP-MS for the first time by Beckett (1991). Since then, many scientists have produced detailed chemical information about colloidal size fractions and associated elements (Murphy et al., 1993; Ranville et al., 1999). As per Table 15.3, Sd-FFF is only capable of separating particles down to ~50 nm. Since many colloids of interest are smaller, Chittleborough et al. (2004) developed the Fl-FFF-ICP-MS technique that can operate across the entire colloidal size range.

15.1.3.1.7 Acoustic Spectroscopy

Acoustic spectroscopy methods measure the propagation velocity and attenuation of sound waves (i.e., 1–100 MHz) passing through a colloidal suspension, providing information about the PSD, rheology, and electrokinetic behavior of the suspension.

Dukhin and Goetz (2002) provide a thorough discussion of acoustic methods and the six mechanisms of acoustic attenuation associated with a colloidal suspension: viscous, thermal, scattering, intrinsic, electrokinetic, and structural signal dissipation. The application of acoustic spectroscopy to colloid characterization assumes that each attenuation process functions independently and that the overall attenuation is the summation of the independent processes.

Viscous and thermal dissipation are the most important because colloids mainly interact with sound waves hydrodynamically through the generation of oscillating shear waves and thermodynamically through temperature losses. The resulting acoustic spectra are generally insensitive to the electrical conductivity of the solution and the charge of suspended particles and provide no information concerning particle morphology (Dukhin and Goetz, 2002; Seaman et al., 2003).

Acoustic attenuation attributed to the suspension is generally determined by measuring the relative change in signal at each frequency with precise changes in gap distance or sample path length in spectroscopic terms. Thermal losses dominate in emulsions and low-density dispersions, so that viscous losses may be neglected. For rigid submicron particles (i.e., soil clays, oxides), viscous attenuation dominates and limited information regarding composition of the particles (density), the media (solution density and shear viscosity), and the relative volume fraction of the two phases is required for estimating particle size, providing a minimum detectable particle size of approximately 10 nm. For complex environmental samples, however, such information may be lacking. Acoustic scattering becomes more important with increasing particle size, and sound speed must also be considered. When characterizing "soft" particles having a limited density contrast compared to the suspending solution, such as latex particles and polymers, additional information concerning their thermal expansion properties is required for interpretation of the attenuation spectra (Dukhin and Goetz, 1996, 1998, 2002).

Acoustic methods offer several advantages compared to other instrumental techniques for evaluating colloid size and surface charge properties. Ultrasound can propagate through suspensions to a much greater degree than light. Therefore, acoustic analysis can be conducted at relatively high solid to solution ratios (up to 30% solids by volume) that are more analogous to soil conditions in the field. However, the analysis may require more colloidal material than may be readily available. Analysis may be indicative of particle interaction within the intact suspension, lessening the impact of trace artifacts, that is, dust, bubbles, suspension heterogeneity due to particle segregation, and filtration artifacts that can bias sizing methods such as SEM and DLS. The applicable sizing range is much larger than light-scattering methods, that is, ≈5 nm–1000 μm. It is also less biased with respect to larger particles, making it more suitable for complex polydisperse systems. Acoustic methods are insensitive to sample convection, allowing stirring or agitation of the sample during analysis as required for reactive titration (e.g., Sun et al., 2006) or for the characterization of low charge,

inherently unstable suspensions, such as materials close to their zero point of charge (ZPC) and/or critical coagulation concentration (CCC) (Babick et al., 2000; Dukhin and Goetz, 2002; Kosmulski et al., 2002; Guerin and Seaman, 2004; Guerin et al., 2004; Delgado et al., 2005). Furthermore, the relatively large sample volume allows for the collection of subsamples throughout acoustic analysis for characterization by other analytical methods (Seaman et al., 2003).

Acoustic spectrometers for use in characterizing colloidal suspensions became commercially available in the 1990s. Despite several advantages when compared to light-scattering techniques, the application of acoustic-based methods to the study of soil colloids has generally been restricted to the characterization of mineral standards or synthesized mineral analogs, such as goethite and hematite (Gunnarsson et al., 2001; Kosmulski, 2002; Kosmulski et al., 2002, 2003; Appel et al., 2003; Guerin and Seaman, 2004; Guerin et al., 2004; Delgado et al., 2005). However, acoustic methods of suspension characterization rely on the interpretation of macroscopic sample properties, that is, the acoustic attenuation spectrum of a suspension, using idealized model algorithms with various simplifying assumptions and a few known system parameters, such as suspension concentration and particle density. The limited information can often result in systems that are "ill defined" and can be described by multiple answers, that is, PSDs (Dukhin and Goetz, 2001; Babick and Ripperger, 2002), a problem that also plagues light-scattering techniques (Schurtenberger and Newman, 1993). As such, acoustic methods are most suitable for evaluating relative changes in colloid aggregation and surface charge for suspensions in response to known changes in solution chemistry rather than for comparing subtle differences between poorly defined environmental samples.

15.1.3.1.8 Applications of Particle Size Methods

PSD is a fundamental soil property, affecting soil surface area, bulk density, porosity, water retention, and hydraulic behavior. Furthermore, precise information about colloidal size and shape is important because submicron-size colloids often act as vehicles that control the transport and fate of adsorbed pollutants (hydrophobic organic compounds, toxic trace metals, and radionuclides [de Jonge et al., 2004a]). Bacteria and viruses are part of the colloidal pool in natural environments; their characterization and transport are significant for understanding biogeochemical processes (Rockhold et al., 2004), epidemic evolution (Bertuzzo et al., 2008), and the spread of diseases in general (Khilar and Fogler, 1984; McDowell-Boyer et al., 1986; Kia et al., 1987; Ryan and Elimelech, 1996; Kretzschmar et al., 1999; de Jonge et al., 2004a; McCarthy and McKay, 2004; Tufenkji, 2007). Submicron-size colloids have been insufficiently studied in the past because methods for their isolation, detection, and characterization have been inadequate, with the exception of a few examples (Kaplan et al., 1993; Kretzschmar et al., 1993; Chanudet and Filella, 2006). However, with the new fractionation methods and the coupling with low detection limit analytical instruments (FFF-ICP-MS), a new era in the characterization of soil colloids

is commencing (Chittleborough et al., 2004; Ranville et al., 2005; Lead and Wilkinson, 2007).

Despite the new advances in the characterization and analysis of colloids, many unanswered questions exist in regard to in situ colloidal behavior in natural environments, in particular, the chemical nature of colloids present and their structure, size, and shape distributions (Filella et al., 1997; Lead and Wilkinson, 2007). New methods of in situ visualization of colloids in porous media are the research focus of several recent publications; light transmission and epifluorescent microscopy are some of the new techniques developed by Crist et al. (2004) and Baumann and Werth (2004) for the observation and modeling of colloidal transport in porous media.

15.1.3.2 Surface Area

Surface area must be regarded as a relative term, in as much as it is scale dependent, as well as often dependent on the chemical and physical conditions of a system. Determinations of surface area range from particle size calculations assuming smooth surfaces and simplified geometry, generally termed geometric surface area, to possible molecular level calculations based on the distances between surface ions in a mineral structure. Because the measurement is scale and system dependent, there is no universally accepted way to measure surface area. Determination as to which measurement system to utilize should consider the scale and chemical conditions required by the application. Kinetic reactions that are diffusion controlled should likely consider geometric surface area, while surface-controlled reactions (such as some adsorption and some dissolution/precipitation reactions) should consider surface area at the scale of the reacting molecule. In most instances, surface area is related to surface reactivity, either adsorption or surface-controlled kinetic processes.

15.1.3.2.1 Measurement Methodology

The results of surface area determinations must be interpreted within the context of the size and orientation of the adsorbate, as well as the attractive forces between the surface and the adsorbate. This distinction is particularly important for clays such as smectites, which can be considered to have internal as well as external surface area. Internal surface area is representative of the surface area of the interlayers. Inert gases such as N_2 are not able to enter the interlayer positions, and thus, measure only external surface area of clay particles. In contrast, polar molecules such as ethylene glycol, ethylene glycol monoethyl ether (EGME), and water are able to cause expansion of the layers and penetrate into interlayer positions. Use of such molecules results in measurement of internal and external surface area. Since water is the solvent in environmental systems, these total surface area measurements are appropriate for adsorption studies. Soil surface area measurements obtained using N_2, water, and EGME were highly correlated with each other and with clay content when considering soils with similar mineralogy but not for soils with differing mineralogy (de Jong, 1999).

To characterize adsorption methods of surface area determination, it is useful to distinguish between chemical and physical adsorption. Physical adsorption is characterized by low heats of adsorption without structural changes at the surface, fully reversible and rapid reactions since no activation energy is required, coverage of the entire surface rather than specific sites, little or no adsorption at elevated temperatures, and potential coverage by more than one layer of adsorbate (Lowell, 1979). Chemisorption is characterized by high heats of adsorption, localization of adsorption at specific surface sites, and irreversible reaction. The term "specific surface area" has been used to denote the surface area of a material expressed on a mass basis ($m^2 \ g^{-1}$).

15.1.3.2.2 Gas Adsorption Using the BET Equation

The BET equation is commonly used in conjunction with physical gas adsorption to measure surface area. The BET equation is named after Brunauer, Emmett, and Teller (1938), who extended the Langmuir theory for monolayer gas adsorption to multilayer adsorption. The Langmuir equation (Langmuir, 1918) is given by:

$$\frac{P}{V} = \frac{1}{kV_m} + \frac{P}{V_m} \tag{15.4}$$

where
  P is the pressure
  V is the volume of gas adsorbed per kilogram of adsorbent at that pressure
  k is a constant
  V_m is the volume of gas adsorbed per kilogram of adsorbent at monolayer surface coverage

The surface area is obtained by determination of $1/V_m$, which is the slope of the P/V versus P plot. The specific surface area is then equal to $1/V_m$ multiplied by the cross-sectional area of the adsorbate and the number of molecules in volume V_m.

The BET relation assumes that there is a dynamic equilibrium between the molecules in the various layers such that the number of molecules in each layer remains constant, although different sites may or may not be occupied at any given time. Use of the equation enables calculation of the number of molecules in a monolayer despite the fact that complete monolayer coverage may not have occurred. The BET equation is written as (Lowell, 1979):

$$\frac{1}{W[P_0/P]} = \frac{1}{W_m C} + \frac{C-1}{W_m C}\frac{P}{P_0} \tag{15.5}$$

where
  P is the adsorbate gas pressure
  P_0 is the adsorbate pressure at saturation for the temperature of the experiment
  W is the weight adsorbed in the monolayer
  W_m is the weight adsorbed in the complete monolayer
  C is the BET constant

Multipoint BET plots are created by plotting $1/(W(P_0/P - 1))$ on the y axis and P/P_0 on the x axis. The value of W_m is calculated from the slope and intercept. The specific surface area is determined by dividing the total surface area by the sample weight. The region of P/P_0 between 0.05 and 0.35 is usually linear and within the region of pressures corresponding to sufficient adsorption to complete monolayer coverage, and thus, best suited for determination of W_m (Lowell, 1979).

Often the BET surface area can be determined from a single pressure measurement without much loss of accuracy. For relatively high values of C, the intercept value is small relative to the slope and can be approximated by zero. The BET equation is thus reduced to (Lowell, 1979):

$$W_m = W\left(1 - \frac{P}{P_0}\right) \qquad (15.6)$$

Soil and mineral surface areas are most commonly measured by N_2 adsorption, using the BET equation. The calculation is made using the N_2 cross-sectional area of $0.162\,nm^2$ (Gregg and Sing, 1982). Often this area is referred to as the effective or occupied area. Alternatively for surface area $<1\,m^2\,g^{-1}$ the use of Kr is recommended.

Most commonly, the BET method consists of adsorption of N_2 at a fixed partial pressure P in a He–N_2 mixture and measurement of the desorbed N_2 in a pure He gas stream using gas chromatography. Alternative methods include measurement of the mass of N_2 adsorbed. In this method, the sample is evacuated to high vacuum, heated, then cooled to liquid N_2 temperature, and weighed. Quantities of N_2 are then added to the system and a series of weighings is made at various pressures. In this instance, the N_2 partial pressures are equal to the total pressure in the system.

Surface area can also be determined from the sorption of water at one or more vapor pressures (Newman, 1983). In this instance, air-dried samples are reacted in evacuated desiccators containing saline solutions with relative vapor pressures on the order of 0.2–0.4. Samples are equilibrated until there is no further weight change, then samples are dried at 105°C and weighed again. Sorbed water is taken as the difference between the oven-dry weight and the desiccator-equilibrated weight. Surface area is then calculated as with N_2, using Equation 15.5. This method gives values comparable to EGME values for nonexpanding clays but underestimates surface area for smectites due to limited water uptake in the interlayers (de Jong, 1999). For smectitic soils, de Jong (1999) recommended using the Langmuir equation (monolayer) for water sorption indicating that the reduced BET equation (Equation 15.6) is not applicable because interlayer water uptake is limited. de Jong (1999) demonstrated a correspondence close to 1:1 between surface area determined with EGME using the BET equation and with water using the Langmuir expression. It can be argued that water sorption values may be more realistic than EGME values (Pennell et al., 1995), since the interlayer spaces of smectites are not accessible to nonpolar molecules until at least two layers of water are present (Quirk and Murray, 1991).

15.1.3.2.3 Organic Molecules

15.1.3.2.3.1 Ethylene Glycol Ethylene glycol was utilized by Dyal and Hendricks (1950) for determination of total surface area of clays. The method consists of adding excess ethylene glycol to soil or clays and allowing the excess to evaporate under vacuum. It is assumed that when the rate of weight loss of the sample decreases, only a monolayer of ethylene glycol remains. Dyal and Hendricks (1950) calibrated the method assuming a bentonite surface area of $810\,m^2\,g^{-1}$ and calculated that 0.31 mg of adsorbed ethylene glycol corresponded to each square meter of surface area. The method was modified by Bower and Goertzen (1959) using $CaCl_2$-monoglycolate to maintain an ethylene glycol vapor pressure just below that of the saturation vapor pressure. In this method, the sample and the liquid are placed in separate open vessels in an evacuated system and the sample is weighed until it is in equilibrium with the vapor pressure of the ethylene glycol.

15.1.3.2.3.2 Ethylene Glycol Monoethyl Ether Ethylene glycol monoethyl ether (EGME) has replaced ethylene glycol as the polar solvent of choice for determination of surface area. Since EMGE has a higher vapor pressure than ethylene glycol, it requires a shorter reaction time to equilibrate the sample (Carter et al., 1986). A solvate of EGME and $CaCl_2$ is used in the evacuated chamber to lower the vapor pressure of EGME to just below the saturation pressure. Open vessels of EGME/$CaCl_2$ and soil are placed in the chamber and the soil is periodically weighed until no further weight gain is observed. It is assumed that the EGME surface coverage is $5.2 \times 10^{-19}\,m^2$ per molecule and that 0.286 mg adsorbed corresponds to $1\,m^2$ of surface area (Carter et al., 1986). The method is limited in that the EGME affinity for cations results in greater than monolayer coverage at those sites, the assumption that EGME covers all surfaces cannot be properly evaluated, and the large size of the molecule may prevent coverage in small surface voids. A serious reservation of the procedure is the assumption that the average EGME occupancy of smectite surface applies equally well to all soil surfaces, regardless of mineralogy (Tiller and Smith, 1990). These authors found that more EGME was retained per unit area by nonexpanding soil clays such as illites and kaolinites than by smectites resulting in an overestimation in surface area of 50%–100% when using smectite as a reference. Measurement of total surface area of soils with mixed mineralogy using EGME and a single conversion factor based on smectite leads to significant under- and overestimation of surface area of many soils, even including smectitic soils (Tiller and Smith, 1990).

15.1.3.2.3.3 Methylene Blue Methylene blue, an organic cation, is reacted at various concentrations with soil suspensions (typically with organic material removed) under pH-buffered conditions. Measurement of methylene blue concentration before and after reaction with soil is made spectrophotometrically at a wavelength of 665 nm. The adsorbed concentration of methylene

blue is used to calculate surface area assuming an area of 1.3 nm^2 per methylene blue molecule (Hang and Brindley, 1970). This surface area corresponds to the molecule attaching parallel to its long axis. This method can only be used in hydrated systems. Aringhieri et al. (1992) found surface areas determined with methylene blue to be unrealistically lower than those obtained with water adsorption and suggested that the major limitation of the method is its dependence on the soil surface charge characteristics. Alternatively, Borkovec et al. (1993) obtained relatively good agreement between methylene blue and N_2 BET surface areas for four soil samples assuming a methylene blue surface area of 0.247 nm^2, corresponding to a molecular attachment of methylene blue perpendicular to its long axis. Orientation of the organic molecules may be related to surface site characteristics, as well as concentration of adsorbing molecules. These differences illustrate one of the disadvantages of using charged molecules for surface area determination; their use is not recommended.

15.1.3.2.4 Electron Microscopy Image Analysis

Transmission electron microscopy as well as SEM can be used for determination of geometric surface area using a variety of methods, including calculation of planar surface area and assumptions regarding geometry to calculate total external surface area. With this method, only edge roughness can be measured. The SEM method, which offers the possibility of measuring the external surface roughness, consists of collecting two images at different sample tilt angles and constructing a three-dimensional representation of the surface. Surface area is calculated within a grid of fixed lines by summation of the planar surfaces in the grid. The ratio of the calculated surface to the area within the grid gives a measure of surface roughness. Measurements can also be made at different scales, providing information about the size distribution of the surface features. Since these measurements are typically made at the micrometer scale, and measure external surface area, it is not surprising that the values are intermediate between geometric surface areas based on particle size and BET values. Determination of specific surface area requires conversion of particle surface area to a mass basis using the particle density. The assumption that the particle density is equal to the density of specimen samples of the mineral is reasonable and introduces relatively minor errors in comparison to other assumptions made in the calculation. Increasing computer capacity and suitable software makes this method the most useful of the geometric methods.

15.1.3.2.5 Small Angle X-Ray Scattering

In this application, $K\alpha$ radiation from a conventional x-ray tube is scattered by freeze-dried samples in glass capillary tubes. The background-corrected x-ray-scattering intensity is plotted against the scattering vector (q) to obtain the apparent surface fractal dimension (D_s) using the relation (Borkovec et al., 1993):

$$I_q = Aq^{D_s-5} + B \tag{15.7}$$

where
  I_q is the scattering intensity
  A is a proportionality constant
  B is the background correction

Values of D_s obtained for soils using this method are in relatively good agreement with values calculated from gas adsorption surface area (a) for soil particle radii (r) using the relation (Borkovec et al., 1993):

$$a = C\lambda^{2-D_s} r^{D_s-3} \tag{15.8}$$

where
  λ is the size of the probing molecules
  C is the proportionality constant

15.1.3.2.6 Negative Adsorption

Negative adsorption refers to a deficiency in the concentration of an ion in solution adjacent to a solid surface relative to the concentration of the ion in the bulk solution. The deficit is caused by electrostatic repulsion between ions and surfaces of similar charge. Since most charged surfaces in soils are negatively charged, negative adsorption usually relates to solution anion concentrations and negative surfaces. For a 1:1 electrolyte, the diffuse double layer model produces the following approximation (Sposito, 1984):

$$d_e \approx \frac{2}{\sqrt{\beta c}} - \delta \tag{15.9}$$

where
  d_e is the exclusion distance
  c is the concentration of the bulk electrolyte
  β is a constant equal to $1.08 \times 10^{16} \text{ m mol}^{-1}$
  δ is the distance between two planes

The exclusion volume and area (S_e) are related by

$$V_e = S_e d_e \tag{15.10}$$

where V_e is the exclusion volume (the hypothetical volume from which the ion is completely excluded). Substituting into Equation 15.9 yields

$$V_e \approx \frac{2S_e}{\sqrt{\beta c}} - \delta S_e \tag{15.11}$$

The exclusion volume is calculated by first removing the bulk solution, determining the remaining liquid volume, displacing the liquid, and analyzing for the bulk solution and extract concentration. The exclusion volume is then equal to the mass deficit in the extract divided by the concentration in the bulk solution.

A plot of V_e vs. $c^{-0.5}$ should yield a straight line whose slope is proportional to the exclusion specific surface area as shown by inspection of Equation 15.11.

This method was first described by Schofield (1949), who used it to determine montmorillonite surface area from reactions with NaCl, NaNO$_3$, and Na$_2$SO$_4$ solutions. Subsequent work by Edwards et al. (1965a, 1965b) demonstrated that the specific surface area determined with this method varied with cation selected. Calculated values for illites ranged from values close to the N$_2$ BET values with Li to 0 with Cs. In contrast, for montmorillonite, LiCl exclusion volumes were 10 times greater than N$_2$ BET surface areas, while Cs values were comparable to those obtained by N$_2$ BET. Trends in surface area values measured using anion exclusion for Li, Na, K, NH$_4$, Cs, and Ca-montmorillonites were in agreement with those calculated from a tactoid model (Schramm and Kwak, 1982a, 1982b).

15.1.3.2.7 Applications

Surface area measurements are required for a variety of calculations. In most soils, the bulk soil surface area is dominated by the surface area of the clay minerals. Surface charge density, which requires measurement of both surface area and particle charge, has been related to cation exchange selectivity. As expected, increasing surface charge density favors adsorption of the higher valence cation in heterovalent exchange (Maes and Cremers, 1977). Surface charge density is also required for calculation of double layer thickness and for use in a variety of adsorption relationships. Determination of specific surface area is required for chemical studies on many different minerals. In this case, bulk surface area is not appropriate. Controlled laboratory studies are often performed using addition of quantities of a well-characterized mineral having a known surface area. This method is used to study a variety of chemical reactions, such as kinetics of calcite, gypsum, and dolomite dissolution, surface area of calcite for prediction of phosphate adsorption, and addition of various Fe or Mn oxides for study of adsorption and redox processes. Among the various applications of the EGME method, Ross (1978) related shrink/swell properties of soils to surface area and Supak et al. (1978) related the specific surface area of clays to the adsorption of the pesticide aldicarb.

15.1.3.3 Surface Charge

The total net surface charge on a particle (Φ_p) is

$$\sigma_p = \sigma_s + \sigma_H + \sigma_{is} = \sigma_{os} - \sigma_d \tag{15.12}$$

where

σ_s is the permanent structural charge

σ_H is the proton surface charge resulting from the specific adsorption of protons and hydroxyl ions

σ_{is} is the inner-sphere complex charge resulting from specific ion adsorption

σ_{os} is the outer-sphere complex charge resulting from non-specific adsorption

σ_d is the dissociated charge

This definition is similar to the one provided by Sposito (1984) with the exception that total net surface charge results from isomorphic substitution and is generated by specifically adsorbing ions (Hunter, 1981). An inner-sphere surface complex contains no water between the adsorbing ion and the surface functional group; while an outer-sphere surface complex contains at least one water molecule between the adsorbing ion and the surface functional group (Sposito, 1984). Examples of surface functional groups are reactive surface hydroxyl groups on oxide minerals, aluminol and silanol groups on clay minerals, and carboxyl and phenol groups on SOM.

15.1.3.3.1 Measurement Methodology

The total net surface charge can be measured directly using electrokinetic experiments. The point of zero charge (p.z.c.) of a particle is the solution pH value where total net particle charge is zero. The p.z.c. can be measured directly using electrokinetic experiments or indirectly from potentiometric titrations under certain experimental conditions (Sposito, 1984).

Electrokinetic phenomena are processes where a relative velocity exists between two parts of the electrical double layer (Hiemenz, 1977; Hiemenz and Rajagopalan, 1997). In this motion, a thin layer of liquid remains with the solid and a shear plane is located between the solid and liquid phases at some distance from the solid surface (van Olphen, 1977). The electric double layer potential at the shear plane is called the zeta potential (ζ). The assumption that ζ is equal to or very close to the diffuse double layer potential (ψ_d) is supported indirectly by a large body of data on a variety of surfaces (Hunter, 1981, 1989). The principle electrokinetic phenomena that measure zeta potential are discussed later.

15.1.3.3.2 Electrophoresis

Electrophoresis measures the movement of a suspended charged particle in response to an applied electric field. This movement is called electrophoretic mobility (μ_E) and is given by the Smoluchowski equation (Hunter, 1989):

$$\mu_E = \frac{\varepsilon \zeta}{\eta} \tag{15.13}$$

where

ε is the relative permittivity

η is the viscosity

The Smoluchowski equation applies when the particle dimensions are much greater than the double layer thickness (Hunter, 1987). The complete formula relating ζ and μ_E is derived by the theoretical evaluation of the electric force on the charged particle (f_1), the hydrodynamic frictional force on the particle by the liquid (f_2), the electrophoretic retardation force (f_3), a frictional force resulting from the movement of water with the counterions, and the relaxation force (f_4), caused by distortion of the double layer around the particle (van Olphen, 1977). Additional considerations arise

for nonspherical particles and those carrying two double layers such as clays. For these reasons, the Smoluchowski equation is, in general, only approximate and it is advisable to report electrophoresis results as electrophoretic mobility rather than to attempt to convert to zeta potential (van Olphen, 1977).

Electrophoresis is the most common method of determining zeta potential. For colloidal systems, the most appropriate technique is microelectrophoresis where the movement of individual particles is followed directly by microscopy (Hunter, 1981). Microelectrophoresis is only applicable at very low particle concentrations. Electrophoresis can also be studied using laser Doppler velocimetry and photon correlation spectroscopy. The mass transport mobility apparatus measures electrophoretic mobility from the mass of colloids transported to a suitable electrode compartment (Hunter, 1981). This apparatus can be used at much higher particle concentrations than microelectrophoresis.

15.1.3.3.3 Electroosmosis

Electroosmosis measures the movement of the liquid adjacent to a flat, charged surface in response to an electric field applied parallel to the surface. This movement is called electroosmotic velocity (v_{eo}) and is also obtained from the Smoluchowski equation (Hunter, 1987):

$$v_{eo} = \frac{-\varepsilon \zeta E}{\eta} \qquad (15.14)$$

where E is the electric field strength. While it is possible to measure electroosmotic velocity directly using microscopy, it is more common to measure the volume of liquid transported per unit time (Hunter, 1981):

$$\frac{V}{i} = \frac{\varepsilon \zeta}{\eta \lambda_0} \qquad (15.15)$$

where

    V is volume
    i is the electric current
    λ_0 is the electrical conductivity

The material whose zeta potential is being measured is formed into a porous plug and the transport of liquid across a tube in response to an electric field may be obtained by measuring the movement of an air bubble in the capillary providing the return path (Adamson, 1976). It is also possible to measure the electroosmotic flow by applying a counter pressure until the flow is exactly compensated (Hunter, 1981).

15.1.3.3.4 Electroacoustic Spectroscopy

A general discussion of the application of acoustic spectroscopy for characterizing colloidal suspensions was presented earlier in this chapter (Section 15.1.3.1). Both acoustic and electroacoustic spectrometers are commercially available separately or combined in a single instrument. In electroacoustic spectroscopy,

one measures either (1) the colloid vibration potential (CVP) or colloid vibration current (CVI), the induced dipole moment created by the displacement or polarization of the electrical double layer in response to an acoustic wave, or (2) the electrokinetic sonic amplitude (ESA) generated by the movement of a charged particle in response to an electrical field as indicators of colloidal zeta potential. The individual dipole moments created by the interaction of sound waves with a colloidal suspension can be measured as an alternating electrical field that varies with the amplitude of the sound wave. Conversely, the electrophoretic movement of charged particles in response to an applied electrical field gives rise to sound waves, known as the ESA effect. Although electroacoustic spectroscopy can provide PSD, such information is best derived from the conventional acoustic spectra (Dukhin and Goetz, 2002; Seaman et al., 2003).

In contrast to microelectrophoretic light-scattering methods, electroacoustic measurement of zeta potential can be conducted at much higher suspension concentrations (30%–50% solids), an obvious advantage given that sample dilution and changes in solution chemistry can impact the expression of zeta potential and apparent particle size (Dukhin et al., 2001; Dukhin and Goetz, 2002; Delgado et al., 2005). Acoustic methods have been used to evaluate shifts in the isoelectric point (IEP) of materials at high-ionic strengths as indicators of specific and nonspecific sorption mechanisms (Kosmulski, 2002; Kosmulski et al., 2002; Greenwood, 2003). Such applications are best used in combination with other surface active spectroscopic methods for determining sorption mechanisms. In comparing various techniques for evaluating the IEP and p.z.c. of Fe (hydr)oxides, Kosmulski et al. (2003) noted that electroacoustic methods produced higher IEP values than observed for potentiometric titration methods. The cause of the discrepancy was not discussed.

15.1.3.3.5 Streaming Potential

The streaming potential is an electric potential difference generated when liquid adjacent to a charged surface is set in motion by an applied pressure gradient (Hunter, 1987). The streaming potential (Φ_{st}) is also governed by the Smoluchowski equation and given by (Sposito, 1984)

$$\Phi_{st} = \frac{\varepsilon \zeta}{\lambda_0 \eta} \Delta P \qquad (15.16)$$

where ΔP is the applied pressure difference. The streaming potential can be measured in similar fashion as the electroosmotic velocity. Liquid is forced under pressure through a porous plug and Φ_{st} is measured by electrodes in the solution on either end (Adamson, 1976).

15.1.3.3.6 Sedimentation Potential

When particles having charged surfaces settle in a liquid under the force of gravity, a plane of shear is developed. As the particles settle, the interfacial charge is separated since a portion inside the shear plane moves with the particle and the remainder is left behind (Sposito, 1984). An electric potential difference arises

from the separation of charge called the sedimentation potential. The gradient for the sedimentation potential, $d\Phi_{sed}$, is given by (Sposito, 1984)

$$\frac{d\Phi_{sed}}{dz} = \frac{\varepsilon\zeta}{\lambda_0\eta}n\Delta\rho g \qquad (15.17)$$

where
   n is the number of particles per unit volume
   $\Delta\rho$ is the difference in mass density between the particles and the liquid phase
   g is gravitational acceleration
   z is distance

The potential difference is measured by inserting reversible electrode probes at two different heights in the column of settling particles (Hunter, 1981). For low particle concentration, the sedimentation potential is also governed by the Smoluchowski equation (Hunter, 1981).

15.1.3.3.7 Potentiometric Titration

Potentiometric titration measures the surface density of proton surface charge (σ_H) defined as (Sposito, 1984)

$$\sigma_H = \frac{F}{A}(q_H - q_{OH}) \qquad (15.18)$$

where
   F is the Faraday constant
   A is the specific surface area
   q_H is the complexed proton charge (mol)
   q_{OH} is the complexed hydroxyl charge (mol) per unit mass of solid

Titration data consist of pH readings obtained while known amounts of acid or base are added to a solid suspension. A net titration curve is obtained by subtracting a calibration curve obtained by titrating the equivalent supernatant solution. The values of $q_H - q_{OH}$ are given by (Sposito, 1984)

$$q_H - q_{OH} = \frac{C_A - C_B - [H^+] + [OH^-]}{C_s} \qquad (15.19)$$

where
   C_A is the molar concentration of acid added
   C_B is the molar concentration of base added
   $[H^+]$ is the molar proton concentration
   $[OH^-]$ is the molar hydroxyl concentration obtained from pH measurement
   C_s is the particle concentration

In order for Equation 15.19 to be valid, added protons and hydroxyl ions must only react with surface-reactive functional groups whose charge is pH dependent. Usually, other reactions that are also pH dependent occur, such as soluble complex formation, dissolution of solid phases, or complexation with surfaces whose charge is not pH dependent (Parker et al., 1979). Without these corrections, no surface chemical significance can be provided by Equation 15.19. A detailed description of the use of potentiometric titration to determine surface charge is provided by Huang (1981).

15.1.3.3.8 Applications

One of the most important applications of electrokinetic experiments and potentiometric titrations is the determination of the p.z.c. A characteristic of variable charge minerals is the p.z.c. obtained in the presence of an inert electrolyte. This p.z.c. is determined electrokinetically as the pH value where the zeta potential is zero or indirectly from the point of zero net proton charge (p.z.n.p.c.) or from the point of zero salt effect (p.z.s.e.) obtained potentiometrically. The p.z.n.p.c. and the p.z.s.e. are discussed in detail in Chapter 16. The p.z.n.p.c. and the p.z.s.e. are equivalent to the p.z.c. in the absence of surface complex formation. Table 15.4 provides characteristic values of p.z.c. obtained using electrokinetic experiments and potentiometric titrations for a variety of variable charge minerals.

TABLE 15.4   Representative Points of Zero Charge for Various Minerals

Solid	p.z.c.
Electrophoresis	
Goethite	8.8
Hematite	8.5
Magnetite	6.9
Amorphous iron oxide	8.0
Gibbsite	9.8
Bayerite	9.2
Boehmite	9.4
Pseudoboehmite	9.2
Amorphous aluminum oxide	9.3
δ-MnO_2	2.3
Rutile	4.8
Anatase	5.9
Kaolinite	2.9
Streaming potential	
γ-Al_2O_3	9.1
α-Al_2O_3	9.2
Titration	
Goethite	8.7
Hematite	8.6
Magnetite	6.9
Gibbsite	9.8
Boehmite	8.5
Pseudoboehmite	9.3
Amorphous aluminum oxide	9.5
δ-MnO_2	3.6
SiO_2	3.0
Rutile	5.8
Anatase	6.0
Kaolinite	2.9
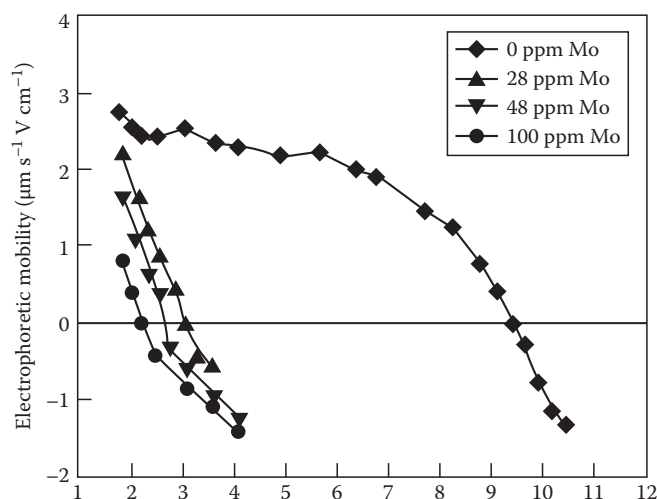
FIGURE 15.1 Shifts in p.z.c. and charge reversal of gibbsite in the presence of molybdate. (Adapted from Goldberg, S., H.S. Forster, and C.L. Godfrey. 1996. Molybdenum adsorption on oxides, clay minerals, and soils. Soil Sci. Soc. Am. J. 60:425–432.)

FIGURE 15.2 Relation between zeta potential and dispersible clay of Na-clay minerals obtained using electrophoresis. (Reprinted from Chorom, M., and P. Rengasamy. 1995. Dispersion and zeta potential of pure clays as related to net particle charge under varying pH, electrolyte concentration and cation type. Eur. J. Soil Sci. 46:657–665. With permission of Blackwell Science Ltd.)

Electrokinetic experiments and potentiometric titrations can be used to infer adsorption mechanisms for adsorbing ions on surfaces. Adsorption of ions that form inner-sphere surface complexes is characterized by shifts in the p.z.c. of the particles and reversals of their electrophoretic mobility with increasing ion concentration (Hunter, 1981). Adsorption of ions that form outer-sphere surface complexes does not produce p.z.c. shifts since they are assumed to lie outside the shear plane. Figure 15.1 presents the shifts in p.z.c. and charge reversals observed for gibbsite upon the specific adsorption of increasing amounts of molybdate. These results are indirect evidence for inner-sphere surface complexation of this ion.

Net particle surface charge is a primary factor in dispersion of clay minerals. Zeta potential as a measure of particle surface charge was also related to percentage of dispersible clay (Chorom and Rengasamy, 1995). Figure 15.2 indicates the relationship between zeta potential measured using electrophoresis, and dispersible clay for Na-saturated kaolinite, montmorillonite, and illite.

The plane interface technique determines the electroosmotic velocity at large plane interfaces and can be used to determine the zeta potential of two different surfaces at the same time under the same conditions (Nishimura et al., 1992). These authors used the plane interface technique to simultaneously study silica plates and muscovite mica basal planes. The zeta potential values for silica presented in Figure 15.3 indicate that the asymmetric silica–mica cell provides results comparable to those of the symmetrical silica–silica cell.

Zeta potentials of clay minerals have also been determined using a flat plate streaming potential apparatus for muscovite mica (Scales et al., 1990), saponite, and hectorite (Nishimura et al., 2002a). Zeta potential of clays becomes less negative with increasing electrolyte concentration as a result of double layer compression (Figure 15.4). The resulting charge reduction causes clay flocculation. Streaming potential measurements of sodium montmorillonite using a flexible wall permeameter were
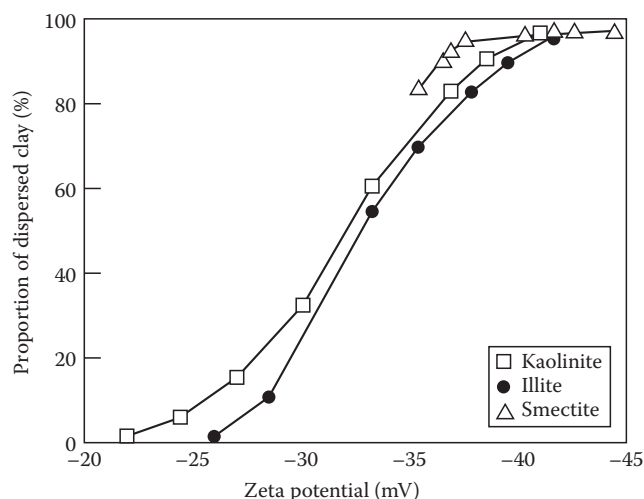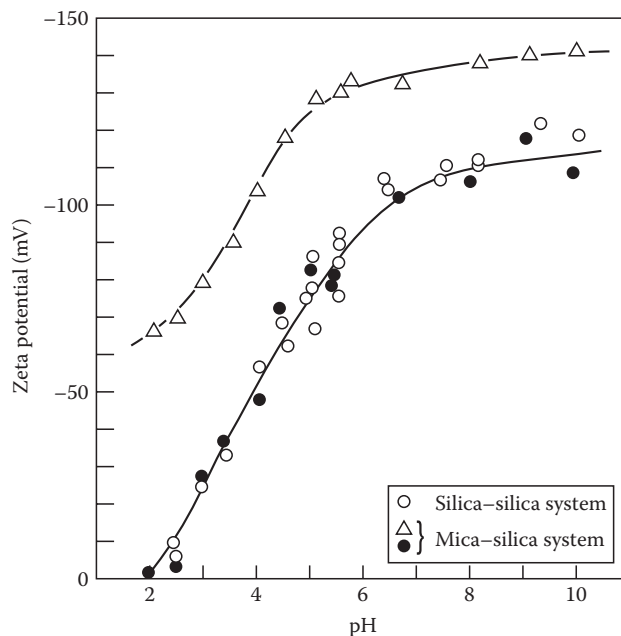
FIGURE 15.3 Zeta potentials as a function of solution pH for muscovite mica basal plane– and silica plate–aqueous solution interfaces obtained using electroosmosis. (Reprinted from Nishimura, S., H. Tateyama, K. Tsunematsu, and K. Jinnai. 1992. Zeta potential measurement of muscovite mica basal plane-aqueous solution interface by means of plane interface technique. J. Colloid Interface Sci. 152:359–367. With permission of Academic Press, Inc.)

found to be dependent on the salt concentration of the permeating solution (Heister et al., 2005).

A vastly different application of electrokinetic experiments is the application to dewatering and decontamination of soils
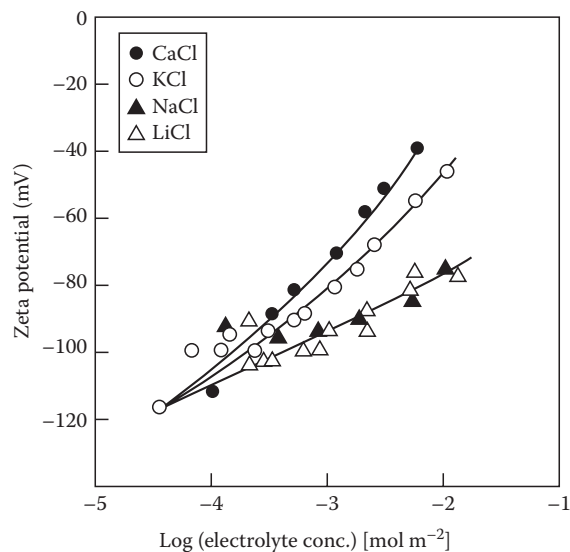
FIGURE 15.4 Relation between zeta potential and electrolyte concentration for various 1:1 electrolytes obtained from streaming potential. (Reprinted with permission from Scales, P.J., F. Grieser, and T.W. Healy. 1990. Electrokinetics of the muscovite mica–aqueous solution interface. Langmuir 6:582–589. Copyright 1990 American Chemical Society.)

and clays. For example, electroosmosis has been used for removing organic contaminants from kaolinite and soil clays (Shapiro and Probstein, 1993; Schultz, 1997; Kim et al., 2001). These authors present applications for in situ hazardous waste remediation of soils. Economic analysis indicated that electro-osmosis compared favorably with the cost of excavation and ex situ treatment (Schultz, 1997).

A potentiometric titration method has been developed to account for changes in solubility of the solid with changes in pH (Schulthess and Sparks, 1986). This is a batch method where the reference for each sample is the supernatant specific to that sample back titrated to pH 7. This method is considered to account for all sources of proton consumption (Schulthess and Sparks, 1986).

15.1.4 Thermodynamics of Colloid Surfaces

To develop the thermodynamic treatment of the surface region, a few definitions are useful. The interfacial region is a space between two adjoining phases (gas–liquid, gas–solid, liquid–liquid, liquid–solid, solid–solid), which is characterized by inhomogeneity in its properties. The Gibbs surface is a mathematical dividing surface, without volume, drawn parallel to the boundaries of the interfacial region, which is used to define the volumes of the two adjoining bulk phases. A schematic of the interfacial region and the Gibbs surface is presented in Figure 15.5. The actual values for the system as a whole will differ from the sum of the values for the bulk phases by an excess or deficiency due to the Gibbs surface (Adamson, 1976). The following relations hold for the variables of state:
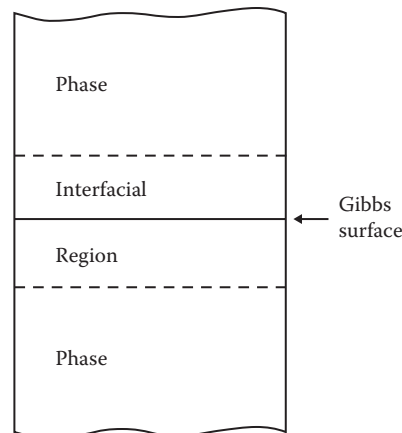
FIGURE 15.5 Representation of the interfacial region and the Gibbs surface.

$$\begin{aligned} \textit{Volume}: &\quad V = V^{\alpha} - V^{\beta} \\ \textit{Internal Energy}: &\quad E = E^{\alpha} + E^{\beta} + E^{\sigma} \\ \textit{Entropy}: &\quad S = S^{\alpha} + S^{\beta} + S^{\sigma} \\ \textit{Moles}: &\quad n_i = n_i^{\alpha} + n_i^{\beta} + n_i^{\sigma} \end{aligned} \qquad (15.20)$$

where
   α and β denote the bulk phases
   σ denotes the Gibbs surface (Adamson, 1976)

Additional variables of state are defined as the surface tension or surface free energy (γ) and the area of the Gibbs surface (A).

The three fundamental thermodynamic relationships of surface chemistry are the Young–Laplace equation, the Kelvin equation, and the Gibbs equation (Adamson, 1976). The Young–Laplace equation is the fundamental equation of capillarity for a curved Gibbs surface:

$$P^{\beta} - P^{\alpha} = \gamma \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \qquad (15.21)$$

where
   $P^{\beta} - P^{\alpha}$ is the capillary pressure
   r_1 and r_2 are the radii of curvature

The Kelvin equation gives the effect of surface curvature on the molar free energy of a substance. The free energy of a substance can be related to its vapor pressure assuming the vapor to be ideal (Adamson, 1976). The Kelvin equation is

$$\ln \left(\frac{P}{P_0} \right) = \frac{\gamma V}{RT} \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \qquad (15.22)$$

where
   P_0 is the normal vapor pressure of the liquid
   P is the vapor pressure observed over the curved surface
   R is the molar gas constant
   T is temperature

For a small, reversible change dE in the energy of the system (Adamson, 1976)

$$dE = dE^\alpha + dE^\beta + dE^\sigma$$

$$= TdS^\alpha + \sum \mu_i dn_i{}^\alpha - P^\alpha dV^\alpha + TdS^\beta$$

$$+ \sum \mu_i dn_i{}^\beta - P^\beta dV^\beta + TdS^\sigma + \sum \mu_i dn_i{}^\sigma - P^\sigma dV^\sigma + \gamma dA$$

$$(15.23)$$

where μ_i is chemical potential. Substituting for dE^α and dE^β and manipulating the equation for dE^σ lead to the expression (Adamson, 1976)

$$S^\alpha dT = Ad\gamma + \sum n_i{}^\sigma d\mu_i = 0 \qquad (15.24)$$

At constant T and A, the Gibbs equation is

$$-d\gamma = \sum \frac{n_i{}^\sigma}{A} d\mu_i = \sum \Gamma_i{}^\sigma d\mu_i \qquad (15.25)$$

where $\Gamma_i{}^\sigma$ is a surface excess concentration per unit area defined as $\Gamma_i{}^\sigma = n_i{}^\sigma/A$. The Gibbs equation can be applied to liquid–liquid and liquid–vapor interfaces where the surface tension can be measured to calculate the surface concentration of the adsorbed species causing the surface tension change. Similarly, if the surface concentration can be measured directly but the surface tension cannot, the Gibbs equation can be used to calculate the lowering of γ from the measured adsorption in solid–gas and solid–liquid systems (Hunter, 1987).

15.2 Interparticle Forces

15.2.1 Electrical Double Layer

Double layer theory describes the distribution of ionic concentrations near electrostatically charged particles. The charge on the colloidal particles is due to isomorphic substitution in the particle lattices or arises from the preferential adsorption of one ionic species from the solution phase (Babcock, 1963). Such a charge requires the presence of a layer of ions of opposite charge. The double layer consists of an excess of ions of opposite sign and a deficiency of ions of the same sign that are electrostatically repelled by the particle. Double layer theory assumes that the surface of the colloidal particles is represented by an infinite flat surface having continuous and uniform electrostatic charge density immersed in an electrolyte with a uniform dielectric constant (Babcock, 1963). All electrolyte ions are assumed to be point charges. The electrical potential, ion charge, and ion distributions can be calculated from the Poisson–Boltzmann equation:

$$\frac{d^2\psi}{dx^2} = -\frac{1}{\varepsilon_0 D} \sum cFz_i \exp\left(-z_i F\psi(x)/RT\right) \qquad (15.26)$$
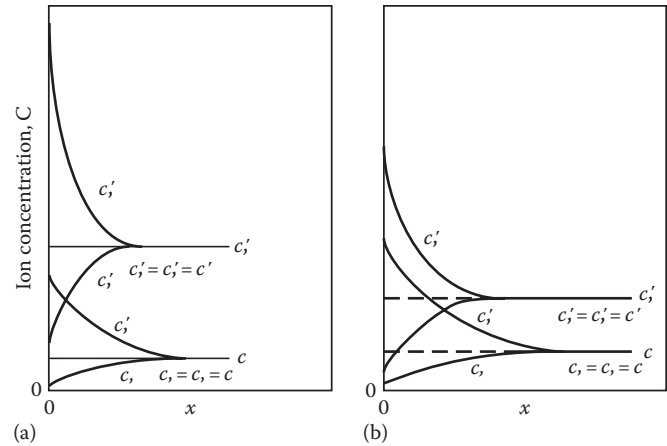
FIGURE 15.6 Distribution of ions in the electric double layer at two electrolyte concentrations ($c' > c$). (a) Constant potential surface. (b) Constant surface charge. (Adapted from van Olphen, H. 1977. An introduction to clay colloid chemistry, 2nd Ed. John Wiley & Sons, New York.)

where

   $\psi(x)$ is the inner potential at a distance x from the surface
   ε_0 is the permittivity of free space
   D is the dielectric constant of water
   c is the concentration
   R is the gas constant
   T is the temperature (K)
   F is the Faraday constant
   z_i is the valence of the charged species (Sposito, 1984)

Figures 15.6 and 15.7 present the ion distribution and the electric potential distribution, respectively, in the double layer at two electrolyte concentrations. The extent of the double layer is given by the distance $(1/\kappa)$ in units of meters:

$$\frac{1}{\kappa} = \sqrt{\frac{\varepsilon_0 RT}{2000 F^2 I}} \qquad (15.27)$$

where I is the ionic strength $\left(= 1/2 \sum c_i z_i^2\right)$. Important findings of diffuse double layer theory are (1) the excess cations near the negative surface neutralize more of the charge than the anion deficit, (2) the electric potential decreases as the electrolyte concentration increases, and (3) the double layer distance $(1/\kappa)$ decreases as the electrolyte concentration increases. Some important limitations of double layer theory are that it applies only to infinitely dilute suspensions and to low surface charge densities (Babcock, 1963).

15.2.2 Attractive Force

The attractive force acting on colloidal particles is called the van der Waals force and acts to bring particles closer together. The basis of the attractive force is that the fluctuating dipole of one atom polarizes another one and the two atoms attract
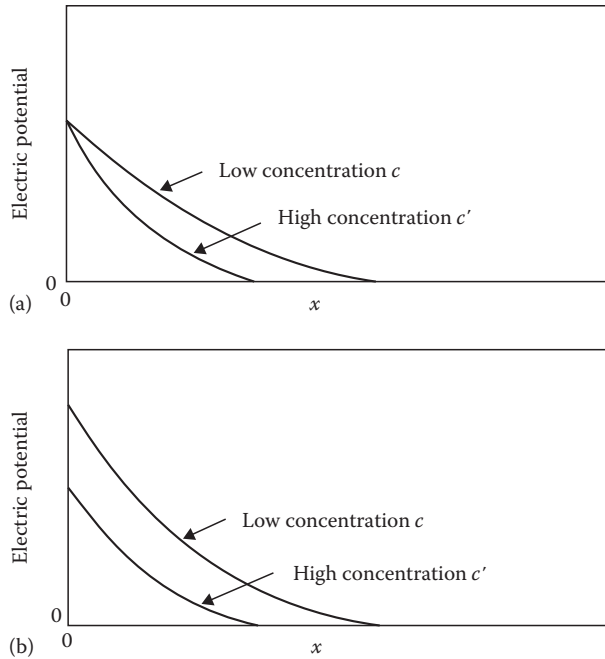
FIGURE 15.7 Electric potential distribution in the electric double layer at two electrolyte concentrations. (a) Constant potential surface. (b) Constant surface charge. (Adapted from van Olphen, H. 1977. An introduction to clay colloid chemistry, 2nd Ed. John Wiley & Sons, New York.)

FIGURE 15.8 van der Waals forces between two mica surfaces in aqueous electrolyte solutions. The measured Hamaker constant is $A = 2.2 \times 10^{-20}$ J. Retarded van der Waals forces are observed above 5 nm. (Reprinted from Israelachvili, J.N. 1992. Intermolecular and surface forces, 2nd Ed. Academic Press, San Diego, CA. With permission of Academic Press Ltd.)

each other. This attraction between atom pairs is additive, and, therefore, the energy of interaction between particles decreases much more slowly with distance than that between individual atoms (Quirk, 1994). The interaction energy per unit area between two opposing planar solids for the van der Waals force (Φ_{vdW}) is (Israelachvili, 1992)

$$\Phi_{vdW} = -\frac{A_H}{12\pi d^2} \qquad (15.28)$$

where
A_H is the Hamaker constant
d is the distance separating the solid surfaces

At distances >5 nm, the correlations between the induced dipole distributions weaken and the interaction energy per unit area corresponds to the retarded van der Waals force (Φ_{RvdW}):

$$\Phi_{RvdW} = -\frac{B_{rH}}{3d^3} \qquad (15.29)$$

where B_{rH} is the retarded Hamaker constant. Figure 15.8 shows values of the van der Waals force obtained experimentally between two mica surfaces. Attractive forces were observed between mica particles in the range of 0.6–2 nm in $CaCl_2$ solution both experimentally and theoretically with statistical mechanics and Monte Carlo simulations (Kjellander et al., 1990).

15.2.3 Repulsive Force

The electrostatic force results from the charge on the colloidal particles and acts to repel them. A force operates on charged surfaces as a result of their interacting double layers. This force is repulsive if the charges on the particles are the same. The repulsion described in terms of interaction energy per unit area (Φ_R) is given by the force times the distance through which it operates (Hiemenz, 1977; Sposito, 1984; Hiemenz and Rajagopalan, 1997):

$$\Phi_R = -\frac{64a^2}{z} cRT \exp(-z\kappa d) \qquad (15.30)$$

where
$a = \tanh(z\psi_0/4RT)$
z is the charge on the electrolyte ions
c is the concentration of the electrolyte ions
κ is the inverse double layer distance
d is half of the surface separation

This equation is valid only when the surface separation $2d \gg 1/\kappa$ and $a \approx 1$. The electrostatic force between two plates for different electrolyte concentrations is presented in Figure 15.9.

When surfaces are brought closer together, an additional repulsive force becomes important. This force is called the solvation force or, when water is the medium, the hydration force. Solvation forces are of short range and oscillatory and arise whenever liquid molecules are induced to order between surfaces. Between colloid surfaces, repulsive hydration forces arise when water molecules strongly bind to hydrogen bonding surface groups such as hydrated ions or hydroxyl groups (Israelachvili, 1992). The effective range of hydration forces in clays and silicas is 3–5 nm. The interaction of mica surfaces in dilute solution obeys double layer theory, but at higher electrolyte concentration, a hydration force
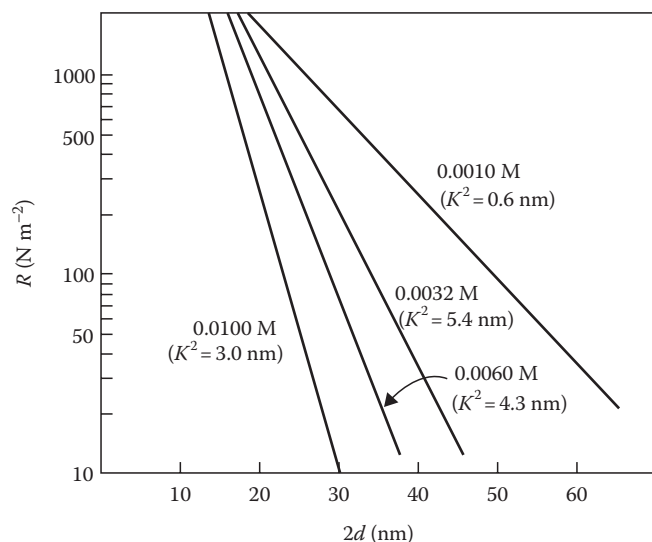
FIGURE 15.9 Repulsive force between two plates for different concentrations of a 1:1 electrolyte where κ is the inverse double layer distance. (Reprinted from Hiemenz, P.C. 1977. Principles of colloid and surface chemistry. Marcel Dekker Inc., New York. With permission of Marcel Dekker, Inc.)

develops due to the energy needed to dehydrate surface-bound cations. The strength of the hydration force, which increases with the hydration number of the cation: $Mg^{2+} > Ca^{2+} > Li^+ \approx Na^+ > K^+ > Cs^+$ (Pashley and Quirk, 1984; Israelachvili, 1992), is illustrated for two mica surfaces in Figure 15.10.
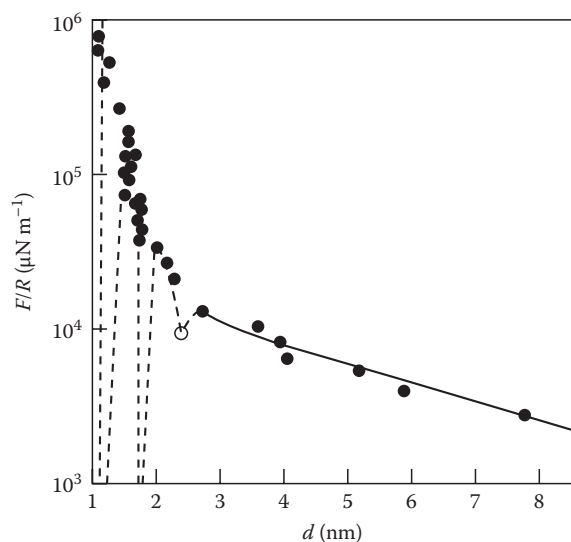
FIGURE 15.10 Forces between two mica surfaces in 7 mmol L^{-1} NaCl electrolyte. *F* is the total force scaled by the radius (*R*) of the curved surfaces. At distances > 1.5 nm, the forces are described by double layer theory. Hydration forces are observed at distances < 1.5 nm. (Reprinted with permission from Ducker, W.A., and R.M. Pashley. 1992. Forces between mica surfaces in the presence of rod-shaped divalent counterions. Langmuir 8:109–112. Copyright 1992 American Chemical Society.)

Clay swelling is the result of double layer repulsion between the surfaces of individual particles. Under confining conditions, a fluid pressure or swelling pressure is created that is a direct measure of the balance of forces between particles (van Olphen, 1977). The swelling pressure is obtained by measuring the confining force that must be applied to keep the clay layers at a given distance. The swelling pressure (Π) is

$$\Pi = P_{vdW} + P_R \tag{15.31}$$

where

P_{vdW} is the van der Waals force
P_R is the electrostatic force (Greathouse et al., 1994)

At short distances, hydration forces become significant in the swelling pressure. At greater distances, measured swelling pressures are of similar magnitude to calculated double layer repulsions (van Olphen, 1977).

Interparticle forces can be measured experimentally using the surface force apparatus (SFA), total internal reflectance microscopy, and the AFM (Israelachvili, 1992). The SFA can measure forces between surfaces at the 10^{-10} m level of resolution. This apparatus has been used to measure attractive van der Waals forces, repulsive double layer forces, and repulsive hydration forces in aqueous solutions (Israelachvili, 1992). Because of its smooth surface and ease of handling, mica has been the primary solid used in SFA studies. Total internal reflection microscopy has been used to study forces between a surface and an individual colloidal particle. AFM has been used to measure both short- and long-range forces (Ducker et al., 1991; Nishimura et al., 2002b; Zhao et al., 2008). Interactions between charged mica surfaces have been investigated using a combination of SFA and AFM experiments. Results from both methods agree with theoretical predictions (Kékicheff et al., 1993).

15.3 Colloidal Stability

15.3.1 Flocculation and Dispersion

The International Union of Pure and Applied Chemistry defines flocculation as "a process of contact and adhesion whereby the particles of a dispersion form larger-size clusters," other terms used interchangeably with flocculation are agglomeration and coagulation (IUPAC, 2009). The stability of colloidal suspensions is a balance between repulsive and attractive forces acting among the suspended particles. If net repulsive forces predominate, particles do not coagulate and remain dispersed. When the attractive forces are dominant, interacting particles coagulate and the resulting floccules settle more rapidly from the suspension than the smaller dispersed particles. Different theories in the literature attempt to describe colloid behavior (Sogami and Ise, 1984; Smalley, 1990; Ise and Smalley, 1994; McBride and Baveye, 2002). A classical continuum scale theory of the forces between particles is the DLVO theory named after Derjaguin, Landau, Verwey, and Overbeek, largely responsible

for its development (Hunter, 1987). Through the use of the SFA, it has been shown that the DLVO theory is an oversimplification of the forces acting at the mineral surface. For example, DLVO theory does not take into account short-range solvation forces (Israelachvili, 1992).

Continuum DLVO theory provides a simplified assumption of a phyllosilicate surface with permanent negative charge. Conceptually, the negative charge is developed from isomorphous substitution of lower valent cations for higher valent cations in either the tetrahedral or octahedral layers of the crystal. With regard to soils, a further complication exists in terms of mineral/organic interaction. Mounting evidence suggests that the primary reactive phases in soils are complex mineral assemblages where phyllosilicates are coated by natural organic matter and Fe-oxyhydroxide phases (Bertsch and Seaman, 1999). In this case, the theoretical models are too simplified to provide a good representation of natural colloids. Advanced computer molecular simulations must be used to begin to understand the complexities of mineral/fluid and mineral/mineral interaction, even for clean systems, for surface force behavior on the colloidal scale (Hsu, 1999). Consequently, DLVO theory, acknowledging its simplifications, will be used as a frame of reference for the discussion of flocculation dispersion.

Flocculation is a thermodynamically favorable process; however, the kinetics of coagulation determine the stability of colloidal suspensions. Generally, all colloidal suspensions will spontaneously flocculate given sufficient time, but potential energy barriers retard the rate of flocculation. These barriers are analogous to activation energies considered in chemical kinetics (Hiemenz and Rajagopalan, 1997). Particles in a primary minimum are adhesive and are not readily separated. In contrast, the dispersion/flocculation transition is the result of a secondary minimum, involves card-house type structures (Figure 15.11), and is readily reversible (Quirk, 1994).

There are many examples in the literature showing that DLVO theory can account for the observed kinetic behavior of dispersed colloidal systems (e.g., Napper and Hunter, 1974). However, this theory must be applied with caution since it treats ions exclusively as point charges ignoring their surface chemical

FIGURE 15.11 Representation of card-house structure. (Reprinted from Hunter, R.J. 1987. *Foundations of colloid science*, Vol. 1. Oxford University Press, New York by permission of Oxford University Press.)

properties and geometry, and should only be applied to dilute systems (Sogami and Ise, 1984).

The potential barrier preventing particles from coagulating is defined by the stability ratio (W), which is the fraction of the total number of collisions between particles that result in coagulation. The rate of coagulation in the absence of a potential barrier or rapid coagulation (R_f) is limited only by the rate of diffusion of the particles toward one another. When the particles have a potential barrier to overcome, the rate of coagulation (R_s) is slow and is related to R_f by

$$R_s = \frac{R_f}{W} \qquad (15.32)$$

Rapid coagulation occurs when no significant repulsive forces act between particles and van der Waals or long-range coulombic attractions predominate. The quantification of the coagulation rate under these conditions was examined by von Smoluchowski (1916, 1917) and is discussed by Overbeek (1952).

Slow coagulation occurs over distances of the order of 1–100 nm when the approaching particles experience a barrier as their double layers overlap. Diffusion over this distance results from many individual Brownian events, some of which bring the particles closer together and some of which take them further apart. Since the rates of rapid and slow coagulation are directly proportional to the number of particles diffusing in the direction of a central particle (J), it follows from Equation 15.32 that the stability ratio is given by

$$W = \frac{R_f}{R_s} = \frac{J_f}{J_s} = 2r \int_{2r}^{\infty} \exp\left(\frac{E_T}{K_B T}\right) \frac{dr}{d^2} \qquad (15.33)$$

where
  r is the particle radius
  $d \cong 2r$

Verwey and Overbeek (1948) showed that W was determined almost entirely by the value of the total potential energy (E_T) at the maximum (Figure 15.12). In Figure 15.12 E_R and E_A are the potential energies due to repulsive and attractive forces. A more complete analysis of the kinetics of colloid flocculation is presented in Hunter (1987).

Gravity removes suspended particles by sedimentation while inducing differential sedimentation coagulation, thereby decreasing Brownian coagulation rates. Ultimately, sedimentation limits the time of a flocculation series test. The test should be long enough to detect relative changes in suspended particle numbers, but not so long that all dispersed particles settle from a stable suspension (Hesterberg and Page, 1990a).

The reverse of flocculation is called dispersion. Ideally, the amount of energy required to separate two particles coagulated into a potential energy minimum is approximately equal to the difference between the interaction energy at the minimum and that at the adjacent maximum (van Olphen, 1977).
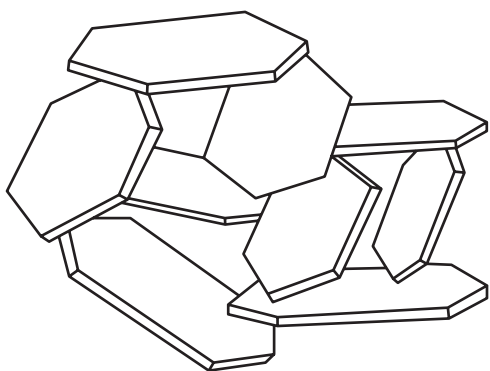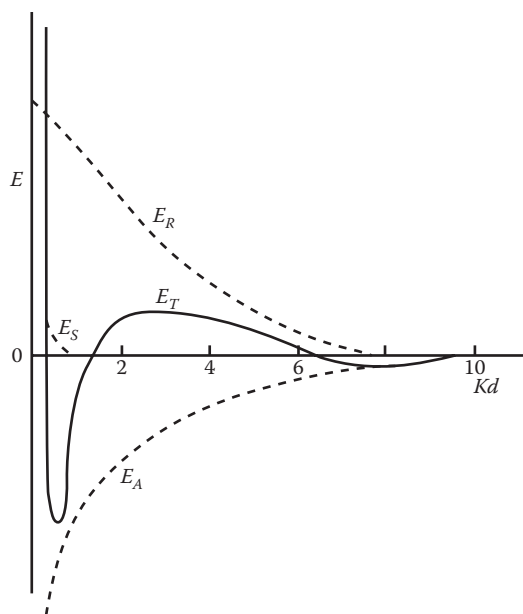
FIGURE 15.12 Total potential energy of interaction of two colloidal particles. $E_T = E_S + E_R + E_A$, where E_S is the potential energy of repulsion due to the solvent layers, E_R is the potential energy due to the repulsive forces, E_A is the potential energy due to the attractive forces, and κd is particle separation. (Reprinted from Hunter, R.J. 1987. Foundations of colloid science, Vol. 1. Oxford University Press, New York by permission of Oxford University Press.)

Under constant chemical conditions, separation of particles could presumably be induced by an input of kinetic energy (e.g., thermal or mechanical shear). Alternatively, changing the chemical conditions of the bulk solution surrounding coagulated particles could provide chemical and/or electrochemical energy to produce dispersion.

15.3.1.1 Modes of Particle Association

15.3.1.1.1 Edge–Edge, Edge–Face, Face–Face

Clay crystals have a net negative surface charge as a consequence of isomorphic substitutions of electropositive ions for ions with a lower valence. This negative charge generates an ionic reorganization in the solution medium that has been described above as the diffuse double layer. Clay particles also have edge surfaces with atomic structure different from the faces. At the edge of the platelet, the tetrahedral layer of Si and the octahedral layer of Al exhibit broken bonds, which, in turn, generate another electric double layer.

Double layer theory assumes that the surfaces of the clay minerals are of semi-infinite spatial extent and show no edge effects, a simplification that is not always satisfactory. Clay mineral particles have finite dimensions. Below the p.z.c., edge surfaces carry a positive charge due to specific adsorption of protons. Using the Poisson–Boltzmann equation, Secor and Radke (1985) calculated the effect of edge–face corners on the electrical potential distribution around an idealized, symmetrical montmorillonite disk. Spillover of the negative electrical potential from the particle faces into the edge region can result in a negative potential everywhere around the particle.

The electrical potential at the edge surface strongly depends on the electrolyte concentration and the ratio of the face to edge charge density. The extent of spillover also is a weak function of the particle shape (Secor and Radke, 1985). This study implies that attraction between positively charged edges and negatively charged faces of phyllosilicate mineral particles will depend on the extent of the edge protonation, electrolyte concentration, and shape of the particle. A phyllosilicate structure that is collapsed in the *c*-dimension, such as mica, should be able to acquire a larger edge surface charge density than a layer silicate like smectite where structural expansion increases the distance between edge surface aluminol groups (Hesterberg, 1988).

When there is a reduction in the thickness of the double layer, particles can associate among themselves in three different ways: face–face, face–edge, or edge–edge. Face–face association is also called parallel aggregation and does not produce floccules, while the other two associations do produce three-dimensional structures called card houses (see Figure 15.11).

In concentrated suspensions of clay, the edge–edge and edge–face associations form a continuum, with chains of particles in the card-house structures mentioned above. The rigidity of the gel depends on the number and strength of the bonds in the continuum structure. Some attempts to characterize the gel structure using freeze-drying techniques have been made (Norrish and Rausell-Colom, 1961). When the water is eliminated from the suspension, the volume of the system does not change and the final product is a dry clay structure with some strength that has been called aerogel (van Olphen, 1977).

15.3.1.1.2 Domains and Quasicrystals

Some colloidal systems may, under certain circumstances, show a reversible clustering among particles. The earliest reported example is the Fe-hydroxide sol described by Cotton and Mouton (1907). The term tactoid was first used by Freundlich (1932) and was more precisely defined by Overbeek (1952) as the association of particles at a certain distance affected by electrolyte concentration and pH. Quirk and Aylmore (1971) proposed the term quasicrystal to describe the regions of parallel alignment of individual aluminosilicate lamellae in montmorillonite and the term domain to describe the regions of parallel alignment of crystals for illite. This terminology is adopted in the present chapter.

The distribution model of adsorbed ions in mixed mono- and divalent systems for smectite quasicrystals was described by Shainberg and Otoh (1968) and Bar-on et al. (1970). According to this theory, called the ion demixing model, when Na is added to Ca saturated montmorillonite, most of the adsorbed Na will concentrate on the external surfaces of the quasicrystals until 10% of the adsorbed Ca has been replaced by Na. Initially, the size and shape of the particles are not altered by the addition of adsorbed Na. As further Na is added to the system, Na penetrates into the quasicrystals and brings about disintegration of the clay packets (Bar-on et al., 1970).

Lebron and Suarez (1992a) used electrophoretic mobility experiments to show that the demixing model can be applied to micaceous clays. The electrophoretic mobility of micaceous

TABLE 15.5 Critical Coagulation Concentrations of Phyllosilicates under Various Conditions

Mineral	CCC (mol$_c$ L^{-1})	Background Electrolyte	pH	Solids Concentration (g kg^{-1})
Kaolinite-4	2–40	NaNO$_3$	4–10	0.025
Kaolinite-9	8, 30	NaHCO$_3$, Na$_2$CO$_3$	8.3, 9.5	0.6–0.9
Kaolinite (Georgia)	5, 245, 75	NaCl, NaHCO$_3$, Na$_2$CO$_3$	7, 8.3, 9.5	0.6–0.9
Kaolinite (KGa-1)	<0.19–54.6	NaCl	5.8–9.1	0.67
Kaolinite-4	0.1–0.3	Ca(NO$_3$)$_2$	4–10	0.025
Kaolinite (KGa-1)	<0.19–0.85	CaCl$_2$	5.5–9.3	0.67
Montmorillonite-23	1–10	NaNO$_3$	3.8–10	0.25
Montmorillonite-23	20, 48, 68	NaCl, NaHCO$_3$, Na$_2$CO$_3$	7, 8.3, 9.5	0.6–0.9
Montmorillonite-27	14, 47, 17	NaCl, NaHCO$_3$, Na$_2$CO$_3$	7, 8.3, 9.5	0.6–0.9
Montmorillonite (SAz-1)	14–28	NaCl	6.4–9.4	0.67
Montmorillonite (SAz-1)	1.09, 1.56	CaCl$_2$	6.1, 7.6	0.67
Montmorillonite (SAz-1)	0.93, 2.02, 0.88	MgCl$_2$	6.1, 8.4, 9.0	0.67
Vermiculite	38, 58, 30	NaCl, NaHCO$_3$, Na$_2$CO$_3$	7, 8.3, 9.5	0.6–0.9
Illite-36	9, 185, 95	NaCl, NaHCO$_3$, Na$_2$CO$_3$	7, 8.3, 9.5	0.6–0.9
Illite (Grundy)	7.24	NaCl	~6	1
Illite (Grundy)	0.2	CaCl$_2$	~6	1

Source: Adapted from Hesterberg, D.L. 1988. Critical coagulation concentrations and rheological properties of illite. Ph.D. Thesis. University of California, Riverside, CA by Oxford University Press.

domains increased when the sodium adsorption ratio (SAR) increased from 5 to approximately 10.

15.3.1.2 Critical Coagulation Concentration

The CCC, sometimes called the critical flocculation concentration (CFC), is the minimum concentration of indifferent electrolyte that induces rapid coagulation. The CCC is strongly dependent on counterion valence (Table 15.5). This observation is known as the Schulze–Hardy rule. An estimate of the CCC (mol L^{-1}) is given by (Hunter, 1987)

$$CCC = \frac{0.107\varepsilon^3 (k_B T)^5 Z^4}{N_A A_H^2 (ze)^6} \tag{15.34}$$

where
  $Z = \tanh(ze\psi_0/4k_B T)$
  N_A is Avogadro's number
  A_H is the Hamaker constant
  ε is the relative dielectric constant
  k_B is the Boltzmann constant
  e is the proton charge

The variation of CCC with counterion valence depends approximately on the inverse sixth power of z. At 25°C in water, using the experimental observation that coagulation usually occurs between low-potential surfaces, the result is

$$CCC \propto \frac{\psi_0^4}{z^2} \tag{15.35}$$

where now the CCC depends on the inverse of the square of the valence. However, the CCC calculated with Equation 15.35

agrees well with experimental results if $\psi_0 \propto 1/z$. When calculating the CCC in soils, one must consider that soils consist of mixtures of permanent and variable charge minerals. The net surface charge, and consequently, the electrical potential around the mixture of particles are dependent upon variables such as pH, specifically adsorbed ions, ionic strength, and mineralogy. A more detailed evaluation of the effect of these variables on the stability of colloids is presented below.

15.3.2 Factors Affecting Colloidal Stability

15.3.2.1 Solution Composition

The magnitude of the repulsion barrier is determined by the nature of the material adsorbed on the particle surface. In the case of a charged colloid, repulsion depends on the magnitude of the surface charge and on the extent of the electrical double layer, which, in turn, depends on the total electrolyte concentration. It is necessary here to distinguish between the concentration of the potential determining ions and that of other ions that have no direct interaction with the surface. If the surface potential of the particles is determined by the concentration of potential determining ions, the magnitude of this potential is not affected by the addition of an indifferent electrolyte. For this type of double layer, when salt concentration increases, the double layer thickness decreases, the surface charge of the particles increases, and the surface potential remains constant (Figure 15.7a).

If the surface charge of the particle is determined by isomorphic substitution, the surface charge does not change with increasing electrolyte concentration. The diffuse double layer compresses, but in this case, the surface potential decreases with increasing electrolyte concentration (Figure 15.7b) (van Olphen, 1977).

15.3.2.2 Exchange Complex Composition

As explained above, clay particles are surrounded by cations as a consequence of the net negative electrical charge on the surface. Cations bonded to the surface can be exchanged for other cations in solution. Consequently, the cations on the exchange complex are dependent on the solution composition.

There is an equilibrium between the cations on the exchange complex and the cations in the solution. Not all cations are adsorbed with the same affinity. Cations with larger hydrated radii are less strongly adsorbed than those with smaller hydrated radii. In solutions with equal initial concentration of different cations, the amounts of Ca and Mg adsorbed are several times greater than the amount of Na adsorbed. In general, polyvalent cations are adsorbed more strongly than monovalent cations and are not easily displaced by other cations. The order of adsorption strength is Al > Ca > Mg > H > K > Na (Duchaufour, 1970).

The flocculation/dispersion behavior of soil colloids depends on salt concentration, exchange complex cation, cation valence, and dominant clay mineralogy. In general, divalent cations are more effective in flocculating colloids than monovalent cations. For example, Quirk and Schofield (1955) found that the flocculating power of $CaCl_2$ is 50–100 times higher than that of NaCl. For monovalent cations, there is also a difference, with KCl showing greater flocculation power than NaCl (Pashley, 1981).

15.3.2.3 pH

The pH is an important determinant of the electrical potential of the clay surface. Changes in pH affect the edge charge on clays and the surface charge of variable charge minerals such as Fe and Al oxides. There is considerable variability depending on structural composition and degree of crystallinity, but Fe and Al oxides generally undergo a surface charge reversal around pH 7–9 (positively charged below that pH and negatively charged above). This is also the region in which kaolinite exhibits its edge charge reversal as evidenced by Cl^- adsorption studies (Schofield and Samson, 1954). Soil colloids consist of a mixture of minerals, each with a different p.z.c. At low pH, edge to face bonding, as well as bonding of positive Fe and Al oxides to negative clay surfaces, is expected to occur (van Olphen, 1977; Kretzschmar et al., 1993, 1997). This type of bonding should hinder dispersion and should thus result in flocculation. With increasing pH, as the p.z.c. is approached, edge to face clay bonding decreases and Fe and Al oxide bonding to clays is also expected to decrease (Suarez et al., 1984). In variable charge systems, flocculation is at a maximum at the p.z.c.

CCC increased at high SAR values with increasing pH for three soil clays whose clay mineralogy was dominantly kaolinite, montmorillonite, or illite (Goldberg and Forster, 1990). Hesterberg and Page (1990b) also found an increase in CCC for a Na- and a K-illite with increasing pH. Similar results were found for illite and three micaceous soil clays when SAR > 15 (Lebron and Suarez, 1992a, 1992b). The electrophoretic mobility of these materials increased when the SAR was greater than 20 and the pH was above the p.z.c. No pH effect was observed at SAR < 15 for either mobility or CCC.
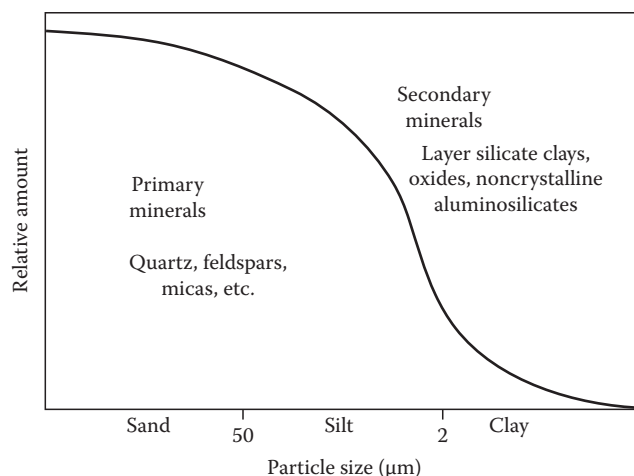
FIGURE 15.13 Typical abundance of primary and secondary minerals in different size fractions of the soil. (Reprinted from McBride, M.B. 1994. Environmental chemistry of soils. Oxford University Press, New York by permission of Oxford University Press.)

15.3.2.4 Mineralogy

The colloidal fraction of a soil consists primarily of secondary minerals (Figure 15.13). Layer silicate minerals differ in chemical composition and charge characteristics leading to different physicochemical behavior (Table 15.5). The stoichiometry of each mineral varies due to isomorphic substitutions in the crystal lattice during the formation or evolution of the mineral structure.

The siloxane (Si–O–Si) surfaces of 2:1 layer silicates without structural charge are hydrophobic. Therefore, the surface oxygens coordinated to Si show little tendency to hydrogen bond with water molecules. Smectites, such as montmorillonite, form weak hydrogen bonds between structural charge located mainly in the octahedral sheet and water. Such bonds are the result of the delocalization of some structural charge into the surface oxygens (Farmer, 1978). Smectites, such as beidellite, with a high proportion of charge in the tetrahedral sheet, form stronger hydrogen bonds. The siloxane surfaces of vermiculite are the most hydrophilic of the 2:1 layer silicate clays because they possess a large tetrahedral charge partially distributed onto surface oxygens (Farmer, 1978). Tetrahedral charge is much more localized on fewer surface oxygens than octahedral charge, explaining the stronger hydrogen bonding of adsorbed water on vermiculite (McBride, 1989). An extensive study of the structural characteristics of soil minerals can be found in Dixon and Weed (1989).

An example of how these mineral differences affect the flocculation/dispersion behavior of soil colloids is shown in Table 15.5. An overview of the data reveals that reported CCC values of kaolinite, montmorillonite, vermiculite, and illite are quite variable within and between these mineral groups.

Many of the differences in Table 15.5 are due to differences in methodology in the determination of CCC and/or differences in the stoichiometry of the silicate minerals. However, these factors do not account for all of the variability. Lebron and Suarez

(1992a) found substantial differences in CCC within samples from the same soil type. Differences in content of organic matter and other minerals can drastically change the behavior of soil colloids. Consequently, general guidelines for reclamation of agricultural land or the use of amendments to maintain colloids in a flocculated state must be implemented with caution because soils usually require higher electrolyte concentrations than the corresponding pure clay minerals to maintain a flocculated condition.

15.3.2.5 Organic Matter

Organic matter constitutes a small portion of the soil mass (0.5%–10%) but is intimately associated with inorganic particles and plays an important role in the improvement of soil structure (Nelson and Oades, 1998). Aeration, water-holding capacity, and permeability increase with increasing soil SOM content. However, adsorbed organic matter can promote dispersion of soil particles. Organic coatings, under certain conditions, maintain a dispersed state for soil colloids in suspension through a combination of electrostatic and steric mechanisms (Stevenson, 1982). Table 15.6 shows the spatial extensions for nonionic polymer molecules of different molecular weights. Like electrical double layers, macromolecules of at least a few thousand molecular weight also extend in space over distances comparable to, or greater than, the van der Waals attraction. In general, it has been shown that organic matter coatings modify the surface properties of minerals, increasing their CEC, generating hydrophobic and hydrophilic surfaces (Hunter, 1987), and significantly altering the point of zero net charge (p.z.n.c., pH at which the CEC and AEC are equal) (Heil and Sposito, 1993; Kaplan et al., 1997; Kretzschmar et al., 1997; Bertsch and Seaman, 1999).

Goldberg and Forster (1990) and Kretzschmar et al. (1993) found that the removal of SOM enhanced soil flocculation (decreased the CCC). Similarly, addition of small amounts of organic material substantially increased dispersion of Na-saturated soil or clay in the order: humic acid > soil polysaccharide ≥ anionic polysaccharide (Gu and Doner, 1993; Kretzschmar et al., 1997). Using smectite, kaolinite, and three soils whose clay fractions were dominated by one of these minerals, Frenkel et al. (1992) showed that the CCC values of Na-soils were much higher, and much more affected by organic matter than those of Ca-soils.

TABLE 15.6 Spatial Extensions of Electrical Double Layers and Polymers of Different Molecular Weight

1:1 Electrolyte Concentration (mol L^{-1})	Double-Layer Thickness, $1/\kappa$ (nm)	Polymer Molecular Weight	Spatial Extension (nm)
10^{-5}	100	1,000,000	60
10^{-4}	30	100,000	20
10^{-3}	10	10,000	6
10^{-2}	3	1000	2
10^{-1}	1		

Source: Reprinted from Hunter, R.J., *Foundations of Colloid Science*, Vol. 1, 1987 by permission of Oxford University Press.
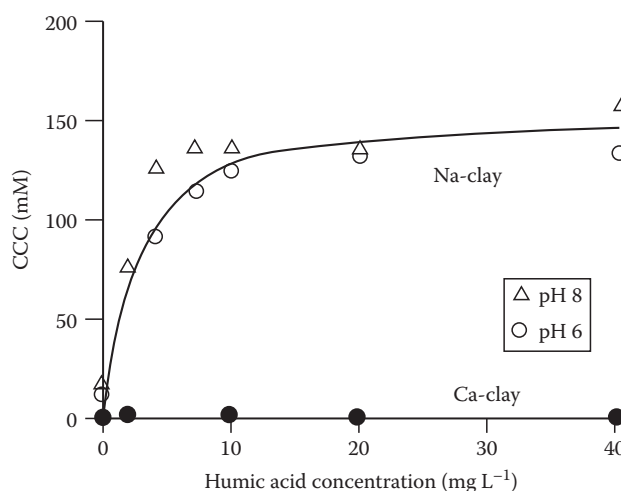
FIGURE 15.14 Effect of the humic acid concentration on the CCC of a Na-clay and a Ca-clay. (After Tarchitzky, J., Y. Chen, and A. Banin. 1993. Humic substances and pH effects on sodium and calcium-montmorillonite flocculation and dispersion. Soil Sci. Soc. Am. J. 57:367–372.)

Tarchitzky et al. (1993) showed similar comparisons between Na- and Ca-montmorillonite suspensions with varying additions of humic and fulvic acids (Figure 15.14). The effect of organic matter on stability of soil colloids is a function of its size. Large organic materials such as polysaccharides and hyphae act to bind colloid particles together. Small organic molecules such as fulvic acid and organic acids increase dispersion of soil colloids through their effect on particle charge. A more detailed review of organic matter chemistry and its effect on soils is provided in Stevenson (1982) and Nelson and Oades (1998).

15.3.3 Measurement of Colloidal Stability

15.3.3.1 Flocculation Series Test

CCC are commonly determined using the flocculation series test. The experiment can be performed by taking a series of test tubes containing the same concentration of the colloid and adding varying amounts of the coagulating electrolyte. The tubes are then shaken and allowed to stand for a given time. If the electrolyte range has been properly chosen, the CCC is defined as the concentration above which the settling material leaves behind a perfectly clear supernatant solution. Below the CCC, the supernatant retains some of the uncoagulated colloid. The concentration of colloidal particles at which the flocculation test must be performed should not exceed 10 g L^{-1}, thus avoiding interferences from different particle interactions such as gel formation.

15.3.3.2 Dispersion Indices

There are several methods (qualitative, semiquantitative, and quantitative) to determine the dispersion or flocculation status of soil colloids. Qualitative analyses of the dispersion state are those based on direct observation of small particles when the soil is immersed in water. This observation can be made with the optical microscope or with the naked eye, as is the case for

the test of Emerson (1967). Methods based on turbidity of a suspension of dispersed soil can be considered semiquantitative when comparative measurements are made. Normally, a standard curve is constructed using known amounts of dispersed clay. The soil under evaluation is assigned a dispersion value by comparison with the standard. The most commonly used quantitative method to determine soil flocculation state is the dispersion index. This index is the ratio of the amount of soil colloids in water to the amount of soil colloids in solution when the soil has been treated with a dispersant. This procedure is recommended by the Soil Conservation Service (Sherard et al., 1977). Different variations regarding particle size, dispersing agent, and manipulation of samples are found in the methods proposed by El-Swaify et al. (1970) and Dong et al. (1983), among others.

15.3.3.3 Bingham Yield Stress

Rheology is the study of the flow and deformation of colloidal systems under the influence of mechanical forces (van Olphen, 1977). A Newtonian liquid when confined between two parallel plates moves at a constant shear rate ($\dot{\gamma}$) proportional to the applied shear stress (τ):

$$\tau = \eta\dot{\gamma} \qquad (15.36)$$

Non-Newtonian fluids obey different relations between shear stress and shear rate. Plastic flow is flow that occurs only above a certain finite stress (τ_0) called the yield stress. Ideal plastic flow exhibits a linear relationship between shear rate and shear stress over all rates of shear. Many colloidal dispersions exhibit Bingham flow, which is characterized by the equation

$$\tau = \tau_B - \eta_\Delta\dot{\gamma} \qquad (15.37)$$

where
  τ_B is the Bingham yield stress found by extrapolating Equation 15.37 to zero shear rate
  η_Δ is the differential or plastic viscosity

The differential viscosity is the derivative of shear stress with respect to shear rate at a given shear rate (van Olphen, 1977). Shear rate versus shear stress relationships for Newtonian flow, ideal plastic flow, and Bingham flow are presented in Figure 15.15.

  Bingham yield stress is a measure of the degree of coagulation of a colloidal suspension and the mode of particle interaction. Bingham yield stress is a function of both the number of particle–particle linkages in the coagulated structure and the energy required to break these linkages (Rand and Melton, 1977). A stable clay dispersion exhibits ideal plastic flow, while flocculated suspensions exhibit Bingham flow as can be seen in Figure 15.16 for kaolinite. Differential viscosity can also be used to assess the extent of particle flocculation, but it is a much less sensitive parameter than Bingham yield stress (Heath and Tadros, 1983).

  Rheological studies have been carried out on kaolinite (Flegman et al., 1969; Rand and Melton, 1977; Yong et al., 1979; Diz and Rand, 1989; Ma and Pierre, 1999), illite (Yong et al., 1979;
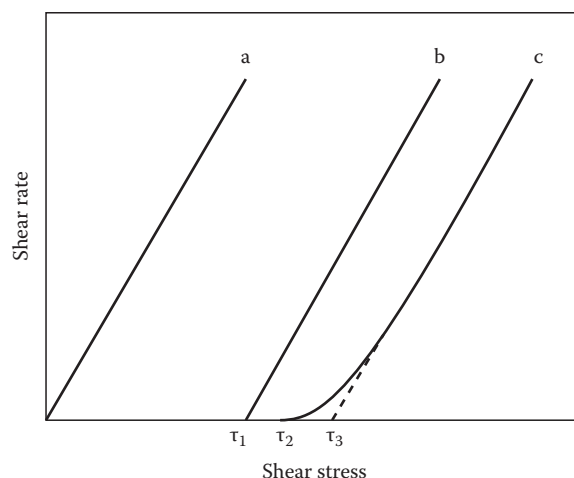
FIGURE 15.15 Shear rate versus shear stress relationships. Curve a represents Newtonian flow, curve b represents ideal plastic flow, and curve c represents Bingham flow. (Adapted from van Olphen, H. 1977. An introduction to clay colloid chemistry, 2nd Ed. John Wiley & Sons, New York.)
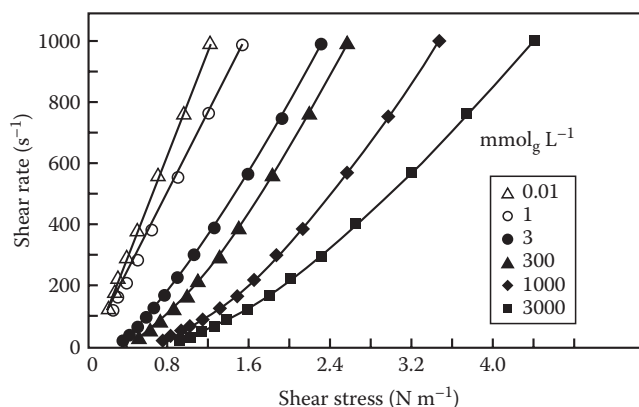
FIGURE 15.16 Shear rate versus shear stress for Na-kaolinite. The open symbols represent dispersed systems and ideal plastic flow. The closed symbols represent coagulated systems and Bingham flow. (Adapted from Yong, R.N., A.J. Sadh, H.P. Ludwig, and M.A. Jorgensen. 1979. Interparticle action and rheology of dispersive clays. J. Geotech. Eng. Div. 105:1193–1209.)

Ohtsubo et al., 1991; Hesterberg and Page, 1993), montmorillonite (Rand et al., 1980; Heath and Tadros, 1983; Keren, 1988, 1989a; Tombácz et al., 1989; Miano and Rabaioli, 1994; Benna et al., 1999; Duran et al., 2000), laponite (Laxton and Berg, 2006), clay mixtures (Yong et al., 1979; Keren, 1989b, 1991), soil clays (Zhao et al., 1991), and soils (Ghezzehei and Or, 2001; Markgraf et al., 2006). A series of curves of Bingham yield stress as a function of pH obtained at successively increasing electrolyte concentration should coincide at one point. This point is characteristic of the pH value of the edge p.z.c. of the mineral (Rand and Melton, 1977). Edge p.z.c. values have been determined in this manner for various kaolinites and range in value from pH 5.6 to 8.8 (Diz and Rand, 1989). Kaolinite particles occur in edge–face associations below the edge p.z.c. and in edge–edge associations around the edge p.z.c. (Rand and Melton, 1977).

Montmorillonite exhibits no edge–face associations over the pH range 4–11; coagulation produced by electrolyte additions is initially the result of edge–edge interactions with face–face interactions occurring at high-electrolyte concentrations (Rand et al., 1980). No edge p.z.c. could be determined on montmorillonite using this method. This is likely because the edge area is small and attraction between edges and faces is small compared to repulsion between faces (Rand et al., 1980). From yield stress and electrophoretic mobility measurements determined without electrolyte addition, Benna et al. (1999) suggested that the p.z.c. of three bentonites occurred in the pH range of 7.5–8.5.

In distilled water, Ca-montmorillonite exhibited Newtonian flow. With increasing exchangeable sodium percentage (ESP), differential viscosity, deviation from Newtonian flow, and Bingham yield stress of montmorillonite all increased (Keren, 1988). These increases are likely due to the increased number of particles in solution as tactoids break down. Differential viscosity and Bingham yield stress decreased with increasing electrostatic charge density of smectites (Keren, 1989a). This result is attributed to the reduced swelling of higher charge density smectites. Kaolinite exhibited Newtonian flow at all ESPs. The introduction of even 5% montmorillonite into the kaolinite systems resulted in deviations from Newtonian flow and increased differential viscosity (Keren, 1989b). Bingham yield stress and deviations from Newtonian flow of kaolinite–montmorillonite mixtures also increased with ESP (Keren, 1991). Bingham yield stress of kaolinite, Ca-montmorillonite, Na-montmorillonite, and two soils increased exponentially with decreasing water content (Ghezzehei and Or, 2001).

15.4 Colloid Transport

As previously discussed, colloidal materials found in soils include phyllosilicate clays, Al, Fe, and other metal oxides, carbonate minerals, microorganisms, and other biological debris, much of which falls into the "nanoparticle" size range (see Section 15.1.2). The fundamental processes involved in colloid migration in the soil environment include mobilization, transport, and deposition, representing primarily a balance between the two opposing processes of colloid mobilization and deposition. The study of colloid transport in the subsurface environment has received considerable attention in recent years because of concern that mobile colloids may enhance the mobility of strongly sorbing contaminants (i.e., facilitated transport) and alter the hydraulic properties of soil and aquifer materials. In addition, the migration of biocolloids such as bacteria and viruses has important implications to in situ bioremediation, the protection of drinking water supplies, and the spread of disease (Khilar and Fogler, 1984; McDowell-Boyer et al., 1986; Kia et al., 1987; Ryan and Elimelech, 1996; Kretzschmar et al., 1999; de Jonge et al., 2004b; McCarthy and McKay, 2004; Tufenkji, 2007). However, our ability to predict colloid movement and deposition is often confounded by the complexities of surface interactions in such heterogeneous natural systems. Understanding colloid transport requires consideration not only of the chemical and biological processes and reactions but also of the physical principles of filtration and deposition in porous media.

Several mechanisms have been identified by which colloidal materials can be mobilized in the soil environment (Figure 15.17) including clay dispersion resulting from changes in pore-water chemistry that increase electrostatic repulsion between the colloid and collector surfaces, water film expansion and air–water interface scouring during episodic wetting events (imbibition), raindrop impact at the soil surface, and increased shear forces associated with transient hydrodynamic events (i.e., precipitation, groundwater pumping, subsurface recharge injection). Mobile colloids can also be formed in situ through precipitation reactions resulting from changes in pore-water chemistry. Clearly, colloid mobilization is associated with a physical and/or chemical perturbation sufficient to overcome the energy barrier limiting particle detachment (McCarthy and Degueldre, 1993; Ryan and Elimelech, 1996; DeNovio et al., 2004).

Compared to saturated aquifer systems, chemical and physical perturbations are common within the soil environment, with the presence of air providing an additional system interface for colloid partitioning (Wan and Wilson, 1994; Sirivithayapakorn and Keller, 2003). The presence of a dynamic air–water interface and fluctuating chemical and hydrological conditions complicate our understanding of colloid transport processes in the soil and vadose environments (Bradford and Torkzaban, 2008). Preferential flow paths can increase the mobility of colloidal material in close proximity to the pathway, while limiting the reactivity of colloidal sorbents isolated from the flow path (DeNovio et al., 2004; McCarthy and McKay, 2004).

The energy barrier limiting colloid detachment is the difference between the interaction minimum and the maximum energy potential described in Figure 15.12. In terms of chemical perturbations, soil colloid mobilization can result from changes in pore-water pH, ionic strength, and/or Na^+/Ca^{2+} ratios (McDowell-Boyer, 1992; Ryan and Gschwend, 1994; Gamerdinger and Kaplan, 2001; Bunn et al., 2002; Khaleel et al., 2002; McCarthy et al., 2002; Grolimund and Borkovec, 2006), alteration of surface charge resulting from a chemical dispersing agent or dissolved organic matter (Ryan and Gschwend, 1994; Seaman and Bertsch, 2000; Johnson et al., 2001), or dissolution of carbonate or Fe cementing agents resulting in the release and transport of silicate clays (Gschwend et al., 1990; Ryan and Gschwend, 1990; Ronen et al., 1992; Swartz and Gschwend, 1998). Furthermore, potentially mobile colloids may precipitate in situ due to altered chemical gradients (Gschwend and Reynolds, 1987; Mashal et al., 2004). In many instances, more than one of the mechanisms may be operative in order to trigger significant colloid mobilization, such as a change in pore-water chemistry that reduces particle attraction combined with an increase in pore-water velocity and shear forces.

Under ideal laminar flow conditions, the hydrodynamic shear force acting on a spherical particle attached to a flat surface can be expressed as follows:

(a)

(b)
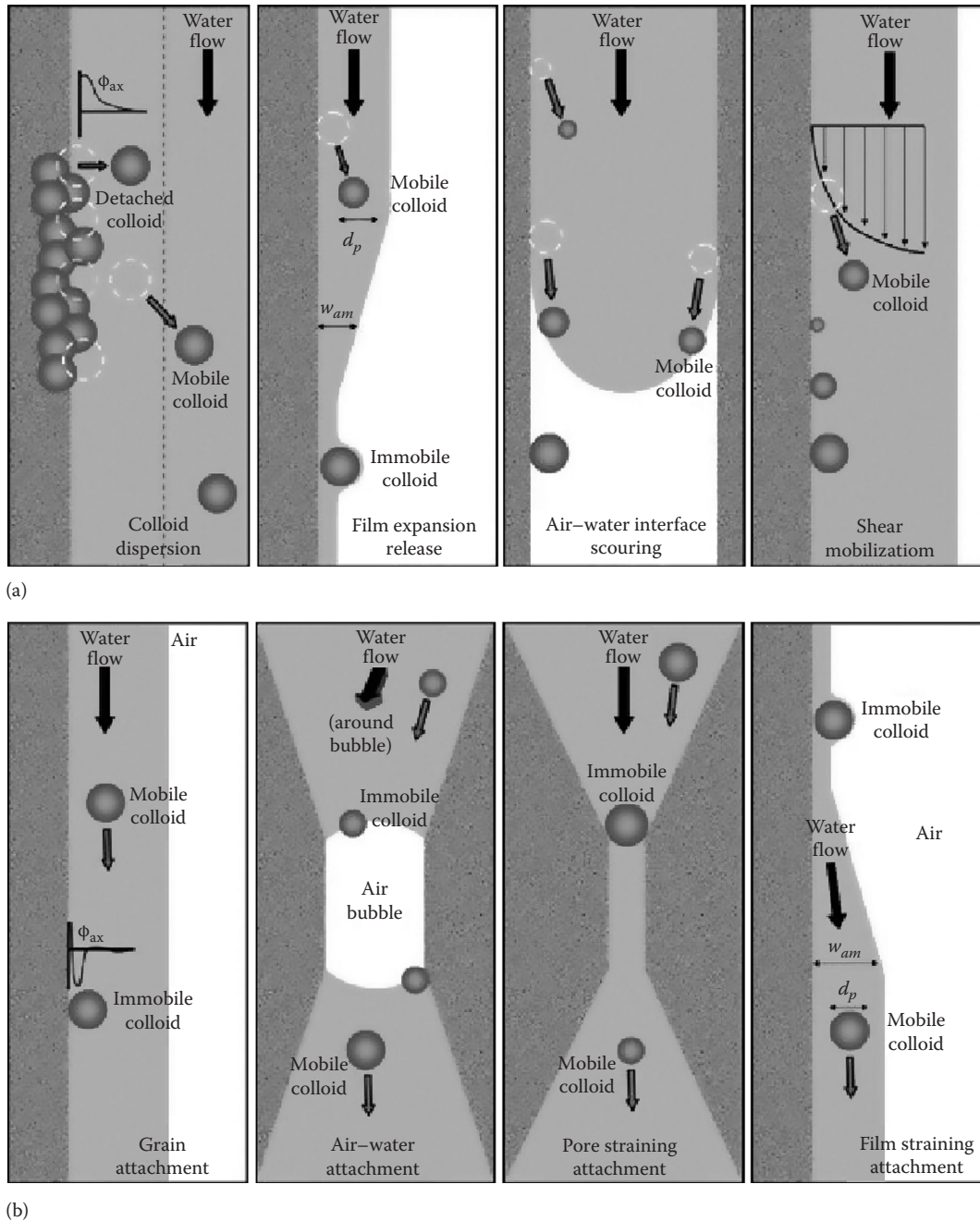
FIGURE 15.17 Colloid mobilization (a) mechanisms include dispersion due to chemical perturbation, expansion of water film during imbibition, air–water interface scouring and shear mobilization. Colloid deposition (b) mechanisms include attachment associated with physicochemical filtration, attachment to the air–water interface, physical pore straining, and water film straining. (Adapted from DeNovio, N.M., J.E. Saiers, and J.N. Ryan. 2004. Colloid movement in unsaturated porous media: Recent advances and future directions. Vadose Zone J. 3:338–351.)

$$F_{shear} = (1.7)6\pi\mu rV \qquad (15.38)$$

where

   V is the flow velocity at the center of the particle
   μ is the fluid viscosity
   r is the particle radius (Sharma et al., 1992; Ryan and Elimelech, 1996; Ryan et al., 1998; DeNovio et al., 2004)

Adhesive forces that oppose shear stress are subject to changes in solution chemistry. The critical velocity for particle release decreases with increasing particle size, even though the required hydrodynamic force increases for non-Brownian particles, that is, >1 μm in diameter (Sharma et al., 1992). As one might expect, the amount of particles mobilized should generally increase with increasing flow velocity, but greater flow velocities are required to mobilize smaller particles that extend to a lesser degree into

the advective stream. However, many assumptions inherent in applying such a physical approach to soil and aquifer systems are likely in question. In more realistic systems, interrelated factors such as contact area, adhesion strength, and surface charge heterogeneity are important in countering shear forces (Sharma et al., 1992).

In addition to the chemical factors affecting colloid stability, transport of colloids is related to soil physical factors such as pore size distribution, geometry, and continuity (Torkzaban et al., 2008; Bradford et al., 2009). Coarse-textured soils have a larger distribution of pore sizes (with larger pores) than fine-textured soils, suggesting greater potential for colloid movement. This is not generalizable since some fine-textured materials experience cracking and formation of continuous structural macropores that provide a preferential pathway for colloid transport.

15.4.1 Colloid Transport Modeling

Physical and chemical processes affect the transport of soil colloids. Colloid transport can be represented by the physical processes of molecular diffusion (Brownian movement), advective flow, and gravitational forces. Molecular diffusion, the random motion of particles caused by thermal effects, is related to temperature and viscosity. Deposition is the process whereby moving particles are attached to stationary surfaces based on the balance of attractive and repulsive forces, that is, attractive van der Waals forces and the combined influence of attractive and repulsive electrostatic forces, as described by DLVO theory of colloid stability (see Sections 15.2.2, 15.2.3, and 15.3.1).

The attachment process involves two steps: transport to the collector surface and attachment that binds the two surfaces together. The transport step reflects a combination of diffusion, convection, and gravity, with electrostatic and van der Waals forces controlling colloid binding. Models used in describing colloid transport are typically extensions of the advection dispersion equation (ADE) for describing solute movement in a homogeneous saturated porous medium:

$$\frac{\partial C}{\partial t} + \frac{\rho_b}{\theta}\frac{\partial S}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - \nu\frac{\partial C}{\partial x} \qquad (15.39)$$

where
  C is the aqueous phase concentration for the solute or suspended colloidal material of interest
  x is distance
  t is time
  D is the hydrodynamic dispersion coefficient
  ν is the pore-water velocity
  S is the sorbed concentration of colloids or solutes
  ρ_b is the matrix bulk density
  θ is the volumetric water content (Grolimund and Borkovec, 2001; Bradford et al., 2003; Tufenkji, 2007)

Hydrodynamic dispersion and dispersivity appear to decrease with increasing particle size (Sinton et al., 2000; Auset and Keller, 2004). As the colloidal material moves through porous medium, it may be removed from solution by various physicochemical filtration processes, with various expressions, such as the equilibrium partitioning (K_{eq}), used to describe reversible colloid attachment processes in a manner analogous to the equilibrium partitioning of solutes. Using a numerical solution to the ADE, equilibrium partitioning can be modified to reflect any one of a variety of phenomenological equilibrium partitioning expressions, that is, Freundlich, Langmuir, etc.

In contrast to equilibrium partitioning, colloid attachment is more often viewed as a kinetically controlled process with terms accounting for both attachment (k_{att}) and detachment (k_{det}) rates:

$$\frac{\rho_b}{\theta}\frac{\partial S}{\partial t} = k_{att}C - \frac{\rho_b}{\theta}k_{det}S \qquad (15.40)$$

yielding an expression that is mathematically equivalent to the reversible first-order kinetic equation for partitioning solutes. Ripening and blocking are terms that are used to describe an increase or decrease in colloid attachment efficiency (i.e., k_{att} and k_{det}) with time or distance.

Figure 15.18 illustrates colloid breakthrough behavior for irreversible (Figure 15.18a) and reversible (Figure 15.18b) attachment kinetics. Colloid attachment is commonly assumed to be irreversible (i.e., $k_{det} = 0$) as described by clean-bed colloid filtration theory (CFT) (O'Melia and Tiller, 1993; Bradford et al., 2003; Simunek et al., 2006; Tufenkji, 2007). For irreversible attachment (i.e., CFT), the kinetic rate coefficient determines the plateau level of colloid breakthrough, while more complex breakthrough behavior is observed when both attachment and detachment occur at differing rates. When $k_{att} = k_{det}$, colloid behavior becomes analogous to that of a conservative tracer, that is, $R = 1$. Colloid retention profiles for irreversible attachment (Figure 15.18c) remain unchanged after the inlet colloid source is removed and an exponential decrease in retained colloids is observed with transport distance. Again, a more complex colloid retention profile is observed for reversible attachment that depends on both the forward and reverse rate coefficients and the treatment duration (Figure 15.18d).

The irreversible colloid attachment/depositional rate constant for a clean-bed filter is

$$k_{att} = \frac{3(1-\theta)\nu}{2d_c}\eta_0\alpha \qquad (15.41)$$

where
  d_c is the average grain size of the matrix
  η_0 is the collector contact efficiency for the three transport mechanisms (diffusion, gravity, convection)
  α is the attachment efficiency (Yao et al., 1971; Bradford et al., 2003; Li et al., 2004)
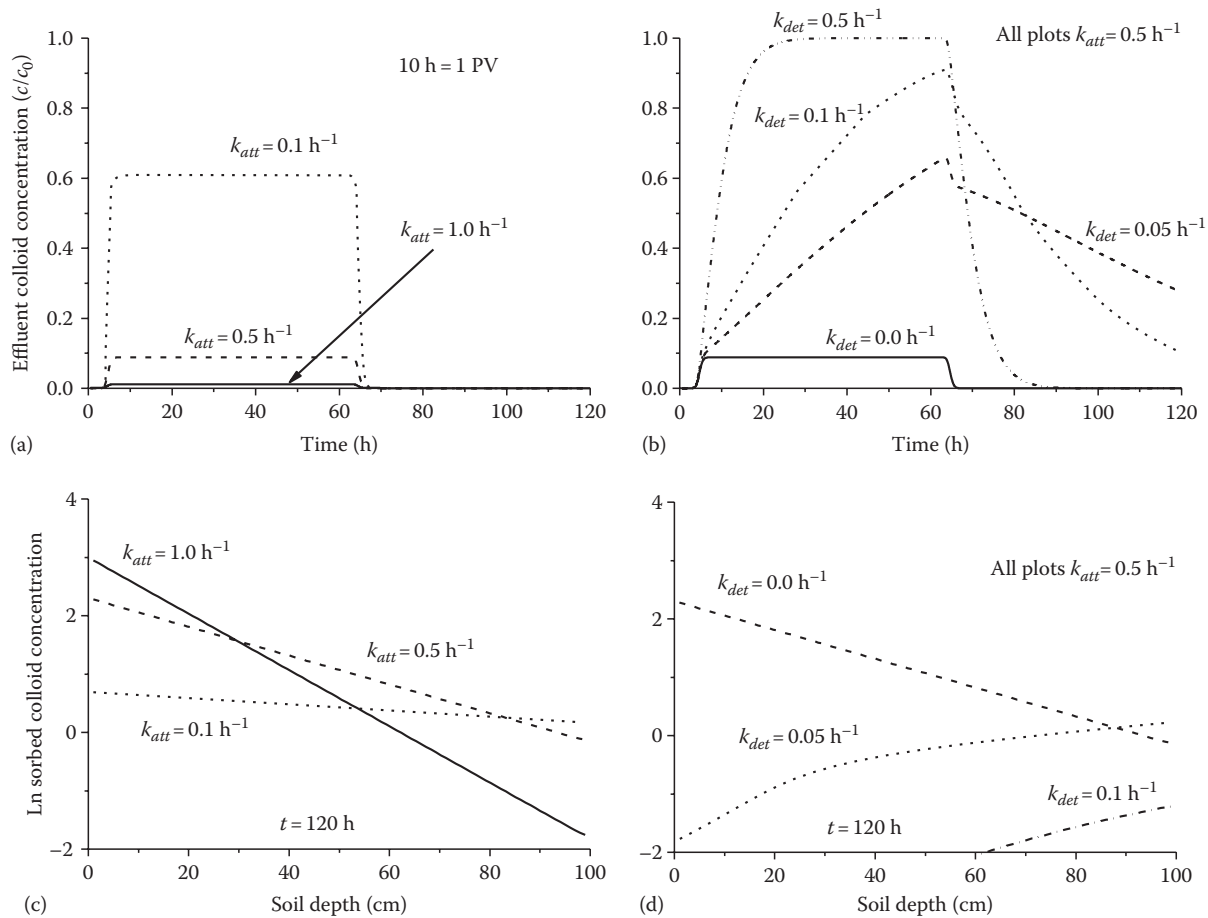
FIGURE 15.18 Simulated colloid breakthrough for a 1 m column with a saturated flow velocity of 2.4 m day^{-1} ($\theta = 0.5$, DL = 0.01 m) exposed to a 60 h colloid pulse with kinetically controlled attachment (a = clean-bed filtration theory), and kinetically controlled attachment and detachment (b). Retained colloid profile for irreversible (c), and reversible attachment (d) after 60 h colloid pulse followed by 60 h of colloid-free inlet solution.

The collector efficiency reflects the colloid collisions with the stationary matrix resulting from diffusion, advective flow, and gravitational forces, and can be calculated (Tufenkji and Elimelech, 2004a). The attachment efficiency term (α) is generally determined experimentally by fitting colloid breakthrough results to the CFT model based on k_{att} and the calculated η (O'Melia and Tiller, 1993; Elimelech et al., 2000; Bradford et al., 2004; Tufenkji and Elimelech, 2004a; Tufenkji, 2007). Additional mechanisms of material loss, such as cell motility, growth, and inactivation, must also be considered when evaluating the transport of microbes. Furthermore, research in the last decade has demonstrated the importance of surface biomolecules, which can vary in response to the local environment, and among microbial strains and species in controlling their adhesion properties (Bolster et al., 2000; Walker et al., 2004; Tufenkji, 2007).

Studies using model colloidal systems indicate that CFT is generally applicable under conditions favoring particle attachment, that is, no significant electrostatic repulsion. However, the CFT approach fails in describing experimental data under apparently unfavorable conditions when significant electrostatic repulsion exists between the mobile colloid and matrix

grain surface. For particles with the same charge, a deep primary attractive force exists at very short separation distances, known as the primary minimum (Figure 15.12). However, electrostatic repulsion supposedly limits particle approach in a manner similar to chemical activation energies. Such discrepancies in deposition studies, and in batch aggregation experiments, have been attributed to porous media heterogeneity that can be difficult to quantify experimentally (i.e., grain size, morphology, and surface charge) and a secondary energy minimum that favors colloid sorption without the necessity to overcome the full electrostatic repulsion required to enter the primary minimum (Figure 15.12), resulting in an apparent change in the attachment rate coefficient (k_{att}) with transport distance (Roy and Dzombak, 1996; Elimelech et al., 2000; Chen et al., 2001; Bradford et al., 2002, 2003; Tufenkji et al., 2003; Hahn and O'Melia, 2004; Li et al., 2004; Redman et al., 2004; Tufenkji and Elimelech, 2005; Tufenkji, 2007). The secondary minimum reflects the exponential decay in electrostatic forces with separation distance, in contrast to van der Waals forces that decrease more slowly with distance. The inability to remobilize particles trapped within the "secondary" minimum by altering solution chemistry, that is, reducing ionic strength, further demonstrates the hydrodynamic

and physicochemical complexity of colloid retention mechanisms (Bradford and Torkzaban, 2008; Torkzaban et al., 2008; Bradford et al., 2009).

Verifying the spatial distribution of retained colloids based on the attachment parameter (α_{BTC}) derived from effluent breakthrough data is a good method for testing the validity of the CFT model. In many instances, the slope of the relationship is greater than CFT predictions, indicating higher initial colloid retention that decreases with increasing transport distance (Li et al., 2004; Tufenkji, 2007). Such discrepancies have been addressed through the use of multiple deposition rate constants, that is, a fast and a slow rate, or a defined statistical distribution of randomly assigned particle deposition rates that reflect the inherent variability in colloidal surface potentials (Li et al., 2004; Tufenkji and Elimelech, 2004b; Tufenkji, 2007). It is interesting to note that the statistical distribution in observed deposition rate coefficients appears to narrow as colloid deposition becomes more favorable, for example, reduced flow velocities and increased ionic strengths (Li et al., 2004).

Attachment kinetics are highly sensitive to colloid–collector interaction potentials. Moderate changes with respect to colloid surface potentials can result in relatively large changes in collision efficiency and attachment rate (Elimelech and O'Melia, 1990; Ryan and Elimelech, 1996; Elimelech et al., 2000; Li et al., 2004). Deviation from the breakthrough plateau with time is indicative of a change in the colloid deposition efficiency during the course of injection, that is, filter ripening or blocking. In addition, surface modifiers such as Fe oxides and soil organics, often present in relatively minor amounts, can alter the expression of electrostatic repulsion on soil mineral surfaces (Goldberg and Forster, 1990; Kretzschmar et al., 1993; Kaplan et al., 1993, 1997; Bertsch and Seaman, 1999; Franchi and O'Melia, 2003; Seaman et al., 1995, 2003).

Recent studies indicate that pore structure and particle straining may also be more significant than previously recognized for intermediate colloid and matrix grain sizes. Colloid straining, trapping colloids in pore regions that are too small to permit passage, is generally considered to be an irreversible process. Within a transport model, straining can be addressed by including an additional first-order expression to the deposition equation given above. In the absence of colloid detachment, the kinetic deposition model accounting for colloid straining becomes

$$\frac{\rho_b}{\theta}\frac{\partial S}{\partial t} = k_{att}C - k_{str}\psi_{str}C \tag{15.42}$$

where
> k_{str} is the straining rate constant
> ψ_{str} is the depth-dependent colloid straining function (Bradford et al., 2002, 2003, 2004)

In contrast to straining, traditionally considered a physical phenomenon insensitive to solution chemistry, recent hydrodynamic modeling efforts have verified particle capture within hydrodynamically disconnected regions of the porous matrix near two collector particle surfaces where the magnitude of fluid velocity is quite small compared to the bulk solution, analogous to the "immobile" regions evident in solute transport behavior. Within the immobile regions, greater colloid/colloid and colloid/collector surface interactions lead to enhanced colloid immobilization in a manner that depends on both solution chemistry and pore geometry. Furthermore, particles weakly retained in the secondary minimum can be subsequently funneled to the immobile capture regions by hydrodynamic forces that are insufficient to induce full colloid detachment (Bradford and Torkzaban, 2008; Torkzaban et al., 2008; Bradford et al., 2009).

15.4.2 Colloid-Mediated Contaminant Transport

Numerous studies have established that organic and inorganic chemicals, which are highly sorbed onto soil nonetheless, are often transported in the subsurface to a greater degree than expected. Such anomalies have been attributed to kinetic effects associated with nonequilibrium sorption and bypass flow, as well as contaminant transport associated with a mobile colloidal fraction that enhances apparent migration beyond the levels controlled by thermodynamically based solubilities or partitioning reactions. Furthermore, colloidal materials may move at rates that are similar or faster than those of nonreactive solute tracers because of physical size or electrostatic exclusion from a fraction of the saturated pore space (Simunek et al., 2006). In general, colloid-facilitated transport is only significant when the contaminant is strongly sorbed to the mobile colloidal phase, approaching irreversible partitioning. Otherwise, the sorbed contaminant will desorb in response to the decrease in the aqueous phase contaminant concentration when the colloidal phase travels faster than the aqueous phase (Grolimund and Borkovec, 2005).

Models that include colloid-facilitated transport generally differ in the manner in which they account for the various colloid release and attachment processes described in Section 15.4.1 (i.e., reversible vs. irreversible attachment, kinetic vs. equilibrium) (Mills et al., 1991; Dunnivant et al., 1992; Corapcioglu and Jiang, 1993; Grolimund and Borkovec, 2005) with additional equilibrium or kinetic expressions used to account for contaminant–colloid interactions. Only recently have models been developed that address partially saturated conditions, colloid straining, and size exclusion (Simunek et al., 2006). The development of such a contaminant transport model, while requiring numerous simplifying assumptions (van de Weerd and Leijnse, 1997; Simunek et al., 2006), still results in a significant list of parameters required to account for both contaminant sorption and colloid attachment. For example, transport and mass balance expressions must account for the aqueous phase contaminant, contaminant sorbed to the immobile matrix, contaminant sorbed to the mobile colloids, and contaminant sorbed to colloids immobilized by various mechanisms, that is, stationary phase attachment, air–water interphase attachment, and straining. Additional expressions are required to account for colloid mass balance as well, illustrating the potential level of model

complexity. Simunek et al. (2006) recommend the use of sensitivity analysis to discern the relative importance of such variables in accurately describing contaminant fate and transport.

Although subject to numerous sampling artifacts, significant concentrations of mobile colloids have been observed in soil and groundwater systems (Ronen et al., 1992; McCarthy and Degueldre, 1993; DeNovio et al., 2004). However, contaminant transport models generally fail to account for the impact of mobile colloids in facilitating the transport of strongly sorbing contaminants. For example, colloid-facilitated transport has been implicated in the migration of organic contaminants (e.g., atrazine, prochloraz, and dichlorodiphenyltrichloroethane [DDT]) (Vinten et al., 1983; Seta and Karathanasis, 1996; Seta and Karathanasis, 1997; de Jonge et al., 1998; Roy and Dzombak, 1998), metals (e.g., Pb, Ni, Cu, and Zn) (Kaplan et al., 1994b, 1995; Grolimund et al., 1996; Roy and Dzombak, 1997; Karathanasis, 1999; Grolimund and Borkovec, 2005; Karathanasis et al., 2005), radionuclides (e.g., ^{137}Cs) (Von Gunten et al., 1988; Kaplan et al., 1994a; Kersting et al., 1999; Flury et al., 2002), and sparingly soluble nutrients, such as phosphate (de Jonge et al., 2004b). Some of the general issues concerning contaminant migration via colloid-facilitated transport are discussed by McDowell-Boyer et al. (1986), McCarthy and Zachara (1989), Ryan and Elimelech (1996), and Kretzschmar et al. (1999).

15.4.3 Effect of Colloid Transport on Hydraulic Conductivity and Soil Formation

Some degree of colloid movement is generally considered a major factor in soil profile development. However, dispersion and transport of soil colloids have been linked to soil crust formation, reduced infiltration and hydraulic conductivity, and erosion (Shainberg et al., 1981; Miller and Baharuddin, 1986; Miller and Radcliffe, 1992). Reductions in hydraulic conductivity caused by colloid transport can be divided into two groups: formation of crusts and migration and deposition within the medium. Crust formation results when the colloids are not able to enter the soil media and form a compact layer above the soil. Alternatively, the particles may enter the medium, flow with the water in a suspension, and be deposited within the medium. This process may result from either mechanical filtration of large colloids or physicochemical processes of attraction/repulsion of small colloids as discussed above (see Section 15.4.1).

Research in soil science related to the effects of colloid transport on hydraulic conductivity has focused primarily on description of the chemical process affecting clay movement, rather than on mathematical expressions for prediction of flow rates. Conditions of low ionic strength, high exchangeable Na, and elevated pH were related to both dispersion of clays (Suarez et al., 1984; Goldberg and Forster, 1990; Miller et al., 1990) and subsequent migration and reduction in soil hydraulic conductivity (Suarez et al., 1984). In highly weathered acid soils where positively charged colloids are important, Seaman et al. (1995, 1997) determined that addition of $CaCl_2$ initially enhanced colloid mobility, likely due to release of exchangeable Al and a decrease in pH.

Soil profile development is often affected by colloid movement. Among these processes are the formation of impermeable clay layers in the subsurface and movement of Fe and Al, most probably as organic metal complexes. Moderately to strongly developed soils are generally characterized by depleted clay contents in the A horizon and larger amounts of clay generally in the upper portion of the B horizon. A substantial portion of the argillic B horizon is related to migration of clays from the upper portion of the profile and subsequent deposition (Birkeland, 1974). Flocculation in the lower part of the profile is enhanced by increased electrolyte concentration relative to the surface horizons. Clay colloid migration is evident by the presence of clay films over ped surfaces and inside voids and presence of oriented clay particles. This process is also observed in aridisols where low organic matter and elevated exchangeable Na enhance colloid transport. Clay deposition is enhanced by increasing electrolyte concentration and removal of water by evapotranspiration. In general, formation of argillic horizons requires that the soil be at least partially dry during some part of the season.

Formation of Fe- and Al-rich layers by translocation (podzolization) is probably caused by transport of metal–fulvic acid complexes, rather than movement of the dissolved metals. The spodic B horizon marks the location at which these chelates either flocculate due to increases in electrolyte concentration or by decomposition of the organic matter (Birkeland, 1974).

References

Adamson, A.W. 1976. Physical chemistry of surfaces, 3rd Ed. John Wiley & Sons, New York.

Appel, C., L.Q. Ma, R.D. Rhue, and E. Kennelley. 2003. Point of zero charge determination in soils and minerals via traditional methods and detection of electroacoustic mobility. Geoderma 113:77–93.

Aringhieri, R., G. Pardini, M. Gispert, and A. Sole. 1992. Testing a simple methylene blue method for surface area estimation in soils. Agrochimica 36:224–232.

Auset, M., and A.A. Keller. 2004. Pore-scale processes that control dispersion of colloids in saturated porous media. Water Resour. Res. 40:W03503, 1–11.

Babcock, K.L. 1963. Theory of chemical properties of soil colloidal systems at equilibrium. Hilgardia 34:417–542.

Babick, F., F. Hinze, and S. Ripperger. 2000. Dependence of ultrasonic attenuation on the material properties. Colloids Surf. A 172:33–46.

Babick, F., and S. Ripperger. 2002. Information content of acoustic attenuation spectra. Part. Part. Syst. Charact. 19:176–185.

Balnois, E., G. Papastavrou, and K.J. Wilkinson. 2007. Force microscopy and force measurements of environmental colloids, p. 405–467. In K.J. Wilkinson and J.R. Lead (eds.) Environmental colloids and particles: Behavior, structure and characterization. IUPAC series on analytical and physical chemistry of environmental systems. John Wiley & Sons, Chichester, U.K.

Bar-on, P., I. Shainberg, and I. Michaeli. 1970. Electrophoretic mobility of montmorillonite particles saturated with Na/Ca ions. J. Colloid Interface Sci. 33:471–472.

Baumann, T., and C.J. Werth. 2004. Visualization and modeling of polystyrol colloid transport in a silicon micromodel. Vadose Zone J. 3:434–443.

Beckett, R. 1991. Field-flow fractionation-ICP-MS: A powerful new analytical tool for characterizing macromolecules and particles. Atomic Spectrosc. 12:215–246.

Beckett, R., and B.T. Hart. 1993. Use of field-flow fractionation techniques to characterize aquatic particles, colloids and macromolecules, p. 165–205. *In* J. Buffle and H.P. van Leeuwen (eds.) Environmental particles, Vol. 2. Lewis Publishers, Ann Arbor, MI.

Beckett, R., D. Murphy, S. Tadjiki, D.J. Chittleborough, J.C. Giddings. 1997. Determination of thickness, aspect ratio and size distribution for platey particles using sedimentation field-flow fractionation and electron microscopy. Colloids Surf. A 120:17–26.

Benna, M., N. Kbir-Ariguib, A. Magnin, and F. Bergaya. 1999. Effect of pH on rheological properties of purified sodium bentonite suspensions. J. Colloid Interface Sci. 218:442–455.

Bertsch, P.M., and J.C. Seaman. 1999. Characterization of complex minerals assemblages: Implications for contaminant transport and environmental remediation. PNAS 96:3350–3357.

Bertuzzo, E., S. Azaele, A. Maritan, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. 2008. On the space-time evolution of a cholera epidemic. Water Resour. Res. W01424, doi: 10.1029/2007WR006211.

Birkeland, P.W. 1974. Pedology, weathering, and geomorphological research. Oxford University Press, New York.

Blo, G., C. Contado, F. Fagioli, M.H. Bollain-Rodriguez, and F. Dondi. 1995. Analysis of kaolin by sedimentation field-flow fractionation and electrothermal atomic absorption spectrometry detection. Chromatographia 41:715–721.

Bohn, H.L., B.L. McNeal, and G.A. O'Connor. 1985. Soil chemistry. John Wiley & Sons, New York.

Bolster, C.H., A.L. Mills, G. Hornberger, and J. Herman. 2000. Effect of intra-population variability on the long-distance transport of bacteria. Ground Water 38:370–375.

Borchardt, G.A. 1977. Montmorillonite and other smectite minerals, p. 293–330. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Borkovec, M., Q. Wu, G. Degovics, P. Laggner, and H. Sticher. 1993. Surface area and size distribution of soil particles. Colloids Surf. A 73:65–76.

Bower, C.A., and J.O. Goertzen. 1959. Surface area of soils by an equilibrium ethylene glycol method. Soil Sci. 87:289–292.

Bradford, S.A., M. Bettahar, J. Simunek, and M.T. van Genuchten. 2004. Straining and attachment of colloids in physically heterogeneous porous media. Vadose Zone J. 3:384–394.

Bradford, S.A., J. Simunek, M. Bettahar, M.T. van Genuchten, and S.R. Yates. 2003. Modeling colloid attachment, straining, and exclusion in saturated porous media. Environ. Sci. Technol. 37:2242–2250.

Bradford, S.A., and S. Torkzaban. 2008. Colloid transport and retention in unsaturated porous media: A review of interface-, collector-, and pore-scale processes and models. Vadose Zone J. 7:667–681.

Bradford, S.A., S. Torkzaban, F. Leij, J. Simunek, and M.T. van Genuchten. 2009. Modeling the coupled effects of pore space geometry and velocity on colloid transport and retention. Water Resour. Res. 45:W02414, doi: 10.1029/2008WR007096.

Bradford, S.A., S.R. Yates, M. Bettahar, and J. Simunek. 2002. Physical factors affecting the transport and fate of colloids in saturated porous media. Water Resour. Res. 38:63/1–63/12.

Brown, G., A.C.D. Newman, J.H. Rayner, and A.H. Weir. 1978. The structure and chemistry of soil clay minerals, p. 29–178. *In* D.G. Greenland and M.H.B. Hayes (eds.) The chemistry of soil constituents. John Wiley & Sons, New York.

Brunauer, S., P.H. Emmett, and E. Teller. 1938. Adsorption of gases in multimolecular layers. J. Am. Chem. Soc. 60:309–319.

Buffle, J., and G.G. Leppard. 1995a. Characterization of aquatic colloids and macromolecules. 1. Structure and behavior of colloidal materials. Environ. Sci. Technol. 29:2169–2175.

Buffle, J., and G.G Leppard. 1995b. Characterization of aquatic colloids and macromolecules. 2. Key role of physical structures on analytical results. Environ. Sci. Technol. 29:2176–2184.

Bunn, R.A., R.D. Magelky, J.N. Ryan, and M. Elimelech. 2002. Mobilization of natural colloids from an iron oxide-coated sand aquifer: Effect of pH and ionic strength. Environ. Sci. Technol. 36:314–322.

Bunville, L.G. 1984. Commercial instrumentation for particle size analysis, p. 1–42. *In* H.G. Barth (ed.) Modern methods of particle size analysis. John Wiley & Sons, New York.

Carter, D.L., M.M. Mortland, and W.D. Kemper. 1986. Specific surface, p. 413–423. *In* A. Klute (ed.) Methods of soil analysis. Part 1—Physical and mineralogical methods, 2nd Ed. SSSA Publ., Madison, WI.

Chanudet, V., and M. Filella. 2006. A non-perturbing scheme for the mineralogical characterization and quantification of inorganic colloids in natural waters. Environ. Sci. Technol. 40:5045–5051.

Chen, J.Y., C. Ko, S. Bhattacharjee, and M. Elimelech. 2001. Role of spatial distribution of porous medium surface charge heterogeneity in colloid transport. Colloids Surf. A 191:3–15.

Chittleborough, D.J., S. Tadjiki, J.F. Ranville, F. Shanks, and R. Beckett. 2004. Soil colloid analysis by field-flow fractionation. Super Soil 2004, 3rd Australian N.Z. Soils Conf., December 5–9, 2004. University of Sydney, Sydney, Australia.

Chorom, M., and P. Rengasamy. 1995. Dispersion and zeta potential of pure clays as related to net particle charge under varying pH, electrolyte concentration and cation type. Eur. J. Soil Sci. 46:657–665.

Coll, H., and L.E. Oppenheimer. 1987. Improved techniques in disc centrifugation, p. 202–214. *In* T. Provder (ed.) Particle size distribution. Assessment and characterization. American Chemical Society, Washington, DC.

Contado, C., G. Blo, F. Fagioli, F. Dondi, and R. Beckett. 1997. Characterisation of river Po particles by sedimentation field-flow fractionation coupled to GFAAS and ICP-MS. Colloids Surf. A 120:47–59.

Corapcioglu, M.Y., and S.Y. Jiang. 1993. Colloid-facilitated groundwater contaminant transport. Water Resour. Res. 29:2215–2226.

Cotton, A., and H. Mouton. 1907. Magneto-optical properties of colloids and heterogeneous liquids. Ann. Chim. Phys. 11:145–203, 289–339.

Crist, J.T., J.F. McCarthy, Y. Zevi, P. Baveye, J.A. Throop, and T.S. Steenhuis. 2004. Pore-scale visualization of colloid transport and retention in partly saturated porous media. Vadose Zone J. 3:444–450.

Davis, J.A., and D.B. Kent. 1990. Surface complexation modeling in aqueous geochemistry. Rev. Mineral. Geochem. 23:177–260.

de Jong, E. 1999. Comparison of three methods of measuring surface area of soils. Can. J. Soil Sci. 79:345–351.

de Jonge, H., O.H. Jacobsen, L.W. de Jonge, and P. Moldrup. 1998. Particle-facilitated transport of prochloraz in undisturbed sandy loam soil columns. J. Environ. Qual. 27:1495–1503.

de Jonge, L.W., C. Kjaergaard, and P. Moldrup. 2004a. Colloids and colloid-facilitated transport of contaminants in soils: An introduction. Vadose Zone J. 3:321–325.

de Jonge, L.W., P. Moldrup, G.H. Rubaek, K. Schelde, and J. Djurhuus. 2004b. Particle leaching and particle-facilitated transport of phosphorus at field scale. Vadose Zone J. 3:462–470.

Delgado, A.V., S. Ahualli, F.J. Arroyo, and F. Carrique. 2005. Dynamic electrophoretic mobility of concentrated suspensions comparison between experimental data and theoretical predictions. Colloids Surf. A 267:95–102.

DeNovio, N.M., J.E. Saiers, and J.N. Ryan. 2004. Colloid movement in unsaturated porous media: Recent advances and future directions. Vadose Zone J. 3:338–351.

Dixon, J.B. 1977. Kaolinite and serpentine group minerals, p. 357–403. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA Inc., Madison, WI.

Dixon, J.B., and S.B. Weed. 1989. Minerals in soil environments. SSSA Inc., Madison, WI.

Diz, H.M.M., and B. Rand. 1989. The variable nature of the isoelectric point of the edge surface of kaolinite. Br. Ceram. Trans. 88:162–166.

Dong, A., G. Chesters, and G.V. Simsiman. 1983. Soil dispersibility. Soil Sci. 136:208–212.

Doucet, F.J., L. Maguire, and J.R. Lead. 2004. Size fractionation of aquatic colloids and particles by cross flow filtration: Analysis by scanning electron and atomic force microscopy. Anal. Chim. Acta 522:59–71.

Douglas, L.A. 1977. Vermiculites, p. 259–292. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA Inc., Madison, WI.

Ducker, W.A., and R.M. Pashley. 1992. Forces between mica surfaces in the presence of rod-shaped divalent counterions. Langmuir 8:109–112.

Ducker, W.A., T.J. Senden, and R.M. Pashley. 1991. Direct measurement of colloidal forces using an atomic force microscope. Nature 353:239–241.

Dukhin, A.S., and P.J. Goetz. 1996. Acoustic and electroacoustic spectroscopy. Langmuir 12:4336–4344.

Dukhin, A.S., and P.J. Goetz. 1998. Characterization of aggregation phenomena by means of acoustic and electroacoustic spectroscopy. Colloids Surf. A 144:49–58.

Dukhin, A.S., and P.J. Goetz. 2001. Acoustic and electroacoustic spectroscopy for characterizing concentrated dispersions and emulsions. Adv. Colloid Interface Sci. 92:71–132.

Dukhin, A.S., and P.J. Goetz. 2002. Ultrasound for characterizing colloids. Elsevier, Amsterdam, the Netherlands.

Dukhin, A.S., P.J. Goetz, and S. Truesdail. 2001. Titration of concentrated dispersions using electroacoustic potential probe. Langmuir 17:964–968.

Dunnivant, F.M., P.M. Jardine, D.L. Taylor, and J.F. McCarthy. 1992. Transport of naturally occurring dissolved organic carbon in laboratory columns containing aquifer material. Soil Sci. Soc. Am. J. 56:437–444.

Duchaufour, P. 1970. Precis de pedologie. Masson et Cie., Paris, France.

Duran, J.D.G., M.M. Ramos-Tejeda, F.J. Arroyo, and F. Gonzalez-Caballaro. 2000. Rheological and electrokinetic properties of sodium montmorillonite suspensions. I. Rheological properties and interparticle energy of interaction. J. Colloid Interface Sci. 219:107–117.

Dyal, R.S., and S.B. Hendricks. 1950. Total surface of clays in polar liquids as a characteristic index. Soil Sci. 69:421–432.

Edwards, D.G., A.M. Posner, and J.P. Quirk. 1965a. Repulsion of chloride ions by negatively charged clay surfaces. Part 1. Monovalent cation Fithian illites. Trans. Faraday Soc. 61:2808–2815.

Edwards, D.G., A.M. Posner, and J.P. Quirk. 1965b. Repulsion of chloride ions by negatively charged clay surfaces. Part 2. Monovalent montmorillonites. Trans. Faraday Soc. 61:2816–2819.

Elimelech, M., and C.R. O'Melia. 1990. Kinetics of deposition of colloidal particles in porous media. Environ. Sci. Technol. 24:1528–1536.

Elimelech, M., M. Nagai, C. Ko, and J.N. Ryan. 2000. Relative insignificance of mineral grain zeta potential to colloid transport in geochemically heterogeneous porous media. Environ. Sci. Technol. 34:2143–2148.

El-Swaify, S.A., S. Ahmed, L.D. Swindale. 1970. Effects of adsorbed cations on physical properties of tropical red and tropical black earths. II. Liquid limit, degree of dispersion, and moisture retention. J. Soil Sci. 21:188–198.

Emerson, W.W. 1967. A classification of soil aggregates based on their coherence in water. Aust. J. Soil Res. 15:255–262.

Fanning, D.S., and V.Z. Keramidas. 1977. Micas, p. 195–258. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA Inc., Madison, WI.

Farmer, V.C. 1978. Water on particle surfaces, p. 405–448. *In* D.J. Greenland and M.H.B. Hayes (eds.) The chemistry of soil constituents. John Wiley & Sons, New York.

Filella, M., and J. Buffle. 1993. Factors controlling the stability of submicron colloids in natural waters. Colloids Surf. A 73:255–273.

Filella, M., J. Zhang, M.E. Newman, and J. Buffle. 1997. Analytical applications of photon correlation spectroscopy for size distribution measurements of natural colloidal suspensions: Capabilities and limitations. Colloids Surf. A 120:27–46.

Finsy, R. 1994. Particle sizing by quasi-elastic light scattering. Adv. Colloid Interface Sci. 52:79–143.

Flegman, A.W., J.W. Goodwin, and R.H. Ottewill. 1969. Rheological studies on kaolinite suspensions. Proc. Br. Ceram. Soc. 13:31–45.

Flury, M., J.B. Mathison, and J.B. Harsh. 2002. In situ mobilization of colloids and transport of cesium in Hanford sediments. Environ. Sci. Technol. 36:5335–5341.

Franchi, A., and C.R. O'Melia. 2003. Effects of natural organic matter and solution chemistry on the deposition and reentrainment of colloids in porous media. Environ. Sci. Technol. 37:1122–1129.

Frenkel, H., G.J. Levy, and M.V. Fey. 1992. Clay dispersion and hydraulic conductivity of clay–sand mixtures as affected by the addition of various anions. Clays Clay Miner. 40:515–521.

Freundlich, H. 1932. Kapillarchemie II. Eine Darstellung der Chemie der Kolloide und verwandter Gebiete. 4. Akademische Verlagsgesellschaft, Leipzig, Germany.

Gamerdinger, A.P., and D.I. Kaplan. 2001. Colloid transport and deposition in water-saturated Yucca Mountain tuff as determined by ionic strength. Environ. Sci. Technol. 35:3326–3331.

Ghezzehei, T.A., and D. Or. 2001. Rheological properties of wet soils and clays under steady and oscillatory stresses. Soil Sci. Soc. Am. J. 65:624–637.

Giddings, J.C. 1993. Field-flow fractionation: Analysis of macromolecular, colloidal, and particulate materials. Science 260:1456–1465.

Goldberg, S., and H.S. Forster. 1990. Flocculation of reference clays and arid-zone soil clays. Soil Sci. Soc. Am. J. 54:714–718.

Goldberg, S., H.S. Forster, and C.L. Godfrey. 1996. Molybdenum adsorption on oxides, clay minerals, and soils. Soil Sci. Soc. Am. J. 60:425–432.

Goldstein, J.I., D.E. Newbury, P. Echlin, D.C. Joy, A.D. Romig, C.E. Lyman, C. Fiori, and E. Lifshin. 1992. Scanning electron microscopy and X-ray microanalysis, 2nd Ed. Plenum Press, New York.

Greathouse, J.A., S.E. Feller, and D.A. McQuarrie. 1994. The modified Gouy–Chapman theory: Comparisons between electrical double layer models of clay swelling. Langmuir 10:2125–2130.

Greenwood, R. 2003. Review of the measurement of zeta potentials in concentrated aqueous suspensions using electroacoustics. Adv. Colloid Interface Sci. 106:55–81.

Gregg, S.J., and K.S.W. Sing. 1982. Adsorption, surface area and porosity. Academic Press, New York.

Grolimund, D., and M. Borkovec. 2001. Release and transport of colloidal particles in natural porous media: 1. Modeling. Water Resour. Res. 37:559–570.

Grolimund, D., and M. Borkovec. 2005. Colloid-facilitated transport of strongly sorbing contaminants in natural porous media: Mathematical modeling and laboratory column experiments. Environ. Sci. Technol. 39:6378–6386.

Grolimund, D., and M. Borkovec. 2006. Release of colloidal particles in natural porous media by monovalent and divalent cations. J. Contam. Hydrol. 87:155–175.

Grolimund, D., M. Borkovec, K. Bartmettler, and H. Sticher. 1996. Colloid-facilitated transport of strongly sorbing contaminants in natural porous media: A laboratory column study. Environ. Sci. Technol. 30:3118–3123.

Groves, M.J. 1984. The application of particle characterization methods to submicron dispersion and emulsions, p. 43–91. *In* H.G. Barth (ed.) Modern methods of particle size analysis. John Wiley & Sons, New York.

Gschwend, P.M., D.A. Backhus, J.K. MacFarlane, and A.L. Page. 1990. Mobilization of colloids in groundwater due to infiltration of water at coal ash disposal site. J. Contam. Hydrol. 6:307–320.

Gschwend, P.M., and M.D. Reynolds. 1987. Monodisperse ferrous phosphate colloids in an anoxic groundwater plume. J. Contam. Hydrol. 1:309–327.

Gu, B., and H.E. Doner. 1993. Dispersion and aggregation of soils as influenced by organic and inorganic polymers. Soil Sci. Soc. Am. J. 57:709–716.

Guerin, M., and J.C. Seaman. 2004. Characterizing clay mineral suspensions using acoustic and electroacoustic spectroscopy. Clays Clay Miner. 52:145–157.

Guerin, M., J.C. Seaman, C. Lehmann, and A. Jurgenson. 2004. Acoustic and electroacoustic characterization of variable-charge mineral suspensions. Clays Clay Miner. 52:158–170.

Gunnarsson, M., M. Rasmusson, S. Wall, E. Ahlberg, and J. Ennis. 2001. Electroacoustic and potentiometric studies of the hematite/water interface. J. Colloid Interface Sci. 240:448–458.

Hahn, M.W., and C.R. O'Melia. 2004. Deposition and reentrainment of Brownian particles in porous media under unfavorable chemical conditions: Some concepts and applications. Environ. Sci. Technol. 38:210–220.

Handy, R.D., F. von der Kammer, J.R. Lead, M. Hassellov, R. Owen, and M. Crane. 2008. The ecotoxicology and chemistry of manufactured nanoparticles. Ecotoxicology 17:287–314.

Hang, P.T., and G.W. Brindley. 1970. Methylene blue adsorption by clay minerals: Determination of surface areas and cation exchange capacities. Clays Clay Miner. 18:203–212.

Hassellov, M., J.W. Readman, J.F. Ranville, and K. Tiede. 2008. Nanoparticle analysis and characterization methodologies in environmental risk assessment of engineered nanoparticles. Ecotoxicology 17:344–361.

Hassellov, M., F. von der Kammer, and R. Beckett. 2007. Characterisation of aquatic colloids and macromolecules by field-flow fractionation, p. 23–276. *In* K.J. Wilkinson

and J.R. Lead (eds.) Environmental colloids and particles: Behavior, structure and characterization. IUPAC series on analytical and physical chemistry of environmental systems. John Wiley & Sons, Chichester, U.K.

Heath, D., and Th.F. Tadros. 1983. Influence of pH, electrolyte and poly(vinyl alcohol) addition on the rheological characteristics of aqueous dispersions of sodium montmorillonite. J. Colloid Interface Sci. 93:307–319.

Heil, D., and G. Sposito. 1993. Organic matter role in illitic soil colloids flocculation: II. Surface charge. Soil Sci. Soc. Am. J. 57:1246–1253.

Heister, K., P.J. Kleingeld, T.J.S. Keijzer, and J.P.G. Loch. 2005. A new laboratory set-up for measurements of electrical, hydraulic, and osmotic fluxes in clays. Eng. Geol. 77:295–303.

Hesterberg, D.L. 1988. Critical coagulation concentrations and rheological properties of illite. Ph.D. Thesis. University of California, Riverside, CA.

Hesterberg, D., and A.L. Page. 1990a. Flocculation series test yielding time-invariant critical coagulation concentrations of sodium illite. Soil Sci. Soc. Am. J. 54:729–735.

Hesterberg, D., and A.L. Page. 1990b. Critical coagulation concentration of sodium and potassium illite as affected by pH. Soil Sci. Soc. Am. J. 54:735–739.

Hesterberg, D., and A.L. Page. 1993. Rheology of sodium and potassium illite suspensions in relation to colloidal stability. Soil Sci. Soc. Am. J. 57:697–704.

Hiemenz, P.C. 1977. Principles of colloid and surface chemistry. Marcel Dekker Inc., New York.

Hiemenz, P.C., and R. Rajagopalan. 1997. Principles of colloid and surface chemistry, 3rd Ed. Marcel Dekker, New York.

Hochella, M.F. 2008. Nanogeoscience: From origins to cutting-edge applications. Elements 4:373–379.

Holsworth, R.M., T. Provder, and J.J. Stansbrey. 1987. External-gradient-formation method for disc centrifuge photosedimentometric particle size distribution analysis, p. 191–201. *In* T. Provder (ed.) Particle size distribution. Assessment and characterization. American Chemical Society, Washington, DC.

Hsu, P.H. 1977. Aluminum hydroxides and oxyhydroxides, p. 99–143. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA Inc., Madison, WI.

Hsu, J.P. 1999. Interfacial forces and fields: Theory and applications. CRC Press, New York.

Huang, C.P. 1981. The surface acidity of hydrous solids, p. 183–217. *In* M.A. Anderson and A.J. Rubin (eds.) Adsorption of inorganics at solid–liquid interfaces. Ann Arbor Science, Ann Arbor, MI.

Hunter, R.J. 1981. Zeta potential in colloid science. Academic Press, London, U.K.

Hunter, R.J. 1987. Foundations of colloid science, Vol. 1. Oxford University Press, New York.

Hunter, R.J. 1989. Foundations of colloid science, Vol. 2. Oxford University Press, New York.

Hutton, J.T. 1977. Titanium and zirconium minerals, p. 673–688. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA Inc., Madison, WI.

Ise, N., and M.V. Smalley. 1994. Thermal compression of colloidal crystals: Paradox of repulsion-only assumption. Phys. Rev. B Condens. Matter. 50:16722–16725.

Israelachvili, J.N. 1992. Intermolecular and surface forces, 2nd Ed. Academic Press, San Diego, CA.

IUPAC. 2009. International Union of Pure and Applied Chemistry. http://www.iupac.org

Johnson, C.R., L.A. Hellerich, N.P. Nikolaidis, and P.M. Gschwend. 2001. Colloid mobilization in the field using citrate to remediate chromium. Ground Water 39:895–903.

Kaplan, D.I., P.M. Bertsch, and D.C. Adriano. 1995. Facilitated transport of contaminant metals through an acidified aquifer. Ground Water 33:708–717.

Kaplan, D.I., P.M. Bertsch, and D.C. Adriano. 1997. Mineralogical and physicochemical differences between mobile and non-mobile colloidal phases in reconstructed pedons. Soil Sci. Soc. Am. J. 61:641–649.

Kaplan, D.I., P.M. Bertsch, D.C. Adriano, and W.P. Miller. 1993. Soil-borne colloids as influenced by water flow and organic carbon. Environ. Sci. Technol. 27:1193–1200.

Kaplan, D.I., P.M. Bertsch, D.C. Adriano, and K.A. Orlandini. 1994a. Actinide association with groundwater colloids in a coastal plain aquifer. Radiochim. Acta 66/67:181–187.

Kaplan, D.I., D.B. Hunter, P.M. Bertsch, S. Bajt, and D.C. Adriano. 1994b. Application of synchrotron x-ray fluorescence spectroscopy and energy dispersive x-ray analysis to identify contaminant metals on groundwater colloids. Environ. Sci. Technol. 28:1186–1189.

Karathanasis, A.D. 1999. Subsurface migration of copper and zinc mediated by soil colloids. Soil Sci. Soc. Am. J. 63:830–838.

Karathanasis, A.D., D.M.C. Johnson, and C.J. Matocha. 2005. Biosolid colloid-mediated transport of copper, zinc, and lead in waste-amended soils. J. Environ. Qual. 34:1153–1164.

Kavanaugh, M.C., and J.O. Leckie. 1980. Particulates in water. Adv. Chem. Ser. No. 189. American Chemical Society, Washington, DC.

Kékicheff, P., S. Marcelja, T.J. Senden, and V.E. Shubin. 1993. Charge reversal seen in electrical double layer interaction of surfaces immersed in 2:1 calcium electrolyte. J. Chem. Phys. 99:6098–6113.

Keren, R. 1988. Rheology of aqueous suspension of sodium/calcium montmorillonite. Soil Sci. Soc. Am. J. 52:924–928.

Keren, R. 1989a. Effect of clay charge density and adsorbed ions on the rheology of montmorillonite suspension. Soil Sci. Soc. Am. J. 53:25–29.

Keren, R. 1989b. Rheology of mixed kaolinite–montmorillonite suspensions. Soil Sci. Soc. Am. J. 53:725–730.

Keren, R. 1991. Adsorbed sodium fraction's effect on rheology of montmorillonite–kaolinite suspensions. Soil Sci. Soc. Am. J. 55:376–379.

Kersting, A.B., D.W. Efurd, D.L. Finnegan, D.J. Rokop, D.K. Smith, and J.L. Thompson. 1999. Migration of plutonium in groundwater at the Nevada Test Site. Nature 397:56–59.

Khaleel, R., T.C.J. Yeh, and Z. Lu. 2002. Upscaled flow and transport properties for heterogeneous unsaturated media. Water Resour. Res. 38:11/1–11/12.

Khilar, K.C., and H.S. Fogler. 1984. The existence of a critical salt concentration for particle release. J. Colloid Interface Sci. 101:214–224.

Kia, S.F., H.S. Fogler, and M.G. Reed. 1987. Effect of pH on colloidally induced fines migration. J. Colloid Interface Sci. 118:158–168.

Kim, S.-O., S.-H. Moon, and K.-W. Kim. 2001. Removal of heavy metals from soils using enhanced electrokinetic soil processing. Water Air Soil Pollut. 125:259–272.

Kjellander, R., S. Marcelja, R.M. Pashley, and J.P. Quirk. 1990. A theoretical and experimental study of forces between charged mica surfaces in aqueous $CaCl_2$ solutions. J. Chem. Phys. 92:4399–4407.

Koehler, M.E., R.A. Zandler, T. Gill, T. Provder, and T.F. Niemann. 1987. An improved disc centrifuge photosedimentometer and data system for particle size distribution analysis, p. 180–190. *In* T. Provder (ed.) Particle size distribution. Assessment and characterization. American Chemical Society, Washington, DC.

Kosmulski, M. 2002. Confirmation of the differentiating effect of small cations in the shift of the isoelectric point of oxides at high ionic strengths. Langmuir 18:785–787.

Kosmulski, M., E. Maczka, E. Jartych, and J.B. Rosenholm. 2003. Synthesis and characterization of goethite and goethite–hematite composite: Experimental study and literature survey. Adv. Colloid Interface Sci. 103:57–76.

Kosmulski, M., E. Maczka, and J.B. Rosenholm. 2002. Isoelectric points of metal oxides at high ionic strengths. J. Phys. Chem. B 106:2918–2921.

Kretzschmar, R., M. Borkovec, D. Grolimund, and M. Elimelech. 1999. Mobile subsurface colloids and their role in contaminant transport. Adv. Agron. 66:121–193.

Kretzschmar, R., D. Hesterberg, and H. Sticher. 1997. Effects of adsorbed humic acid on surface charge and flocculation of kaolinite. Soil Sci. Soc. Am. J. 61:101–108.

Kretzschmar, R., W.P. Robarge, and S.B. Weed. 1993. Flocculation of kaolinitic soil clays: Effects of humic substances and iron oxides. Soil Sci. Soc. Am. J. 57:1277–1283.

Langmuir, I. 1918. The adsorption of gases on plane surfaces of glass, mica, and platinum. J. Am. Chem. Soc. 40:1361–1402.

Laskin, A., and J.P. Cowin. 2001. Automated single-particle SEM/EDX analysis of submicrometer particles down to 0.1 μm. Anal. Chem. 73:1023–1029.

Laxton, P.B., and J.C. Berg. 2006. Relating clay yield stress to colloidal parameters. J. Colloid Interface Sci. 296:749–755.

Lead, J.R., D. Muirhead, and C.T. Gibson. 2005. Characterization of freshwater natural aquatic colloids by atomic force microscopy (AFM). Environ. Sci. Technol. 39:6930–6936.

Lead, J.R., and K.J. Wilkinson. 2007. Environmental colloids and particles: Current knowledge and future developments, p. 1–15. *In* K.J. Wilkinson and J.R. Lead (eds.) Environmental colloids and particles: Behavior, structure and characterization. IUPAC series on analytical and physical chemistry of environmental systems. John Wiley & Sons, Chichester, U.K.

Lead, J.R., K.J. Wilkinson, E. Balnois, B. Cutak, C. Larive, S. Assemi, and R. Beckett. 2000. Diffusion coefficients and polydispersities of the Suwannee River fulvic acid: Comparison of fluorescence correlation spectroscopy, pulsed-field gradient nuclear magnetic resonance and flow field-flow fractionation. Environ. Sci. Technol. 34:3508–3513.

Lebron, I., M.G. Schaap, and D.L. Suarez. 1999. Saturated hydraulic conductivity prediction from microscopic pore geometry measurements and neural network analysis. Water Resour. Res. 35:3149–3157.

Lebron, I., and D.L. Suarez. 1992a. Electrophoretic mobility of illite and micaceous soil clays. Soil Sci. Soc. Am. J. 56:1106–1115.

Lebron, I., and D.L. Suarez. 1992b. Variations in soil stability within and among soil types. Soil Sci. Soc. Am. J. 56:1412–1421.

Lebron, I., D.L. Suarez, C. Amrhein, and J.E. Strong. 1993. Size of mica domains and distribution of the adsorbed Na–Ca ions. Clays Clay Miner. 41:380–388.

Ledin, A., S. Karlsson, A. Duker, and B. Allard. 1993. Applicability of photon correlation spectroscopy for measurement of concentration and size distribution of colloids in natural waters. Anal. Chim. Acta 281:421–428.

Ledin, A., S. Karlsson, A. Duker, and B. Allard. 1994. Measurements in-situ of concentration and size distribution of colloidal matter in deep groundwaters by photon-correlation spectroscopy. Water Res. 28:1539–1545.

Li, X., T.D. Scheibe, and W.P. Johnson. 2004. Apparent decreases in colloid deposition rate coefficients with distance of transport under unfavorable deposition conditions: A general phenomenon. Environ. Sci. Technol. 38:5616–5625.

Lowell, S. 1979. Introduction to powder surface area. John Wiley & Sons, New York.

Ma, K., and A.C. Pierre. 1999. Clay sediment-structure formation in aqueous kaolinite suspensions. Clays Clay Miner. 47:522–526.

Maes, A., and A. Cremers. 1977. Charge density effects in ion exchange. Part 1. Heterovalent exchange equilibria. J. Chem. Soc. Faraday Trans. 1. 73:1807–1814.

Markgraf, W., R. Horn, and S. Peth. 2006. An approach to rheometry in soil mechanics—Structural changes in bentonite, clayey and silty soils. Soil Till. Res. 91:1–14.

Mashal, K., J.B. Harsh, M. Flury, A.R. Felmy, and H. Zhao. 2004. Colloid formation in Hanford sediments reacted with simulated tank waste. Environ. Sci. Technol. 38: 5750–5756.

Mavrocordatos, D., D. Perret, and G.D. Leppard. 2007. Strategies and advances in the characterization of environmental colloids by electron microscopy, p. 345–404. *In* K.J. Wilkinson

and J.R. Lead (eds.) Environmental colloids and particles: Behavior, structure and characterization. IUPAC series on analytical and physical chemistry of environmental systems. John Wiley & Sons, Chichester, U.K.

McBride, M.B. 1989. Surface chemistry of soil minerals, p. 35–88. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments, 2nd Ed. Soil Sci. Soc. Am. Book Ser. 1. SSSA, Madison, WI.

McBride, M.B. 1994. Environmental chemistry of soils. Oxford University Press, New York.

McBride, M.B., and P. Baveye. 2002. Diffuse double-layer models, long-range forces, and ordering in clay colloids. Soil Sci. Soc. Am. J. 66:1207–1217.

McCarthy, J.F., and C. Degueldre. 1993. Sampling and characterization of colloids and particles in groundwater for studying their role in contaminant transport, p. 247–315. *In* J. Buffle and H.P. van Leeuwen (eds.) Environmental particles, Vol. 2. Lewis Publishers, Ann Arbor, MI.

McCarthy, J.F., and L.D. McKay. 2004. Colloid transport in the subsurface: Past, present, and future challenges. Vadose Zone J. 3:326–337.

McCarthy, J.F., L.D. McKay, and D.D. Bruner. 2002. Influence of ionic strength and cation charge on transport of colloidal particles in fractured shale saprolite. Environ. Sci. Technol. 36:3735–3743.

McCarthy, J.E., and J.M. Zachara. 1989. Subsurface transport of contaminants. Environ. Sci. Technol. 23:496–502.

McDowell-Boyer, L.M. 1992. Chemical mobilization of micron-sized particles in saturated porous media under steady flow conditions. Environ. Sci. Technol. 26:586–593.

McDowell-Boyer, L.M., J.R. Hunt, and N. Sitar. 1986. Particle transport through porous media. Water Resour. Res. 22:1901–1921.

McKenzie, R.M. 1989. Manganese oxides and hydroxides, p. 439–465. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments, 2nd Ed. SSSA Inc., Madison, WI.

Miano, F., and M.R. Rabaioli. 1994. Rheological scaling of montmorillonite suspensions: The effect of electrolytes and polyelectrolytes. Colloids Surf. A 84:229–237.

Miller, W.P., and M.K. Baharuddin. 1986. Relationship of soil dispersibility to infiltration and erosion of southeastern soils. Soil Sci. Soc. Am. J. 54:346–351.

Miller, W.P., H. Frenkel, and K.D. Newman. 1990. Flocculation concentration and sodium/calcium exchange of kaolinitic soil clays. Soil Sci. Soc. Am. J. 54:346–351.

Miller, W.P., and D.E. Radcliffe. 1992. Soil crusting in the southeastern United States, p. 233–266. *In* M.E. Sumner and B.A. Stewart (eds.) Soil crusting: Chemical and physical processes. Lewis Publishers, Boca Raton, FL.

Mills, W.B., S. Liu, and F.K. Fong. 1991. Literature review and model (comet) for colloid/metals transport in porous media. Ground Water 29:199–208.

Murphy, D.M., J.R. Garbarino, H.W. Taylor, B.T. Hart, R. Beckett. 1993. Determination of size and element composition distributions of complex colloids by sedimentation field-flow fractionation-inductively coupled plasma mass spectrometry. J. Chromatogr. 642:459–467.

Napper, D.H., and R.J. Hunter. 1974. Hydrosols, p. 161–213. *In* M. Kerker (ed.) M.T.P. Int. Rev. Sci. Phys. Chem. Ser. Butterworths, London, U.K.

Nelson, P.N., and J.M. Oades. 1998. Organic matter, sodicity and soil structure. p. 51–75. *In* M.E. Sumner and R. Naidu (eds.) Sodic soils: Distribution, properties, management and environmental consequences. Oxford University Press, New York.

Newman, A.C.D. 1983. The specific surface of soils determined by water sorption. Soil Sci. 34:23–32.

Newman, M.E., M. Filella, Y.W. Chen, J.C. Negre, D. Perret, and J. Buffle. 1994. Submicron particles in the Rhine river. 2. Comparison of field observations and model predictions. Water Res. 28:107–118.

Nishimura, S., M. Kodama, K. Yao, Y. Imai, and H. Tateyama. 2002b. Direct surface force measurement for synthetic smectites using the atomic force microscope. Langmuir 18:4681–4688.

Nishimura, S., H. Tateyama, K. Tsunematsu, and K. Jinnai. 1992. Zeta potential measurement of muscovite mica basal plane-aqueous solution interface by means of plane interface technique. J. Colloid Interface Sci. 152:359–367.

Nishimura, S., K. Yao, M. Kodama, Y. Imai, K. Ogino, and K. Mishima. 2002a. Electrokinetic study of synthetic smectites by flat plate streaming potential technique. Langmuir 18:188–193.

Norrish, K., and J.A. Rausell-Colom. 1961. Low-angle X-ray diffraction studies of the swelling of montmorillonite and vermiculite. Clays Clay Miner. 10:123–149.

Ohtsubo, M., A. Yoshimura, S.-I. Wada, and R.N. Yong. 1991. Particle interaction and rheology of illite-iron oxide complexes. Clays Clay Miner. 39:347–354.

O'Melia, C.R, and C.L. Tiller. 1993. Physicochemical aggregation and deposition in aquatic environments, p. 353–385. *In* J. Buffle and H. van Leeuwen (eds.) Environmental particles, Vol. 2. Lewis Publishers, Boca Raton, FL.

Overbeek, J.Th.G. 1952. Stability of hydrophobic colloids and emulsions. p. 302–341. *In* H.R. Kryt (ed.) J. Colloid Sci. Vol. 1, Elsevier, Amsterdam.

Parker, J.C., L.W. Zelazny, S. Sampath, and W.G. Harris. 1979. A critical evaluation of the extension of zero point of charge (ZPC) theory to soil systems. Soil Sci. Soc. Am. J. 43:668–674.

Pashley, R.M. 1981. DLVO and hydration forces between mica surfaces in Li^+, Na^+, K^+, and Cs^+ electrolyte solutions: A correlation of double-layer and hydration forces with surface cation. J. Colloid Interface Sci. 83:531–546.

Pashley, R.M., and J.P. Quirk. 1984. The effect of cation valency on DLVO and hydration forces between macroscopic sheets of muscovite mica in relation to clay swelling. Colloids Surf. 9:1–17.

Pecora, R. 1983. Quasi-elastic light scattering of macromolecules and particles in solution and suspension, p. 3–30. *In* B.E. Dahneke (ed.) Measurement of suspended particles by quasi-elastic light scattering. John Wiley & Sons, New York.

Perret, D., M.E. Newman, J.C. Negre, Y.W. Chen, and J. Buffle. 1994. Submicron particles in the Rhine river. 1. Physicochemical characterization. Water Res. 28:91–106.

Pennell, K.D., S.A. Boyd, and L.M. Abriola. 1995. Surface area of soil organic matter reexamined. Soil Sci. Soc. Am. J. 59:1012–1018.

Quirk, J.P. 1994. Interparticle forces: A basis for the interpretation of soil physical behavior. Adv. Agron. 53:121–183.

Quirk, J.P., and L.A.G. Aylmore. 1971. Domain and quasi-crystalline regions in clay systems. Proc. Soil Sci. Soc. Am. 35:652–654.

Quirk, J.P., and R.S. Murray. 1991. Towards a model for soil structural behavior. Aust. J. Soil Res. 29:829–867.

Quirk, J.P., and R.K. Schofield. 1955. The effect of electrolyte concentration on soil permeability. J. Soil Sci. 6:163–178.

Rand, B., and I.E. Melton. 1977. Particle interactions in aqueous kaolinite suspensions. I. Effect of pH and electrolyte upon the mode of particle interaction in homoionic sodium kaolinite suspensions. J. Colloid Interface Sci. 60:308–320.

Rand, B., E. Pekenc, J.W. Goodwin, and R.W. Smith. 1980. Investigation into the existence of edge–face coagulated structures in Na-montmorillonite suspensions. J. Chem. Soc. Faraday Trans. 1 76:225–235.

Ranville, J.F., D.J. Chittleborough, and R. Beckett. 2005. Particle-size and element distributions of soil colloids: Implications for colloid transport. Soil Sci. Soc. Am. J. 69:1173–1184.

Ranville, J.F., D.J. Chittleborough, F. Sanks, R.J.S. Morrison, T. Harris, F. Doss, et al. 1999. Development of sedimentation field-flow fractionation-inductively coupled plasma mass-spectroscopy for the characterization of environmental colloids. Anal. Chim. Acta 381:315–329.

Redman, J.A., S.L. Walker, and M. Elimelech. 2004. Bacterial adhesion and transport in porous media: Role of the secondary energy minimum. Environ. Sci. Technol. 38:1777–1785.

Rees, L.B. 1990. A Monte-Carlo approach to x-ray attenuation corrections for proton-induced x-ray-emission from particulate samples. Nucl. Instrum. Meth. Phys. Res. Sect. A 299:614–617.

Rockhold, M.L., R.R. Yarwood, and J.S. Selker. 2004. Coupled microbial and transport processes in soil. Vadose Zone J. 3:368–383.

Ronen, D., M. Magaritz, U. Weber, A.J. Amiel, and E. Klein. 1992. Characterization of suspended particles collected in groundwater under natural gradient flow conditions. Water Resour. Res. 28:1279–1291.

Ross, G. 1978. Relationships of specific surface area and clay content to shrink–swell potential of soils having different clay mineralogical compositions. Can. J. Soil Sci. 58:159–166.

Roy, S.B., and D.A. Dzombak. 1996. Colloid release and transport processes in natural and model porous media. Colloids Surf. A 107:245–262.

Roy, S.B., and D.A. Dzombak. 1997. Chemical factors influencing colloid-facilitated transport of contaminants in porous media. Environ. Sci. Technol. 37:656–664.

Roy, S.B., and D.A. Dzombak. 1998. Sorption nonequilibrium effects on colloid-enhanced transport of hydrophobic compounds in porous media. J. Contam. Hydrol. 30:179–200.

Ryan, J.N., and M. Elimelech. 1996. Colloid mobilization and transport in groundwater. Colloids Surf. A 107:1–56.

Ryan, J.N., and P.M. Gschwend. 1990. Colloid mobilization in two Atlantic Coastal Plain aquifers: Field studies. Water Resour. Res. 26:307–322.

Ryan, J.N., and P.M. Gschwend. 1994. Effect of solution chemistry on clay colloid release from an iron oxide-coated aquifer sand. Environ. Sci. Technol. 28:1717–1726.

Ryan, J.N., T.H. Illangasekare, M.I. Litaor, and R. Shannon. 1998. Particle and plutonium mobilization in macroporous soils during rainfall simulations. Environ. Sci. Technol. 32:476–482.

Scales, P.J., F. Grieser, and T.W. Healy. 1990. Electrokinetics of the muscovite mica–aqueous solution interface. Langmuir 6:582–589.

Schofield, R.K. 1949. Calculation of surface areas of clays from measurements of negative adsorption. Trans. Br. Ceram. Soc. 48:207–213.

Schofield, R.K., and H.R. Samson. 1954. Flocculation of kaolinite due to the attraction of oppositely charged crystal faces. Discuss. Faraday Soc. 18:135–145.

Schramm, L.L., and J.C.T. Kwak. 1982a. Influence of exchangeable cation composition on the size and shape of montmorillonite particles in dilute suspensions. Clays Clay Miner. 30:40–48.

Schramm, L.L., and J.C.T. Kwak. 1982b. Interactions in clay suspensions: The distribution of ions in suspension and the influence of tactoid formation. Colloids Surf. 3:43–60.

Schulthess, C.P., and D.L. Sparks. 1986. Backtitration technique for proton isotherm modeling of oxide surfaces. Soil Sci. Soc. Am. J. 50:1406–1411.

Schultz, D.S. 1997. Electroosmosis technology for soil remediation: Laboratory results, field trial, and economic modeling. J. Hazard. Mater. 55:81–91.

Schulze, D.G. 2002. An introduction to soil mineralogy, p. 1–35. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Schurtenberger, P., and M.E. Newman. 1993. Characterization of biological and environmental particles using static and dynamic light scattering, p. 37–115. *In* J. Buffle and H.P. van Leeuwen (eds.) Environmental particles, Vol. 2. Lewis Publishers, Ann Arbor, MI.

Schwertmann, U., and R.M. Taylor. 1977. Iron oxides, p. 145–180. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA Inc., Madison, WI.

Seaman, J.C. 2000. Thin-foil SEM analysis of soil and groundwater colloids: Reducing instrument and operator bias. Environ. Sci. Technol. 34:187–191.

Seaman, J.C., and P.M. Bertsch. 2000. Selective colloid mobilization through surface-charge manipulation. Environ. Sci. Technol. 34:3749–3755.

Seaman, J.C., P.M. Bertsch, and W.P. Miller. 1995. Chemical controls on colloid generation and transport in a sandy aquifer. Environ. Sci. Technol. 29:1808–1815.

Seaman, J.C., P.M. Bertsch, and R.N. Strom. 1997. Characterization of colloids mobilized from southeastern coastal plain sediments. Environ. Sci. Technol. 31:2782–2790.

Seaman, J.C., M. Guerin, B.P. Jackson, P.M. Bertsch, and J.F. Ranville. 2003. Analytical techniques for characterizing complex mineral assemblages: Mobile soil and groundwater colloids, p. 271–309. *In* H.M. Selim and W.L. Kingery (eds.) Geochemical and hydrological reactivity of heavy metals in soils. CRC Press, Washington, DC.

Secor, R.B., and C.J. Radke. 1985. Spillover of the diffuse double layer on montmorillonite particles. J. Colloid Interface Sci. 103:237–244.

Seta, A.K., and A.D. Karathanasis. 1996. Colloid-facilitated transport of metolachlor through intact soil columns. J. Environ. Sci. Health B 31:949–968.

Seta, A.K., and A.D. Karathanasis. 1997. Atrazine adsorption by soil colloids and co-transport through subsurface environments. Soil Sci. Soc. Am. J. 61:612–617.

Shainberg, I., and H. Otoh. 1968. Size and shape of montmorillonite particles saturated with Na/Ca ions (inferred from viscosity and optical measurements). Isr. J. Chem. 6:251–259.

Shainberg, I., J.D. Rhoades, and R.J. Prather. 1981. Effect of low electrolyte concentration on clay dispersion and hydraulic conductivity of a sodic soil. Soil Sci. Soc. Am. J. 45:273–277.

Shapiro, A.P., and R.F. Probstein. 1993. Removal of contaminants from saturated clay by electroosmosis. Environ. Sci. Technol. 27:283–291.

Sharma, M.M., H. Chamoun, D.H.S.S.R. Sarma, and R.S. Schecter. 1992. Factors controlling the hydrodynamic attachment of particles from surfaces. J. Colloid Interface Sci. 149:121–134.

Sherard, J.L., L.P. Dunningan, and R.S. Decker. 1977. Identification and nature of dispersive soils. J. Geotech. Eng. Am. Soc. Chem. Eng. 4:287–301.

Simunek, J., C. He, L. Pang, and S.A. Bradford. 2006. Colloid-facilitated solute transport in variably saturated porous media: Numerical model and experimental verification. Vadose Zone J. 5:1035–1047.

Sinton, L.W., M.J. Noonan, R.K. Finlay, L. Pang, and M.E. Close. 2000. Transport and attenuation of bacteria and bacteriophages in an alluvial gravel aquifer. N.Z. J. Mar. Freshwater Res. 34:175–186.

Sirivithayapakorn, S., and A. Keller. 2003. Transport of colloids in unsaturated porous media: A pore-scale observation of processes during the dissolution of air–water interface. Water Resour. Res. 39:SBH 6/1–6/10.

Smalley, M.V. 1990. Electrostatic interaction in macro-ionic solutions and gels. Mol. Phys. 71:1251–1267.

Sogami, I., and N. Ise. 1984. On the electrostatic interaction in macroionic solutions. J. Chem. Phys. 81:6320–6332.

Sposito, G. 1984. The surface chemistry of soils. Oxford University Press, New York.

Stevenson, F.J. 1982. Humus chemistry. Genesis, composition, reactions. John Wiley & Sons, New York.

Suarez, D.L., J.D. Rhoades, R. Lavado, and C.M. Grieve. 1984. Effect of pH on saturated hydraulic conductivity and soil dispersion. Soil Sci. Soc. Am. J. 48:50–55.

Sun, Y.-P., X.-Q. Li, J. Cao, W.-X. Zhang, and H.P. Wang. 2006. Characterization of zero-valent iron nanoparticles. Adv. Colloid Interface Sci. 120:47–56.

Supak, J.R., A.R. Swoboda, and J.B. Dixon. 1978. Adsorption of aldicarb by clays and soil organo-clay complexes. Soil Sci. Soc. Am. J. 42:244–248.

Swartz, C.H., and P.M. Gschwend. 1998. Mechanisms controlling release of colloids to groundwater in a southeastern coastal plain aquifer sand. Environ. Sci. Technol. 32:1779–1785.

Taboada-Serrano, P., V. Vithayaveroj, S. Yiacoumi, and C. Tsouris. 2005. Surface charge heterogeneities measured by atomic force microscopy. Environ. Sci. Technol. 39:6352–6360.

Tarchitzky, J., Y. Chen, and A. Banin. 1993. Humic substances and pH effects on sodium and calcium-montmorillonite flocculation and dispersion. Soil Sci. Soc. Am. J. 57:367–372.

Theng, B.K., and G. Yuan. 2008. Nanoparticles in the soil environment. Elements 4:395–399.

Tiller, K.G., and L.H. Smith. 1990. Limitations of EGME retention to estimate the surface area of soils. Aust. J. Soil Res. 28:1–26.

Tombácz, E., J. Balázs, J. Lakatos, and F. Szántó. 1989. Influence of the exchangeable cations on stability and rheological properties of montmorillonite suspensions. Colloid Polym. Sci. 267:1016–1025.

Torkzaban, S., S.S. Tazehkand, S.L. Walker, and S.A. Bradford. 2008. Transport and fate of bacteria in porous media: Coupled effects of chemical conditions and pore space geometry. Water Resour. Res. 44:W04403, doi: 10.1029/2007WR006541.

Tufenkji, N. 2007. Modeling microbial transport in porous media: Traditional approaches and recent developments. Adv. Water Resour. 30:1455–1469.

Tufenkji, N., and M. Elimelech. 2004a. Correlation equation for predicting single-collector efficiency in physicochemical filtration in saturated porous media. Environ. Sci. Technol. 38:529–536.

Tufenkji, N., and M. Elimelech. 2004b. Deviation from the classical colloid filtration theory in the presence of repulsive DLVO interactions. Langmuir 20:10818–10828.

Tufenkji, N., and M. Elimelech. 2005. Breakdown of colloid filtration theory: Role of the secondary energy minimum and surface charge heterogeneities. Langmuir 21:841–852.

Tufenkji, N., J.A. Redman, and M. Elimelech. 2003. Interpreting deposition patterns of microbial particles in laboratory-scale column experiments. Environ. Sci. Technol. 37:616–623.

van de Weerd, H., and A. Leijnse. 1997. Assessment of the effect of kinetics on colloid facilitated radionuclide transport in porous media. J. Contam. Hydrol. 26:245–256.

van Olphen, H. 1977. An introduction to clay colloid chemistry, 2nd Ed. John Wiley & Sons, New York.

Verwey, E.J.W., and J.Th.G. Overbeek. 1948. Theory of stability of lyophobic colloids. Elsevier, Amsterdam, the Netherlands.

Vinten, A.J.A., B. Yaron, and P.H. Nye. 1983. Vertical transport of pesticides when adsorbed on suspended particles. J. Agric. Food Chem. 31:662–664.

Von Gunten, H.R., U.E. Warber, and U. Krahenbuhl. 1988. The reactor accident at Chernobyl: A possibility to test colloid-controlled transport of radionuclides in a shallow aquifer. J. Contam. Hydrol. 2:237–247.

von Smoluchowski, M. 1916. Three discourses on diffusion, Brownian movements, and the coagulation of colloid particles. Phys. Z. 17:557–571, 585–599.

von Smoluchowski, M. 1917. Mathematical theory of the kinetics of the coagulation of colloidal suspensions. Z. Phys. Chem. 92:129–168.

Walker, S.L., J.A. Redman, and M. Elimelech. 2004. Role of cell surface lipopolysaccharides in *Escherichia coli* K12 adhesion and transport. Langmuir 20:7736–7746.

Wan, J., and J.L. Wilson. 1994. Visualization of the role of the gas–water interface on the fate and transport of colloids in porous media. Water Resour. Res. 30:11–23.

Waychunas, G.A., and H. Zhang. 2008. Structure, chemistry, and properties of mineral nanoparticles. Elements 4:381–387.

Wilding, L.P., N.E. Smeck, and L.R. Drees. 1977. Silica in soils: Quartz, cristobalite, tridymite, and opal, p. 471–552. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA Inc., Madison, WI.

Wilkinson, K.J., and J.R. Lead. 2007. Environmental colloids and particles: Behavior, structure and characterization. IUPAC series on analytical and physical chemistry of environmental systems. John Wiley & Sons, Chichester, U.K.

Yao, K.M., T. Habibian, and C.R. O'Melia. 1971. Water and waste water filtration, concepts and applications. Environ. Sci. Technol. 5:1105–1112.

Yong, R.N., A.J. Sadh, H.P. Ludwig, and M.A. Jorgensen. 1979. Interparticle action and rheology of dispersive clays. J. Geotech. Eng. Div. 105:1193–1209.

Zhao, H., S. Bhattacharjee, R. Chow, D. Wallace, J.H. Masliyah, and Z. Xu. 2008. Probing surface charge potentials of clay basal planes and edges by direct force measurements. Langmuir 24:12899–12910.

Zhao, H., P.F. Low, and J.M. Bradford. 1991. Effects of pH and electrolyte concentration on particle interaction in three homoionic sodium soil clay suspensions. Soil Sci. 151:196–207.

16

Ion Exchange Phenomena

Ian C. Bourg
University of California

Garrison Sposito
University of California, Berkeley

16.1 Introduction

Ion exchange phenomena involve the population of *readily exchangeable ions*, the subset of adsorbed solutes that balance the intrinsic surface charge and can be readily replaced by major background electrolyte ions (Sposito, 2008). These phenomena have occupied a central place in soil chemistry research since Way (1850) first showed that potassium uptake by soils resulted in the release of an equal quantity of moles of charge of calcium and magnesium. Ion exchange phenomena are now routinely modeled in studies of soil formation (White et al., 2005), soil reclamation (Kopittke et al., 2006), soil fertilization (Agbenin and Yakubu, 2006), colloidal dispersion/flocculation (Charlet and Tournassat, 2005), the mechanics of argillaceous media (Gajo and Loret, 2007), aquitard pore water chemistry (Tournassat et al., 2008), and groundwater (Timms and Hendry, 2007; McNab et al., 2009) and contaminant hydrology (Chatterjee et al., 2008; van Oploo et al., 2008; Serrano et al., 2009).

The prototypical chemical reaction equation for the exchange of cations A^{a+} and B^{b+} can be written as follows if X^- represents a mole of negative charge carried by the solid exchanger:

$$bX_aA_{(s)} + aB^{b+}_{(aq)} = aX_bB_{(s)} + bA^{a+}_{(aq)}. \tag{16.1}$$
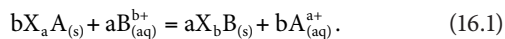
Equation 16.1 can be modified in a straightforward manner to describe anion exchange reactions on positively charged surface sites. An example of Equation 16.1 of importance to sodicity and the physical properties of soils is the heterovalent Na–Ca exchange reaction:

$$2XNa_{(s)} + Ca_{(aq)}^{2+} = X_2Ca_{(s)} + 2Na_{(aq)}^+. \tag{16.2}$$
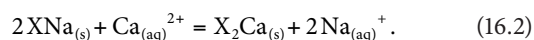
Figure 16.1 shows an experimental *ion exchange isotherm* for the binary Na–Ca exchange reaction on a montmorillonitic soil (Fletcher and Sposito, 1984a), plotted as the fractional contribution of Ca^{2+} to the total adsorbed charge of exchangeable cations ($E_{Ca} = q_{Ca}/Q$, where $q_i = z_in_i$ is the adsorbed charge of species i in mol_c kg^{-1} of solid, z_i and n_i being the valence of i and the moles of adsorbed i per kilogram of solid, respectively, and $Q = \Sigma_iq_i$) against its fractional contribution to the charge concentration of all cations in solution ($\tilde{E}_{Ca} = z_{Ca}C_{Ca}/\tilde{Q}$, where C_i is the molar concentration of species i [mol dm^{-3}] and $\tilde{Q} = \sum z_iC_i$). The convexity of the isotherm in Figure 16.1 is typical of the competitive adsorption of the higher-valence ion in heterovalent exchange reactions. As shown below, however, this convexity does not imply selectivity or thermodynamic preference of the solid exchanger for Ca^{2+} vs. Na^+.

For practical applications, ion exchange isotherms are fitted with a variety of empirical one- or two-parameter models. The most widely used one-parameter models are those introduced by Vanselow (1932), Gapon (1933), and Gaines and Thomas (1953):

$$^VK_A^B = \frac{x_B^a}{x_A^b} \frac{(A^{a+})^b}{(B^{b+})^a} \quad \text{(Vanselow)}, \tag{16.3}$$
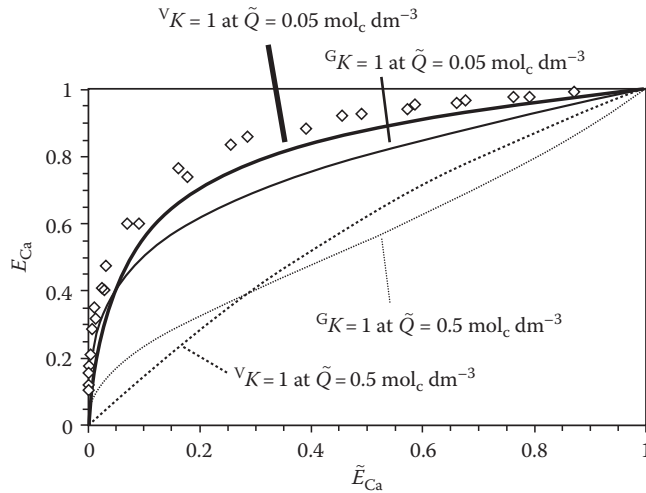
FIGURE 16.1 Binary Na–Ca ion exchange isotherm on a montmorillonitic soil at $\tilde{Q} = 0.05\,\text{mol}_c\,\text{dm}^{-3}$ (diamonds, Fletcher et al. (1984b)) plotted as E_{Ca} vs. $\tilde{E}_{Ca}$. Nonselective isotherms were calculated for the Vanselow (thick lines) and Gapon conventions (thin lines) at $\tilde{Q} = 0.05$ and $0.5\,\text{mol}\,\text{dm}^{-3}$ (solid lines and dashed lines, respectively).

$$^{G}K_{A}^{B} = \frac{E_{B}}{E_{A}}\frac{(A^{a+})^{1/a}}{(B^{b+})^{1/b}} \quad \text{(Gapon)}, \tag{16.4}$$

$$^{GT}K_{A}^{B} = \frac{E_{B}^{a}}{E_{A}^{b}}\frac{(A^{a+})^{b}}{(B^{b+})^{a}} \quad \text{(Gaines and Thomas)}, \tag{16.5}$$

where

- ^{V}K, ^{G}K, and ^{GT}K are the Vanselow, Gapon, and Gaines–Thomas *selectivity coefficients*, respectively
- (A^{a+}) is the thermodynamic activity of A^{a+} in aqueous solution
- x_{A} is the fractional contribution of A^{a+} to the total number of moles of exchangeable cations ($x_{A} = n_{A}/\Sigma_{i}n_{i}$)

By definition, nonselective (or nonpreference) Na–Ca binary exchange isotherms are obtained if $^{V}K = 1$, $^{G}K = 1$, or $^{GT}K = 1$; examples are plotted in Figure 16.1 for $\tilde{Q} = 0.05\,\text{mol}\,\text{dm}^{-3}$ (solid lines) and $\tilde{Q} = 0.5\,\text{mol}\,\text{dm}^{-3}$ (dashed lines). Clearly, despite its name, nonselective exchange can differ among models and produce a strong, $\tilde{Q}$-dependent adsorption of the higher-valence ion; caution should therefore be used in assigning underlying mechanistic significance to isotherm shapes or selectivity coefficient values.

The most popular two-parameter model is the Rothmund–Kornfeld type model, in which the ratio of solute activities in Equations 16.3 through 16.5 is raised to a fitted power n (Bond, 1995). The Rothmund–Kornfeld model based on Equation 16.3, for example, is (Bond, 1995):

$$^{VRK}K_{A}^{B} = \frac{x_{B}^{a}}{x_{A}^{b}}\left[\frac{(A^{a+})^{b}}{(B^{b+})^{a}}\right]^{n}. \tag{16.6}$$

16.2 Surface Charge and Ion Exchange Capacities

As noted above, ion exchange phenomena involve adsorption reactions that balance the intrinsic surface charge of soil particles. The intrinsic surface charge density σ_{in} ($\text{mol}_c\,\text{kg}^{-1}$) is the sum of the net structural surface charge density σ_0 ($\text{mol}\,\text{kg}^{-1}$) and the net proton surface charge density σ_H ($\text{mol}\,\text{kg}^{-1}$) (Sposito, 1998, 2008):

$$\sigma_{in} \equiv \sigma_0 + \sigma_H. \tag{16.7}$$

The net structural ("permanent") surface charge density σ_0 results from crystalline defects, such as isomorphic substitutions of Si(IV), Al(III), or Mg(II) by lower-valence cations in 2:1 phyllosilicates (smectites, vermiculites, illites, and micas; Figure 16.2a); for these 2:1 phyllosilicates, σ_0 contributes dominantly to σ_{in} and is invariably negative. The net proton ("variable") surface charge density σ_H (the difference between the moles of protons and the moles of hydroxide ions complexed by surface functional groups ($\sigma_H = q_H - q_{OH}$)) results from Brønsted acid surface groups with a pH-dependent charge, such as hydroxyl, carboxyl, or phenol groups (Figure 16.2b); it is predominant in natural organic matter, kaolinite, and oxide minerals and can be negative, zero, or positive depending on pH, ionic strength, and other conditions.

Charge balance at solid–water interfaces imposes that Δq, the sum of the adsorbed ion charge densities q_i of all species except surface-complexed H^+ and OH^- ions, equals the opposite of the intrinsic surface charge density (Sposito, 1998, 2008):

$$\sigma_0 + \sigma_H + \Delta q = 0. \tag{16.8}$$

The net adsorbed ion charge Δq can be expressed as a sum of the net charge of ions adsorbed in the Stern layer (σ_S) (ions immobile on timescales >10 ps [Sposito et al., 1999]) or in the diffuse ion swarm (σ_d). The Stern layer charge component can be further divided into the contributions of inner-sphere (σ_{IS}) and outer-sphere (σ_{OS}) surface complexes (formed by direct contact of surface functional groups or through one or more interposed water molecules, respectively) to yield the expression (Sposito, 1998, 2008):

$$\Delta q = \sigma_{IS} + \sigma_{OS} + \sigma_d. \tag{16.9}$$

The utility of Equation 16.9 depends on the extent to which the molecular-scale coordination of adsorbed ions can be determined. Equations 16.7 and 16.8 show that σ_{in} and Δq can vary significantly with pH, ionic strength, and other variables that influence σ_H, especially in soils with low permanent structural charge (i.e., soils poor in 2:1 phyllosilicates). Experimental data on Δq vs. σ_H for a kaolinitic tropical soil suspended in
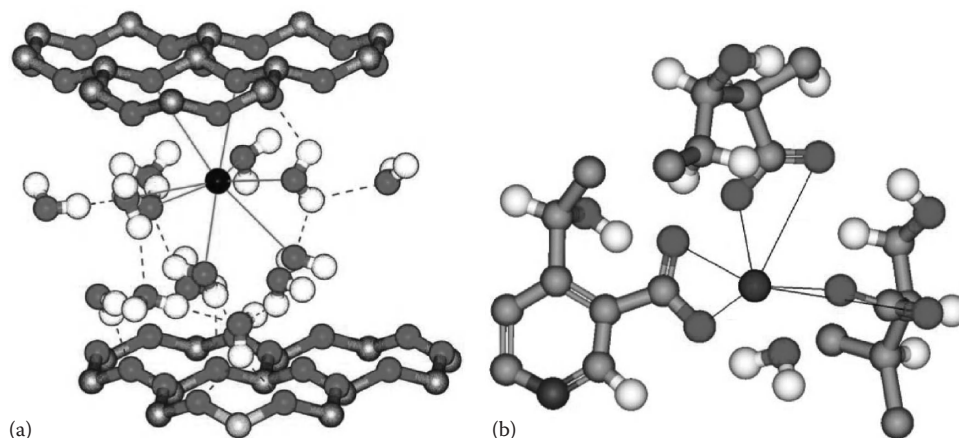
FIGURE 16.2 Major types of ion exchange sites in natural materials: (a) clay siloxane surfaces located near a site of isomorphic substitution (shown only as the basal surface O atoms (matte gray) and tetrahedral sheet Si atoms (shiny gray) of two stacked smectite lamellae with a K^+ ion (black, forming an inner-sphere surface complex) and nearby water molecules in the interlayer space. (From Sposito, G., N.T. Skipper, R. Sutton, S.-H. Park, A.K. Soper, and J.A. Greathouse. 1999. Surface geochemistry of the clay minerals. Proc. Natl. Acad. Sci. U. S. A. 96:3358–3364. Copyright 1999 National Academy of Sciences, U.S.A.) (b) pH-dependent sites resulting from the deprotonation of Lewis acid groups on organic matter or mineral oxide surfaces (here three carboxyl groups of natural organic matter (with O atoms in matte gray) coordinating an almost completely desolvated Ca^{2+} ion (central dark gray atom)). (From Sutton, R., G. Sposito, M.S. Diallo, and H.-R. Schulten. 2005. Molecular simulation of a model of dissolved organic matter. Environ. Toxicol. Chem. 24:1902–1911. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. With permission.)

LiCl solutions of varying ionic strength and pH (Figure 16.3) confirm the inverse relationship between Δq and σ_H in Equation 16.8 and the strong dependence of intrinsic surface charge density on experimental conditions. Thus, measured ion exchange isotherms and Q values for such variable-charge soils may be highly sensitive to pH and other conditions that determine σ_H.

The pH dependence of surface charge is characterized by points of zero charge, pH values at which one or more of the surface charge components in Equations 16.8 and 16.9 vanishes at fixed temperature, applied pressure, and aqueous solution composition (Sposito, 1998, 2008). For example, the pH value at which $\sigma_{in} = 0$ (where adsorbed cation and anion charge densities are equal according to Equations 16.7 and 16.8) is the point of zero net charge (p.z.n.c.). Nomenclature for these points of zero charge is listed in Table 16.1. Unfortunately, previous terminology for the points of zero charge has been highly erratic (e.g., both the p.z.s.e. and the p.z.n.c. have been termed points of zero charge, while the p.z.n.p.c. has been termed zero point of charge). Furthermore, the points of zero charge frequently have been indirectly determined from experimental data on electrophoretic mobility or particle flocculation (calculations sensitive

FIGURE 16.3 "Chorover plot" (Chorover and Sposito, 1995) of the net adsorbed ion charge against the net proton surface charge density for a Brazilian oxisol (Manaus soil). The combined data, for ionic strengths of 0.001 (open circles), 0.005 (crosses), and 0.01 (filled circles) mol kg^{-1}, can be fit to the regression equation (solid line): $\Delta q = -1.01(\pm 0.07)\sigma_H + 12.5(\pm 0.8)$, where Δq and σ_H are in mmol$_c$ kg^{-1}. Charge balance is confirmed by the values of the slope and both intercepts ($\sigma_0 = 12.5 \pm 0.4$ mmol$_c$ kg^{-1} in direct measurement). (From *Geochim. Cosmochim. Acta*, 59, Chorover, J. and Sposito, G., Surface charge characteristics of kaolinitic tropical soils, 875–884, Copyright 1995, with permission from Elsevier.)

TABLE 16.1 Points of Zero Charge

Symbol	Name	Definition
p.z.n.c.	Point of zero net charge	$\sigma_{in} = 0$
p.z.n.p.c.	Point of zero net proton charge	$\sigma_H = 0$
p.z.c.	Point of zero charge	$\sigma_p = 0$[a]
p.z.s.e.	Point of zero salt effect	$\partial\sigma_H/\partial I = 0$

[a] σ_p ($\equiv \sigma_0 + \sigma_H + \sigma_{IS} + \sigma_{OS}$) is the net particle surface charge density.

TABLE 16.2 Typical Cation Exchange Capacities (CEC, in mol_c kg^{-1}) of Soils and Major Soil Constituents

Solid	CEC
Soils	0.01–1.4
Natural organic matter	0.8–10[a]
Vermiculite	1.6–2.5
Smectite	0.7–1.7
Illite	0.2
Mica	0.1
Kaolinite	<0.05

[a] Increasing with pH and in the order peat < humic acid.

Source: Sposito, G. 2004. The surface chemistry of natural particles. Oxford University Press, Oxford, U.K.; Sposito, G. 2008. The surface chemistry of soils. 2nd edn. Oxford University Press, Oxford, U.K.

to the assumed distribution and mobility of ions in the electrical double layer [Fair and Anderson, 1989; Hunter, 1993]) or from acid–base titrations of solid suspensions (calculations that often use untested assumptions on the initial value of the net proton surface charge [Sposito, 1998; Bourg et al., 2007]).

If all adsorbed ions (except surface-complexed H^+ and OH^-) are readily exchangeable, then Δq = CEC − AEC, where CEC is the cation exchange capacity (equal to Q plus the usually small positive equivalent adsorbed charge contributed by anion exclusion from the vicinity of X^- surface sites [Sposito, 2008]) and AEC is the anion exchange capacity (defined equivalently for anion exchange on positively charged surface functional groups). Ranges of the CEC of soils and soil constituents are listed in Table 16.2. The AEC of soils is usually less than 0.05 mol_c kg^{-1} (Sposito, 2008). Major contributors to the CEC of soils are the widely studied smectite clay minerals (Sposito et al., 1999; Sposito, 2008) and the less well-characterized soil organic matter (Helling et al., 1964; Curtin et al., 1998; Sutton and Sposito, 2005; Sposito, 2008). Micaceous minerals (illite, mica) also play an important role in the uptake of small quantities of weakly solvated ions, such as K^+, NH_4^+, and Cs^+ (Maes and Cremers, 1986; Bradbury and Baeyens, 2000).

16.3 Ion Exchange Thermodynamics

If Equation 16.1 describes a true chemical equilibrium and the exchanger sites X^- are all identical (or are taken to represent an average site), a thermodynamic equilibrium constant or exchange equilibrium coefficient ^{ex}K can be defined with the relation (Sposito, 1994):

$$^{ex}K_A^B = \frac{(X_bB)^a(A^{a+})^b}{(X_aA)^b(B^{b+})^a},$$ (16.10)

where (i) represents the thermodynamic activity of species i. The equilibrium constant ^{ex}K is directly related to the difference between the standard-state chemical potentials $\mu^0[\ldots]$ of the products and reactants in Equation 16.1, termed the standard Gibbs energy change of the reaction:

$$\Delta_r G^0 = -RT \ln{^{ex}K_A^B} = a\mu^0[X_bB] + b\mu^0[A^{a+}] - b\mu^0[X_aA] - a\mu^0[B^{b+}],$$

where

 R is the molar gas constant
 T is the absolute temperature

The equilibrium chemical potentials μ of reactants and products typically differ from their standard-state values μ^0, and this difference must enter Equation 16.10 through the thermodynamic activities on its right side. For the aqueous species A^{a+} and B^{b+}, the activity is defined by setting

$$(i) \equiv \gamma_i C_i,$$ (16.11)

where γ_i is an activity coefficient (dm^{-3} mol), commonly expressed relative to the infinite dilution reference state at $T = 298.15\,K$ and $P = 1\,atm$ with the semiempirical Davies equation (Sposito, 1994):

$$\ln\gamma_i = -0.512z_i^2\left\{\frac{\sqrt{I}}{1+\sqrt{I}} - 0.3I\right\} \quad (I < 0.5\,mol\,dm^{-3})$$ (16.12)

where I is the ionic strength ($I = 1/2\sum_i z_i^2 C_i$). For adsorbed species, if the exchanger phase is pictured as analogous to a solid solution of two components X_aA and X_bB, an appropriate model of thermodynamic activity should be (Argersinger et al., 1950; Sposito, 1994)

$$(i) \equiv f_i x_i,$$ (16.13)

where the rational activity coefficient f_i (dimensionless) is equal to 1 in the conventionally chosen reference state $x_i = 1$, $I = 0$, $T = 298.15\,K$, and $P = 1\,atm$ (Gaines and Thomas, 1953; Sposito, 1994). No model for f_i of similar applicability and simplicity to Equation 16.12 currently exists for f_i. However, Equations 16.3, 16.10, and 16.13, and the Gibbs–Duhem relation at fixed T, P $[x_A d\ln f_A + x_B d\ln f_B = 0]$ yield closed-form expressions for calculating ^{ex}K and f_i from experimental ion exchange isotherms at fixed $\tilde{Q}$ (Argersinger et al., 1950; Sposito, 1994):

$$\ln{^{ex}K_A^B} = \int_0^1 \ln{^V K_A^B}\, dE_B,$$ (16.14)

$$b\ln f_A = E_B \ln{^V K_A^B} - \int_0^{E_B} \ln{^V K_A^B}\, dE_B.$$ (16.15)

In the context of the solid-solution picture, the A–B exchange is defined as *ideal* if $f_A = f_B = 1$, that is, if $^VK_A^B = {}^{ex}K_A^B$, and *nonpreference* if $^{ex}K_A^B = 1$. The Vanselow model in Equation 16.3 therefore describes ideal binary exchange and the nonselective isotherms with $^VK = 1$ in Figure 16.1 are *ideal thermodynamic nonpreference isotherms*. Binary exchange ^{ex}K values should obey the "triangle rule": $c \log {}^{ex}K_A^B + a \log {}^{ex}K_B^C + b \log {}^{ex}K_C^A = 0$, as has been verified within ±0.1 log units for montmorillonite (Lewis and Thomas, 1963; Gast, 1969), vermiculite (Wild and Keay, 1964), and illite (Brouwer et al., 1983).

Since the reference state for surface species includes the ionic strength condition $I = 0$, Equations 16.14 and 16.15 are strictly valid only if applied to $^VK_A^B$ vs. E_B data measured at several $\tilde{Q}$ values and then extrapolated to $\tilde{Q} = 0$ (Gaines and Thomas, 1953). In practice, this extrapolation is rarely done, but ^VK values for exchange reactions on smectites and soils have been shown to have a rather small $\tilde{Q}$-dependence (≤0.1 log units) if $\tilde{Q} \le 0.2 \, mol_c \, dm^{-3}$ (Laudelout et al., 1972; Jensen and Babcock, 1973). Furthermore, the rational activity coefficients f_i calculated with Equation 16.15 are strictly valid only for the binary systems in which they were measured; expressions for f_i in ternary or more complex exchange systems are much more complicated (Chu and Sposito, 1981; Sposito, 1994). However, several models of comparable accuracy have been proposed for estimating ternary-system activity coefficients from binary ion exchange isotherm data (Bond and Verburg, 1997).

16.4 Trends in ^VK and ^{ex}K

Broad syntheses of ion exchange data for natural materials are scarce, despite the large number of reported experimental studies. This scarcity results in part from the complexity of soils. Even for smectite minerals (the most widely studied soil constituent), analyses of ion exchange data are complicated by the difficulty of isolating these clays (exchangeable cation homogenization and removal of carbonate, organic matter, and Al and Fe hydroxide impurities require a careful choice of sample pretreatment and storage procedures [Duc et al., 2005]) and of accurately calculating the selectivity coefficient (^VK calculated without measuring C_i and q_i for all competing ions may be highly imprecise [Pabalan and Bertetti, 1999]). In addition, the measured selectivity coefficients may vary with experimental conditions, such as $\tilde{Q}$, pH, solid–liquid ratio m_s, or type of background anion because of poorly understood processes, such as ion-pair adsorption (Sposito et al., 1983a, 1983b; Griffoen and Appelo, 1993; Charlet and Tournassat, 2005), adsorption on variable-charge sites on the edge surfaces of smectite lamellae (Fletcher and Sposito, 1989; Chen and Hayes, 1999), or the influence of experimental conditions on exchanger structure (Laird and Shang, 1997).

In this section, we summarize current knowledge of the ion exchange selectivity of smectites, micaceous minerals, and soils based on reported ^VK and ^{ex}K values. In the case of smectites, we base our analysis as much as possible on studies that (1) used solid pretreatment practices known to produce pure, homoionic

materials (a series of several acid washes [pH ≈ 4], exchangeable cation homogenization [$I ≈ 1$ M], and rinsing steps followed by storage in liquid water at low temperature [Duc et al., 2005; Bourg et al., 2007]), and (2) measured q_i and C_i for all competing ions (Table 16.3). For the sake of brevity, we focus on results obtained at $T ≈ 298$ K and do not discuss the temperature dependence of ^VK or the estimation of enthalpic and entropic contributions to the Gibbs energy of exchange (Gast, 1972; Maes and Cremers, 1978; Morel et al., 2007).

16.4.1 Smectites

Ion exchange reactions on smectites that involve only strongly hydrated cations (Li$^+$, Na$^+$, and divalent metal cations M^{2+}) have ^VK values that display no hysteresis (Verburg and Baveye, 1994) and are independent of E_i within 0.1 log units (Gast, 1969; Sposito et al., 1981, 1983a, 1983b, 1983c; Tang and Sparks, 1993; Zhang and Sparks, 1996), that is, they are *ideal* within experimental precision. (Erroneous reports of "nonideal" heterovalent exchange reactions have been based on plots of $^{GT}K_A^B$ vs. E_B [Banin, 1968; Keren, 1979; McBride, 1980]; for example, the significant dependence of $^{GT}K_{Na}^M$ on E_M [M = Cu, Cu, Ni, or Zn] observed for montmorillonite is in fact consistent with an ideal exchange reaction since $^VK_{Na}^M$ is independent of E_M [Sposito and Mattigod, 1979].) Exchange reactions between strongly hydrated cations are mildly selective (i.e., slightly favor the adsorption of cations of larger ionic radius or larger valence) and weakly affected by the type of smectite (Table 16.4). As expected from the near ideality and weak selectivity of cation exchange reactions on smectites in the absence of weakly hydrated ions, chemical speciation in such systems can be reasonably well described using log $K_V ≈ 0$ for all ion exchange reactions, for example, the systems Na–H (Tournassat et al., 2004), Na–Cd, and Na–UO$_2$–UO$_2$(OH)–(UO$_2$)$_3$(OH)$_5$ (Zachara and McKinley, 1993).

Binary reactions that involve at least one weakly hydrated cation (K$^+$, NH$_4^+$, Rb$^+$, Cs$^+$, large organic cations, and, to a smaller extent, Ba^{2+}) are typically *nonideal*, and they frequently display hysteresis (Verburg and Baveye, 1994; Laird and Shang, 1997; Chatterjee et al., 2008) along with a strong dependence of ^VK on E_i (Gast, 1969; Maes and Cremers, 1978; McBride, 1979; Shainberg et al., 1987; Amrhein and Suarez, 1991), especially if the exchanging cations have very different hydration energies. This nonideality may result in part (but not entirely [Maes and Cremers, 1978; Laird and Shang, 1997]) from the greater selectivity of *octahedral-charge sites* vs. *tetrahedral-charge sites* (i.e., sites resulting from isomorphic substitutions in the octahedral or tetrahedral sheet, respectively, of phyllosilicate minerals) for weakly hydrated cations (Xu and Harsh, 1992; Onodera et al., 1998). Among alkali metals and organic cations, the adsorption selectivity of weakly hydrated cations increases with ionic radius (Gast, 1972; Maes and Cremers, 1986; Teppen and Aggarwal, 2007) because larger, less strongly hydrated ions have lower affinity for the aqueous solution phase (Mizutani et al., 1995; Teppen and Miller, 2006; Teppen and Aggarwal, 2007). Selectivity also increases

TABLE 16.3   Compilation of the Best Available Data Sets on Cation Exchange on Smectite Clay Minerals

Reference	Solid	Cations	Comments
Gast (1972)	Arizona montmorillonite	Na–Li, Na–K, Na–Rb, and Na–Cs	$\tilde{Q} = 1\,\mathrm{mmol_c\,dm^{-3}}$ (Cl$^-$ electrolyte); $T = 298\,\mathrm{K}$; m_s *not specified*; q_i *values not measured*
Maes and Cremers (1977)	Otay and RCCB[a] montmorillonites	Na–Ca	$\tilde{Q} = 10\,\mathrm{mmol_c\,dm^{-3}}$ (Cl$^-$ electrolyte); $T = 298\,\mathrm{K}$; $m_s \approx 10\,\mathrm{g\,kg^{-1}}$; q_i *values not measured*
Maes and Cremers (1978)	Otay and RCCB montmorillonites, hectorite	Na–Cs	$\tilde{Q} = 10\,\mathrm{mmol\,dm^{-3}}$ (Cl$^-$ electrolyte); $T = 298\,\mathrm{K}$; m_s *not specified*; pH = 5.5–6; *clay stored in freeze-dried form*; q_i *values not measured*
Sposito et al. (1981)	Wyoming montmorillonite	Na–Cu	$\tilde{Q} = 10\,\mathrm{mmol_c\,dm^{-3}}$ (Cl$^-$ or ClO$_4^-$ electrolytes); $T = 298\,\mathrm{K}$; $m_s = 13$–$21\,\mathrm{g\,kg^{-1}}$; pH = 5–6 to avoid Cu adsorption on oxide-type edge surface sites
Sposito et al. (1983a, 1983b, 1983c)	Wyoming montmorillonite	Na–Ca, Na–Mg, Ca–Mg, and Na–Ca–Mg	$\tilde{Q} = 50\,\mathrm{mmol_c\,dm^{-3}}$ (Cl$^-$ or ClO$_4^-$ electrolytes); $T = 298\,\mathrm{K}$; $m_s = 20$–$30\,\mathrm{g\,kg^{-1}}$; pH = 6.8–7.1
Xu and Harsh (1992)	Cameron montmorillonite	Na–Li, Na–K, Na–Rb, and Na–Cs	$\tilde{Q} = 10\,\mathrm{mmol_c\,dm^{-3}}$ (Cl$^-$ electrolyte); $T = 298\,\mathrm{K}$; m_s *not specified*; C_i and q_i *values not reported*; ^VK *values reported only at* $E_{Na} = 0.5$
Tang and Sparks (1993)	Wyoming montmorillonite	Na–Ca and K–Ca	$I = 10\,\mathrm{mmol\,dm^{-3}}$ (Cl$^-$ electrolyte); $T = 296\,\mathrm{K}$; $m_s = 10.2\,\mathrm{g\,dm^{-3}}$; pH = 6.8; *clay stored in freeze-dried form*
Zhang and Sparks (1996)	Wyoming montmorillonite	Na–Cu	$\tilde{Q} = 20\,\mathrm{mmol\,dm^{-3}}$ (ClO$_4^-$, Cl$^-$, NO$_3^-$, or SO$_4^{2-}$ electrolyte); $T = 298\,\mathrm{K}$; $m_s \approx 16$–$17\,\mathrm{g\,dm^{-3}}$; pH = 5.2–6.5, *decreasing with* q_{Cu}; *clay stored in freeze-dried form*
Laird and Shang (1997)	Synthetic fluorohectorite	Mg–Ba	$\tilde{Q} = 20\,\mathrm{mmol\,dm^{-3}}$ (Cl$^-$ electrolyte); T = room temperature; $m_s = 10\,\mathrm{g\,dm^{-3}}$; no acid wash, but the material (a synthetic magnesium silicate) should not contain Al or Fe hydroxides or organic impurities; *clay stored in freeze-dried form*
Charlet and Tournassat (2005)	Wyoming montmorillonite	Na–Fe(II), Ca–Fe(II), Na–Ca–Fe(II)	$\tilde{Q} = 50$ or $130\,\mathrm{mmol_c\,dm^{-3}}$ (Cl$^-$ electrolyte); T *not specified*; $m_s = 4$–$8\,\mathrm{g\,dm^{-3}}$; pH = 2.1–3.6 to avoid Fe adsorption on oxide-type edge surface sites; q_{Fe} *not measured (estimated by mass balance)*

[a] RCCB montmorillonite: reduced-charge Camp-Berteau (Morocco) montmorillonite. Layer charge was reduced by 5%–41% using the Hofmann–Klemen effect (Maes and Cremers, 1977, 1978).

The studies by Gast (1972) and Maes and Cremers (1977, 1978) are tentatively included, despite the fact that they did not include measurements of q_i values, because of the good quality of their pretreatment methods. Studies that used smectites stored by freeze-drying also are tentatively included, although storage in liquid water at low temperature is preferred (Duc et al., 2005). Experimental procedures that may adversely affect the quality of experimental results are reported in italics in the last column.

with surface charge density (Gast, 1972; Maes and Cremers, 1978; Shainberg et al., 1987; Xu and Harsh, 1992) (Table 16.4) and perhaps especially with the charge density of octahedral-charge sites (Xu and Harsh, 1992).

Metal–ligand complexes also can adsorb by ion exchange, but this process has not been deeply studied. The strong adsorption of copper-ethylenediamine complexes $\mathrm{Cu(en)_2^{2+}}$ (Maes and Cremers, 1986) and the complexation of nitroaromatic compound with exchangeable cations on smectite surfaces (Chatterjee et al., 2008) are consistent with the expectation that metal–organic complexes should have an affinity similar to large organic cations for smectite surfaces. Divalent metals are known to coadsorb with anions, such as Cl$^-$ (Sposito et al., 1983a, 1983b; Charlet and Tournassat, 2005) or HCO$_3^-$ (Griffoen and Appelo, 1993), but the mechanism of this coadsorption (ion-pair formation or diffuse layer process) and its dependence on experimental conditions are not well understood (Sposito, 1991).

16.4.2 Micaceous Minerals

Illites and micas have a lamellar morphology similar to that of smectites but with collapsed, K$^+$-filled interlayers (Sposito, 2008).

The reactivity of external basal surfaces of illite and mica crystals is thought to be similar to that of smectites (Brouwer et al., 1983), but the crystals also carry a small population of sites (0.1%–10%) with a very high affinity for weakly hydrated cations (K$^+$, NH$_4^+$, Rb$^+$, and Cs$^+$) (Brouwer et al., 1983; Thellier and Sposito, 1989; Liu et al., 2004; Tournassat et al., 2007). These so-called *frayed-edge sites* occur at partially propped-open, K$^+$-depleted edges of collapsed interlayers (Rajec et al., 1999; McKinley et al., 2004). Adsorption on frayed-edge sites may be determined to a certain extent by nonequilibrium processes, such as diffusion-controlled adsorption over timescales of weeks or more (Comans et al., 1991). Such processes can cause difficulties in defining and measuring ion exchange selectivity, since the population of "accessible" frayed-edge sites may vary with time, experimental conditions, the type of exchangeable cation, and the choice of agent used to extract adsorbed cations for measuring q_i (Brouwer et al., 1983; Comans et al., 1991; Baeyens and Bradbury, 2004; Tournassat et al., 2007). Nevertheless, the finding that frayed-edge sites obey the "triangle rule" (Brouwer et al., 1983) suggests that ion exchange on these sites can be reasonably well described on certain timescales as a thermodynamic process. If ion B = K, Rb, or Cs, experimental $^VK_A^B$ values decrease sharply with E_B

TABLE 16.4 Recommended Values for ^VK on Smectite Based on the Studies Listed in Table 16.3

Variable	log K	References
Exchanges of two strongly hydrated cations		
$^VK_{Na}^{Li}$	-0.08 ± 0.06	Gast (1972), Xu and Harsh (1992)
$^VK_{Na}^{M(II)}$	0.11 ± 0.11^a (M = Mg, Ca, and Cu)	Maes and Cremers (1977), Sposito et al. (1981, 1983a), Tang and Sparks (1993), Zhang and Sparks (1996)
$^VK_{Mg}^{Ca}$	0.01 ± 0.02 (E_{Na} = 0, 0.016, or 0.036)	Sposito et al. (1983b, 1983c)
$^VK_{Fe(II)}^{Ca}$	~0.01	Charlet and Tournassat (2005)[b]
Exchanges involving at least one weakly hydrated cation		
$^{ex}K_{Na}^{K}$	0.58 ± 0.14^c	Gast (1972), Xu and Harsh (1992)
$^{ex}K_{Na}^{Rb}$	1.08 ± 0.25^c	Gast (1972), Xu and Harsh (1992)
$^{ex}K_{Na}^{Cs}$	$1.23 \times CEC - 0.06^d$	Gast (1972), Maes and Cremers (1978), Xu and Harsh (1992)
$^VK_{Mg}^{Ba}$	~0.5 in the three-layer hydrate, increasing to ~1.4 in the two-layer hydrate	Laird and Shang (1997)

Confidence intervals calculated as $\pm 2\sigma$ where several ^VK values were available.

[a] Neglecting Na–Ca exchange data obtained with reduced-charge montmorillonites, which show $^VK_{Na}^{Ca}$ values closer to 0 (Maes and Cremers, 1977).

[b] Charlet and Tournassat (2005) erroneously reported their ^{GT}K values as ^VK values, but for homoionic Ca–Fe(II) exchange $^{GT}K = {}^VK$.

[c] log $^{ex}K_{Na}^{K}$ and log $^{ex}K_{Na}^{Rb}$ are thought to increase with Q (Shainberg et al., 1987; Xu and Harsh, 1992).

[d] Linear regression with $r^2 = 0.91$, $p < 0.001$; CEC = 0.66–$1.44 \, mol_c \, kg^{-1}$, estimated from the number of octahedral and tetrahedral substitutions in the clay unit cell formula.

when the frayed-edge sites become B-saturated (Brouwer et al., 1983), a behavior that can be described with a two- or three-site model, that is, one or two types of highly selective frayed-edge site plus weakly selective basal surface sites (Maes and Cremers, 1986; Liu et al., 2004; Tournassat et al., 2007).

16.4.3 Soils and Sediments

For highly heterogeneous, multiphase media such as soils and sediments, Equations 16.14 and 16.15 have no clear thermodynamic meaning and any selectivity model may be suitable on a case-by-case basis. For example, the Gapon selectivity coefficient has been found to be less variable than ^VK over broad ranges of exchanger phase composition for Na–Ca, Mg–K, and Ca–K exchange reactions in some soils (Naylor and Overstreet, 1969; Jensen and Babcock, 1973; Evangelou and Coale, 1987; Feigenbaum et al., 1991; Agbenin and Yakubu, 2006). Experimental data on log $^VK_{Ca}^{M}$ vs. E_M (M = Na, K, or Mg) obtained for a range of conditions of pH, $\tilde{Q}$, and soil type illustrate the variability of the ion exchange selectivities of soils (Figure 16.4). The scatter of log $^VK_{Ca}^{Na}$ values at $E_{Na} \leq 0.2$ in Figure 16.4b may reflect in part the high sensitivity of ^VK at small E_i values to sources of error such as the dissolution of soil materials (Carlson and Buchanan, 1973; Sheta et al., 1981).

Despite the inherent complexity of soil and sediment ion exchange properties, these have been modeled conceptually with some success assuming their exchange properties result from their smectite constituent (Charlet and Tournassat, 2005) or from the sum of their smectite, organic matter (Curtin et al., 1998), and/or micaceous constituents (Bradbury and Baeyens, 2000).

The first type of approximation is exemplified by the ion exchange properties of Amazon river suspended matter in the system Na–K–Ca–Mg, which are essentially identical to those of montmorillonite, even though only half of the CEC of the suspended matter originates from its clay-size fraction (Charlet and Tournassat, 2005). The second type of approximation is well illustrated by the fact that the Mg–Ca exchange selectivity of soils is correlated with their ratio of organic carbon to clay content (Curtin et al., 1998). Curtin et al. (1998) successfully described the Mg–Ca selectivity of several soils with a two-site model where the soil organic matter and clay fractions were represented by log $^VK_{Mg}^{Ca}$ = 0.6 and 0.1, respectively. Similar additive behavior of soil smectites and soil organic matter has been used to interpret the ion exchange selectivity of montmorillonitic soils in the system Na–Ca–Mg (Fletcher et al., 1984a; Sposito and Fletcher, 1985) and may also explain the Ca–Sr selectivity of soils (Juo and Barber, 1969) and the Ca–Cd and Ca–Zn selectivity of acidic soils (Voegelin and Kretzschmar, 2003). Additive behavior of micaceous materials and other soil constituents may explain the strong adsorption of K^+ and Cs^+ at low E_K or E_{Cs} values in numerous soils and sediments (Feigenbaum et al., 1991; Bradbury and Baeyens, 2000; Sinanis et al., 2003; Liu et al., 2004; Agbenin and Yakubu, 2006). The concept that soil ion exchange selectivity is a sum of the selectivities of component phases evidently can be quite powerful (Sposito and Fletcher, 1985; Curtin et al., 1998; Bradbury and Baeyens, 2000; Charlet and Tournassat, 2005), although it neglects nonadditive behavior such as the enhanced (about 0.3 log unit greater) Na–Ca selectivity of smectites when their interlayers contain Al-hydroxide polymers (Keren, 1979; Janssen et al., 2003).
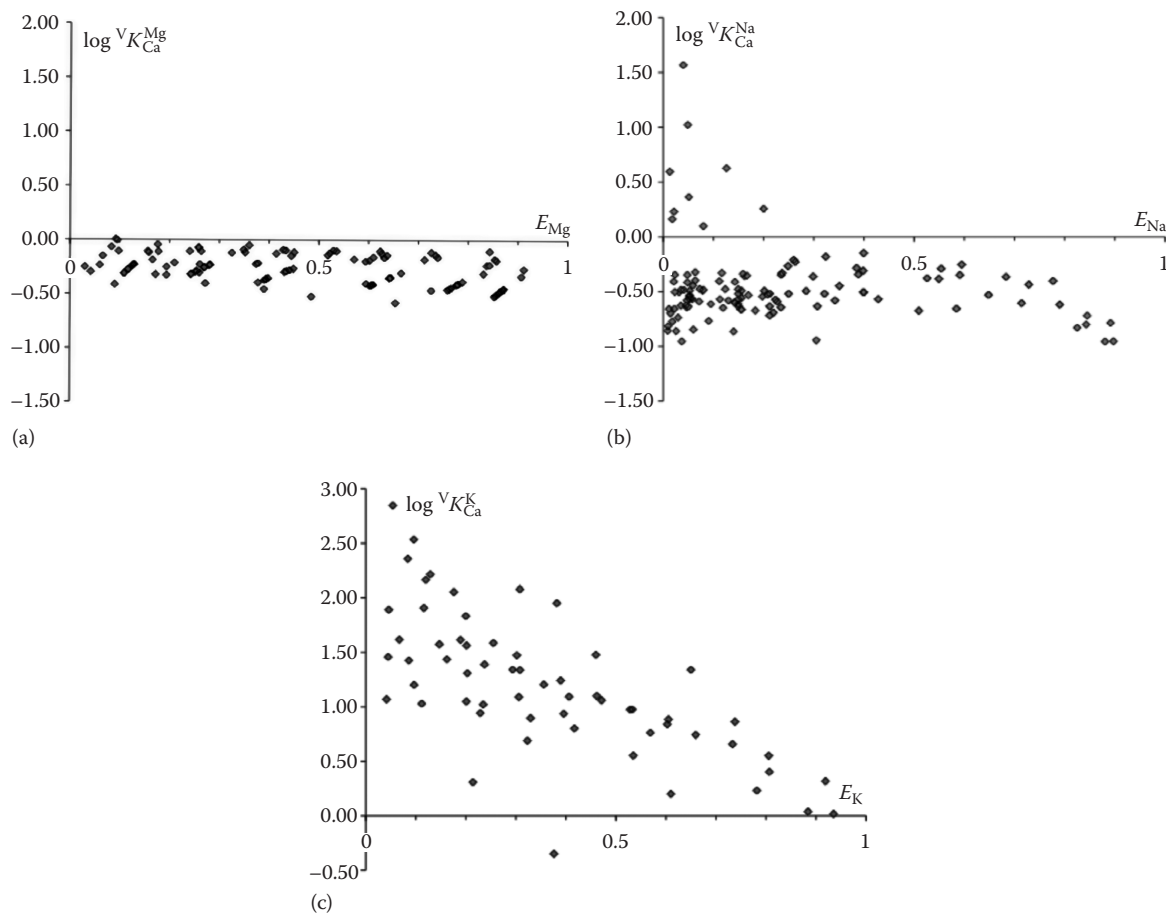
FIGURE 16.4 Compilation of experimental data on $\log {}^{V}K_{Ca}^{M}$ vs. E_M (M = Mg^{2+}, Na^+, or K^+) in soils if $\tilde{Q} \leq 0.2$ mol dm^{-3} and $T = 298$ K; (a) $\log {}^{V}K_{Ca}^{Mg}$ data for a loam soil (Jensen and Babcock, 1973), calcareous clay soils (Van Bladel and Gheyi, 1980), smectitic soils with 27–96 g kg^{-1} organic carbon (Curtin et al., 1998), and a montmorillonitic soil (DeSutter et al., 2006); (b) $\log {}^{V}K_{Ca}^{Na}$ for a montmorillonitic soil (Fletcher et al., 1984b), kaolinitic soils (Levy et al., 1988), a kaolinitic sandy loam soil with 25.7 g kg^{-1} organic carbon (Rhue and Mansell, 1988), a calcareous, smectitic clay soil (Amrhein and Suarez, 1991), and an illite–kaolinite clay soil (Bond, 1995); (c) $\log {}^{V}K_{Ca}^{K}$ for a loam soil (Jensen and Babcock, 1973), kaolinitic soils (Levy et al., 1988), a kaolinitic sandy loam soil with 25.7 g kg^{-1} organic carbon (Rhue and Mansell, 1988), an illite–kaolinite clay soil (Bond, 1995), and a tropical soil (Agbenin and Yakubu, 2006).

16.5 Ion Exchange and Chemical Speciation Models

Analytical solutions of thermodynamic equations to determine exchanger phase composition rapidly become intractable in systems that involve more than three exchangeable ions (Bond and Verburg, 1997). For such complex systems, a numerical solution can be implemented in chemical speciation models after dividing Equation 16.1 into two hypothetical half-reactions involving a fictitious anionic species X^- (Sposito and Mattigod, 1977; Shaviv and Mattigod, 1985; Fletcher and Sposito, 1989):

$$aX^- + A_{(aq)}^{a+} = X_aA_{(s)}, \tag{16.16a}$$

$$bX^- + B_{(aq)}^{b+} = X_bB_{(s)}. \tag{16.16b}$$
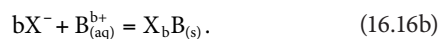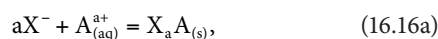
Equation 16.16 can be incorporated into conventional chemical speciation models using its analogy with surface complexation or solid-solution reactions, as described below.

16.5.1 Modeling Ion Exchange as a Surface Complexation Process

Most chemical speciation programs include a model of surface complexation reactions (similar in form to Equation 16.16) where the thermodynamic activities of uncharged surface species arbitrarily are set equal to their concentrations $[X_i i]$ (Sposito, 2004). On this model, the half-reaction selectivity coefficients associated with Equation 16.16 are the following (Fletcher and Sposito, 1989; Stadler and Schindler, 1993):

$$^{half}K^A = \frac{[X_aA]}{(X^-)^a(A^{a+})}, \tag{16.17a}$$

$$^{\text{half}}K^{\text{B}} = \frac{[\text{X}_{\text{b}}\text{B}]}{(\text{X}^-)^{\text{b}}(\text{B}^{\text{b}+})}. \tag{16.17b}$$

Equation 16.17 can be incorporated into any program that models surface complexation reactions; typically, $^{\text{half}}K$ is set to a very large value (10^{10}–10^{20}) for a reference cation, such that the wholly uncomplexed X^- sites contribute negligibly to the total mass balance on X, and the $^{\text{half}}K$ values of other cations are chosen to fit experimental equilibrium constants (Flechter and Sposito, 1989; Stadler and Schindler, 1993). Certain chemical speciation programs include a specific model of ion exchange reactions based on Equation 16.17. For example, in the program PHREEQC2 (Appelo and Postma, 2005), the fictitious species X^- is excluded from the total mass balance on X and Na^+ is chosen as a reference species with log $^{\text{half}}K^{\text{Na}} = 0$. Suggested values of $^{\text{half}}K$ for other cations on smectite are provided in the PREEQC2 database (Appelo and Postma, 2005). The use of Equation 16.17 with fixed $^{\text{half}}K$ values corresponds to an implicit assumption that $^{\text{GT}}K$ is equal to an equilibrium constant, since $^{\text{GT}}K$ and $^{\text{half}}K$ are related through

$$^{\text{GT}}K_{\text{A}}^{\text{B}} = \left[\frac{(^{\text{half}}K^{\text{B}})^{\text{a}}}{(^{\text{half}}K^{\text{A}})^{\text{b}}}\right]\left(\frac{\text{b}^{\text{a}}}{\text{a}^{\text{b}}}\right)Q^{(\text{b}-\text{a})}. \tag{16.18}$$

Equation 16.17 has been used to model ion exchange reactions involving protons, alkali metals, trace metals, and actinides on clay minerals (Stadler and Schindler, 1993; Zachara and McKinley, 1993; Poinssot et al., 1999; Bradbury and Baeyens 2000, 2005, 2009; Baeyens and Bradbury, 2004; Charlet and Tournassat, 2005; Heidmann et al., 2005; Bourg et al., 2007; Gu and Evans, 2008). Most of these studies (with the notable exception of Charlet and Tournassat, 2005) investigated systems where $^{\text{half}}K$ should be constant if ion exchange is ideal, either because the main background electrolyte cation was $\text{A}^{\text{a}+}$ and occupied most ion exchange sites, such that $^{\text{GT}}K_{\text{A}}^{\text{B}} \oplus {}^{\text{V}}K_{\text{A}}^{\text{B}}(\text{b/a})^{\text{a}}$, or because the system studied was homovalent, in which case $^{\text{GT}}K_{\text{A}}^{\text{B}} = {}^{\text{V}}K_{\text{A}}^{\text{B}}$.

16.5.2 Modeling Ion Exchange as a Solid-Solution Process

Some chemical speciation programs include a model of solid-solution formation, which is also similar in form to Equation 16.16. The "solubility coefficients" associated with Equation 16.16 are described as follows in the solid-solution formalism (Appelo and Postma, 2005):

$$^{\text{solubility}}K^{\text{A}} = \frac{(\text{X}^-)^{\text{a}}(\text{A}^{\text{a}+})}{x_{\text{A}}f_{\text{A}}}, \tag{16.19a}$$

$$^{\text{solubility}}K^{\text{B}} = \frac{(\text{X}^-)^{\text{b}}(\text{B}^{\text{b}+})}{x_{\text{B}}f_{\text{B}}}. \tag{16.19b}$$

where (X^-) is the activity of a fictitious dissolved species. In the program PHREEQC2, Equation 16.19 can be solved by defining X^-, $\text{A}^{\text{a}+}$, and $\text{B}^{\text{b}+}$ as dissolved species and $\text{X}_{\text{a}}\text{A}$ and $\text{X}_{\text{b}}\text{B}$ as solid phases and by setting $^{\text{solubility}}K \ll 1$ for a chosen reference cation, such that X^- contributes negligibly to the mass balance on X. If $f_{\text{A}} = f_{\text{B}} = 1$, Equation 16.19 is identical to the Vanselow model with

$$^{\text{V}}K_{\text{A}}^{\text{B}} = \left[\frac{(^{\text{solubility}}K^{\text{B}})^{\text{a}}}{(^{\text{solubility}}K^{\text{A}})^{\text{b}}}\right]. \tag{16.20}$$

The ideality of ion exchange reactions between strongly hydrated cations on smectite (Table 16.4) suggests that Equation 16.19 may be more accurate than Equation 16.17 for multicomponent ion exchange reactions on smectite over broad ranges of exchanger phase composition. However, to our knowledge, Equation 16.19 has never been used for this purpose.

16.6 Micro- and Nanoscale Perspectives on Ion Exchange Selectivity

Micro- and nanoscale studies have yielded insight into the processes that determine $^{\text{V}}K$ using theoretical estimates of long-range electrostatic forces (Barak, 1989; Rytwo et al., 1996), short-range interactions (Shainberg and Kemper, 1966; Eberl, 1980), or statistical mechanical quantities (Sposito, 1993; Benjamin, 2002). In the present chapter, we focus on two subfields illustrating micro- and nanoscale studies: (1) the molecular-scale coordination and dynamics of exchangeable ions and (2) the influence of exchanger microstructure on ion exchange selectivity.

16.6.1 Molecular-Scale Coordination and Dynamics of Exchangeable Ions

Spectroscopic and molecular simulation methods have been widely used to probe the coordination environment and dynamics of exchangeable cations in smectites (Sposito and Prost, 1982; Sposito, 2004; Skipper et al., 2006) and have been increasingly applied to other exchanger phases (Skipper et al., 1995; Kim and Kirkpatrick, 1998; Schlegel et al., 2006; Xu et al., 2006). Spectroscopic techniques can probe the molecular structure of the interface as a function of distance from a solid surface (by x-ray reflectivity [Schlegel et al., 2006; Park et al., 2008]), the dynamics of water H atoms (by quasielastic neutron scattering [QENS] [Marry et al., 2008]), the local coordination environment of certain atomic probes (such as water or Li by neutron diffraction with isotopic substitution [NDIS] [Skipper et al., 1995; Powell et al., 1998] or Co, Sr, Pb, Cu, Cs, and other atoms by x-ray absorption spectroscopy [XAS] [Papelis and Hayes, 1996; Chen and Hayes, 1999; Strawn and Sparks, 1999; Morton et al., 2001; Bostick et al., 2002]), the coordination environment and

rotational dynamics of paramagnetic atoms such as Cu (by electron spin resonance [ESR] or electron spin-echo modulation [ESEM] [McBride et al., 1975; Brown and Kevan, 1988]), and the local molecular environment ("chemical shielding") of atoms with an odd number of nucleons or protons such as ^{23}Na, ^{133}Cs, ^{113}Cd, or ^{35}Cl (by nuclear magnetic resonance [NMR] [Weiss et al., 1990; Kim and Kirkpatrick, 1998; Xu et al., 2006]). Molecular simulation techniques, primarily Monte Carlo (MC) and molecular dynamics (MD), have been used for over a decade to complement spectroscopic results (Chang et al., 1995; Park and Sposito, 2002). Most spectroscopic and simulation studies have been used to probe homoionic exchangers, but a few studies have investigated the behavior of a "reporter" cation doped into a smectite that is almost homoionic in another cation (McBride et al., 1975; Brown and Kevan, 1988; Marry and Turq, 2003; Bourg and Sposito, 2010).

Spectroscopic and molecular simulation studies have shown that the aqueous phase in smectite interlayers is analogous to a concentrated ionic solution (Sposito and Prost, 1982; Powell et al., 1998); in the two- and three-layer hydrates and on external basal surfaces, it diffuses about 30%–40% as fast as bulk liquid water (Chang et al., 1997; Marry et al., 2008; Bourg and Sposito, 2010). Exchangeable cations in smectite one-layer hydrates form inner-sphere surface complexes (ISSC) for obvious steric reasons, and they diffuse very slowly (McBride et al., 1975; Chang et al., 1995, 1997; Bourg and Sposito, 2010); at higher hydration levels, they can be divided into three categories based on their interaction with smectite surfaces: divalent metal cations (M^{2+}), strongly hydrated alkali metal cations (Li^+, Na^+), and weakly hydrated cations (K^+, Rb^+, Cs^+). The divalent cations are adsorbed in fully solvated form as outer-sphere surface complexes (OSSC) or in the diffuse layer (DL) (McBride et al., 1975; Brown and Kevan, 1988; Papelis and Hayes, 1996; Chen and Hayes, 1999; Strawn and Sparks, 1999; Greathouse et al., 2000; Chávez-Páez et al., 2001; Morton et al., 2001; Whitley and Smith, 2004); they tumble and diffuse slowly, about 1%–10% as fast as in bulk water (McBride et al., 1975; Brown and Kevan, 1988; Greathouse et al., 2000; Bourg and Sposito, 2010). The strongly hydrated monovalent cations (Li^+, Na^+) also adsorb as OSSC or DL species on octahedral-charge sites, but they form inner-sphere surface complexes (ISSC) on tetrahedral-charge sites (Chang et al., 1995, 1997; Leote de Carvalho and Skipper, 2001; Marry and Turq, 2003; Marry et al., 2003; Tambach et al., 2004, 2006; Greathouse and Cygan, 2005); they diffuse about 20%–60% as fast as in bulk water if adsorbed on octahedral-charge sites, but are essentially immobile (on subnanosecond timescales) on tetrahedral-charge sites (Chang et al., 1995, 1997; Leote de Carvalho and Skipper, 2001; Marry and Turq, 2003; Bourg and Sposito, 2010). Finally, the weakly hydrated cations form primarily ISSC on smectite surfaces along with small amounts of OSSC or DL species (Chang et al., 1998; Nakano et al., 2003; Whitley and Smith, 2004; Tambach et al., 2006; Liu et al., 2008); two populations of ISSC exist in Cs- and K-smectite interlayers that may correspond to cations located above ditrigonal cavities or "triads" of O atoms of the siloxane surface (Weiss et al., 1990; Onodera

et al., 1998; Park and Sposito, 2002; Nakano et al., 2003); the mobility of these cations along smectite basal surfaces is not well characterized (Kosakowski et al., 2008), but recent results suggest that they diffuse only 5%–8% as fast as in bulk liquid water (Bourg and Sposito, 2010). Clearly, the molecular-scale behavior of cations on smectites parallels their ion exchange selectivity: The strongly hydrated cations (M^{2+}, Li^+, Na^+) adsorb mainly as fully solvated OSSC or DL species and display ideal, weakly selective ion exchange behavior, whereas the weakly hydrated ions adsorb mainly as ISSC and display nonideal, strongly selective ion exchange behavior. The ISSC formed by Li^+ and Na^+ on tetrahedral-charge sites do not fit this simple classification, suggesting that Li^+ and Na^+ may show nonideal exchange behavior on smectites with high tetrahedral charge. This point may not have been noticed previously because reference smectites typically carry 80% ± 20% octahedral-charge sites (Xu and Harsh, 1992).

Thanks to increases in computational capabilities, MD and MC simulations may soon be able to "bridge the gap" from the time and length scales (or number of MC simulation steps) required to probe molecular exchanger phase structure and dynamics (<1 ns and ≤1 nm) to the time and length scales (tens of nanoseconds and nanometers) on which ion exchange equilibria become established (Greathouse and Cygan, 2005; Rotenberg et al., 2007). Molecular simulation methods could then be used to test Equation 16.13 directly or to predict ^{V}K values (Greathouse and Cygan, 2005; Teppen and Miller, 2006). Teppen and Miller (2006) showed that the Gibbs energy difference between K-, Rb-, and Cs-montmorillonite (one component of the Gibbs energy of exchange) at fixed interlayer spacing and interlayer water content could be determined by MD simulation. Greathouse and Cygan (2005) found that ten 1 ns MD simulations of a 4 nm thick Na^+–UO_2^{2+}–CO_3^{2-} aqueous solution on a montmorillonite basal surface were too short and small scale to determine ^{V}K accurately, but were sufficient to identify important processes, such as the formation of $\left[Na_2UO_2(CO_3)_3\right]^{2-}$ complexes.

16.6.2 Coupling between Cation Exchange and Exchanger Structure

As pointed out above, Equation 16.10 has strict thermodynamic meaning only if all exchanger sites X^- are identical. For smectites (Sposito, 1992; Laird and Shang, 1997) and natural organic matter (Sutton et al., 2005), however, the microstructure of the exchanger phase depends on experimental conditions and, therefore, the X^- may *not* be identical at different points along a binary exchange isotherm, or even in different regions of the exchanger phase within a single sample. In the case of smectite clay minerals, the stacking arrangement of smectite lamellae (number of layers per stack, interlayer spacing) is a dynamic, nonuniform property (several interlayer hydration levels may coexist [Tamura et al., 2000; Wilson et al., 2004; Ferrage et al., 2005]), which depends on the magnitude and location of clay structural charge (Slade et al., 1991; Tambach et al., 2004), the

population of adsorbed cations (Schramm and Kwak, 1982; Sposito, 1992; Laird and Shang, 1997; Ferrage et al., 2005; Table 16.5), the thermodynamic activity of water (Norrish, 1954; Slade et al., 1991; Laird et al., 1995), as well as the previous history of the clay (Verburg and Baveye, 1994; Laird et al., 1995; Chatterjee et al., 2008) and its degree of compaction (Kozaki et al., 1998). Molecular-scale simulations confirm that homoionic montmorillonites frequently exhibit several stable states of crystalline swelling separated by energy barriers (Whitley and Smith, 2004; Tambach et al., 2004, 2006; Smith et al., 2006). Transitions between these states are predominantly enthalpic (Whitley and Smith, 2004), driven by cationic solvation energy (Whitley and Smith, 2004)—as opposed to being determined by the hydrogen bond network of interlayer water (Tambach et al., 2006)—strongly hysteretic (Tambach et al., 2006), and in fact thermodynamically analogous to a phase transition (Laird and Shang, 1997; Whitley and Smith, 2004).

The stacking arrangement, in turn, determines the fraction of X^- sites located on external vs. internal basal surfaces, which may
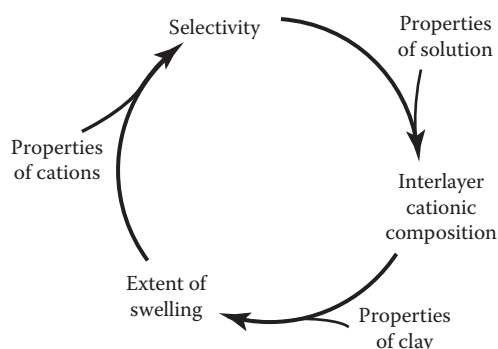
FIGURE 16.5 Conceptual model of the feedback between the ion exchange selectivity (^VK), population of exchangeable ions (E_i), and arrangement of clay particles (interlayer spacing, number of layers per stack) in smectite clay minerals. (Reproduced from Laird, D.A., and C. Shang. 1997. Relationship between cation exchange selectivity and crystalline swelling in expanding 2:1 phyllosilicates. Clays Clay Miner. 45:681–689. With permission of The Clay Minerals Society, publisher of *Clays and Clay Minerals*.)

TABLE 16.5 Interlayer Hydration Level (Norrish, 1954; Laird and Shang, 1997; Chatterjee et al., 2008) and Number of Lamellae per Stack (Sposito, 1992; Verburg and Baveye, 1994; Verburg et al., 1995) in Dilute Aqueous Suspensions of Homoionic Montmorillonite

Cation	Hydration Level[a]	Lamellae per Stack
Li^+	>3	1.0
Na^+	>3	1.0–1.7
K^+	2	1.0–7.0
Cs^+	1–2	1.4–4.0
Mg^{2+}	3	2.7–14.0
Ca^{2+}	3	3.0–20.0
Ba^{2+}	2–3	2.7–7.0

[a] Number of statistical water monolayers in each smectite interlayer.

have unequal cation exchange selectivities (Keren, 1979; Sposito et al., 1983a), and the fraction of internal-surface X^- sites in contact with one-, two- or three-layer interlayer hydrates, which also may have unequal cation exchange selectivities (Barak, 1989; Laird and Shang, 1997; Van Loon and Glaus, 2008). Thus, cation exchange on swelling clay minerals involves a feedback loop among adsorption, exchanger structure, and selectivity (Laird and Shang, 1997; Figure 16.5) that is reminiscent of the behavior of certain ion-selective pores (Kuyucak et al., 2001). This feedback loop may explain the occurrence of exchangeable cation "de-mixing" (Shainberg and Otoh, 1968; Fink et al., 1971; Iwasaki and Watanabe, 1988), increased selectivity for the preferred cation at high surface loading of the same cation (Shainberg et al., 1980; Laird and Shang, 1997; Janssen et al., 2003), cation exchange hysteresis (Fripiat et al., 1965; Verburg et al., 1995), and increased selectivity for weakly hydrated ions as smectite charge density increases (Maes and Cremers, 1978; Shainberg et al., 1987). Smectite ^{ex}K values would then have thermodynamic meaning only if defined for a certain particle arrangement (Laird and Shang, 1997). Thus, the nonideality of cation exchange reactions that involve weakly hydrated cations may be related to the fact that these cations cause a sequential collapse of the interlayer from a three-layer hydrate or more (for Li^+, Na^+, and strongly hydrated divalent cations M^{2+}) to a two-layer hydrate (for K^+ and Ba^{2+}) to a one-layer hydrate (for Cs^+; Table 16.5).

References

Agbenin, J.O., and S. Yakubu. 2006. Potassium-calcium and potassium-magnesium exchange equilibria in an acid savanna soil from northern Nigeria. Geoderma 136:542–554.

Amrhein, C., and D.L. Suarez. 1991. Sodium-calcium exchange with anion exclusion and weathering corrections. Soil Sci. Soc. Am. J. 55:698–706.

Appelo, C.A.J., and D. Postma. 2005. Geochemistry, groundwater and pollution. 2nd edn. Balkema Publications, New York.

Argersinger, W.J., Jr., A.W. Davidson, and O.D. Bonner. 1950. Thermodynamics and ion exchange phenomena. Trans. Kansas Acad. Sci. 53:404–410.

Baeyens, B., and M.H. Bradbury. 2004. Cation exchange capacity measurements on illite using the sodium and cesium isotope dilution technique: Effects of the index cation, electrolyte concentration and competition: Modeling. Clays Clay Miner. 52:421–431.

Banin, A. 1968. Ion exchange isotherms of montmorillonite and structure factors affecting them. Isr. J. Chem. 6:27–36.

Barak, P. 1989. Double layer theory prediction of Al–Ca exchange on clay and soil. J. Colloid Interface Sci. 133:479–490.

Benjamin, M.M. 2002. Modeling the mass-action expression for bidentate adsorption. Environ. Sci. Technol. 36:307–313.

Bond, W.J. 1995. On the Rothmund–Kornfeld description of cation exchange. Soil Sci. Soc. Am. J. 59:436–443.

Bond, W.J., and K. Verburg. 1997. Comparison of methods for predicting ternary exchange from binary isotherms. Soil Sci. Soc. Am. J. 61:444–454.

Bostick, B.C., M.A. Vairavamurthy, K.G. Karthikeyan, and J. Chorover. 2002. Cesium adsorption on clay minerals: An EXAFS spectroscopic investigation. Environ. Sci. Technol. 36:2670–2676.

Bourg, I.C., and G. Sposito. 2010. Connecting the molecular scale to the continuum scale for diffusion processes in smectite-rich porous media. Environ. Sci. Technol. 44:2085–2091.

Bourg, I.C., G. Sposito, and A.C.M. Bourg. 2007. Modeling the acid–base surface chemistry of montmorillonite. J. Colloid Interface Sci. 312:297–310.

Bradbury, M.H., and B. Baeyens. 2000. A generalized sorption model for the concentration dependent uptake of caesium by argillaceous rocks. J. Contam. Hydrol. 42:141–163.

Bradbury, M.H., and B. Baeyens. 2005. Modelling the sorption of Mn(II), Co(II), Ni(II), Zn(II), Cd(II), Eu(III), Am(III), Sn(IV), Th(IV), Np(V) and U(VI) on montmorillonite: Linear free energy relationships and estimates of surface binding constants for some selected heavy metals and actinides. Geochim. Cosmochim. Acta 69:875–892.

Bradbury, M.H., and B. Baeyens. 2009. Sorption modelling on illite Part I: Titration measurements and the sorption of Ni, Co, Eu and Sn. Geochim. Cosmochim. Acta 73:990–1003.

Brouwer, E., B. Baeyens, A. Maes, and A. Cremers. 1983. Cesium and rubidium ion equilibria in illite clay. J. Phys. Chem. 87:1213–1219.

Brown, D.R., and L. Kevan. 1988. Aqueous coordination and location of exchangeable Cu^{2+} cations in montmorillonite clay studied by electron spin resonance and electron spin-echo modulation. J. Am. Chem. Soc. 110:2743–2748.

Carlson, R.M., and J.R. Buchanan. 1973. Calcium-magnesium-potassium equilibria in some California soils. Soil Sci. Soc. Am. Proc. 37:851–855.

Chang, F.-R.C., N.T. Skipper, and G. Sposito. 1995. Computer simulation of interlayer molecular structure in sodium montmorillonite hydrates. Langmuir 11:2734–2741.

Chang, F.-R.C., N.T. Skipper, and G. Sposito. 1997. Monte Carlo and molecular dynamics simulations of interfacial structure in lithium-montmorillonite hydrates. Langmuir 13:2074–2082.

Chang, F.-R.C., N.T. Skipper, and G. Sposito. 1998. Monte Carlo and molecular dynamics simulations of electrical double-layer structure in potassium montmorillonite hydrates. Langmuir 14:1201–1207.

Charlet, L., and C. Tournassat. 2005. Fe(II)-Na(I)-Ca(II) cation exchange on montmorillonite in chloride medium: Evidence for preferential clay adsorption of chloride—Metal ion pairs in seawater. Aquat. Geochem. 11:115–137.

Chatterjee, R., D.A. Laird, and M.L. Thompson. 2008. Interactions among K^+–Ca^{2+} exchange, sorption of m-dinitrobenzene, and smectite quasicrystal dynamics. Environ. Sci. Technol. 42:9099–9103.

Chávez-Páez, M., L. de Pablo, and J.J. de Pablo. 2001. Monte Carlo simulations of Ca-montmorillonite hydrates. J. Chem. Phys. 114:10948–10953.

Chen, C.-C., and K.F. Hayes. 1999. X-ray absorption spectroscopy investigation of aqueous Co(II) and Sr(II) sorption at clay-water interfaces. Geochim. Cosmochim. Acta 63:3205–3215.

Chorover, J., and G. Sposito. 1995. Surface charge characteristics of kaolinitic tropical soils. Geochim. Cosmochim. Acta 59:875–884.

Chu, S.-Y., and G. Sposito. 1981. The thermodynamics of ternary cation exchange systems and the subregular model. Soil Sci. Soc. Am. J. 45:1084–1089.

Comans, R.N.J., M. Haller, and P. de Preter. 1991. Sorption of cesium on illite: Non-equilibrium behaviour and reversibility. Geochim. Cosmochim. Acta 55:433–440.

Curtin, D., F. Selles, and H. Steppuhn. 1998. Estimating calcium-magnesium selectivity in smectitic soils from organic matter and texture. Soil Sci. Soc. Am. J. 62:1280–1285.

DeSutter, T.M., G.M. Pierzynski, and L.R. Baker. 2006. Flow-through and batch methods for determining calcium-magnesium and magnesium-calcium selectivity. Soil Sci. Soc. Am. J. 70:550–554.

Duc, M., F. Gaboriaud, and F. Thomas. 2005. Sensitivity of the acid-base properties of clays to the methods of preparation and measurement. 1. Literature review. J. Colloid Interface Sci. 289:139–147.

Eberl, D.D. 1980. Alkali cation selectivity and fixation by clay minerals. Clays Clay Miner. 28:161–172.

Evangelou, V.P., and F.J. Coale. 1987. Dependence of the Gapon coefficient on exchangeable sodium for mineralogically different soils. Soil Sci. Soc. Am. J. 51:68–72.

Fair, M.C., and J.L. Anderson. 1989. Electrophoresis of nonuniformly charged ellipsoidal particles. J. Colloid Interface Sci. 127:388–400.

Feigenbaum, S., A. Bar-Tal, R. Portnoy, and D.L. Sparks. 1991. Binary and ternary exchange of potassium on calcareous montmorillonitic soils. Soil Sci. Soc. Am. J. 55:49–56.

Ferrage, E., C. Tournassat, E. Rinnert, and B. Lanson. 2005. Influence of pH on the interlayer cationic composition and hydration state of Ca-montmorillonite: Analytical chemistry, chemical modelling and XRD profile modelling study. Geochim. Cosmochim. Acta 69:2797–2812.

Fink, D.H., F.S. Nakayama, and B.L. McNeal. 1971. Demixing of exchangeable cations in free-swelling bentonite clay. Soil Sci. Soc. Am. Proc. 35:552–555.

Fletcher, P., K.M. Holtzclaw, C. Jouany, G. Sposito, and C.S. LeVesque. 1984a. Sodium-calcium-magnesium exchange reactions on a montmorillonitic soil: II. Ternary exchange reactions. Soil Sci. Soc. Am. J. 48:1022–1025.

Fletcher, P., and G. Sposito. 1989. The chemical modelling of clay/electrolyte interactions for montmorillonite. Clay Miner. 24:375–391.

Fletcher, P., G. Sposito, and C.S. LeVesque. 1984b. Sodium-calcium-magnesium exchange reactions on a montmorillonitic soil: I. Binary exchange reactions. Soil Sci. Soc. Am. J. 48:1016–1021.

Fripiat, J.J., P. Cloos, and A. Poncelet. 1965. Comparison entre les propriétés d'échange de la montmorillonite et d'une résine vis-à-vis des cations alcalins et alcalino-terreux. I. Réversibilité des processus. Bull. Soc. Chim. Fr. 1:208–215.

Gaines, G.L., and H.C. Thomas. 1953. Adsorption studies on clay minerals. II. A formulation of the thermodynamics of exchange adsorption. J. Chem. Phys. 21:714–718.

Gajo, A., and B. Loret. 2007. The mechanics of active clays circulated by salts, acids and bases. J. Mech. Phys. Solids 55:1762–1801.

Gapon, Y.N. 1933. On the theory of exchange adsorption in soils. J. Gen. Chem. USSR (Engl. Trans.) 3:144–160, cited by Sposito and Mattigod (1977).

Gast, R.G. 1969. Standard free energies of exchange for alkali metal cations on Wyoming bentonite. Soil Sci. Soc. Am. Proc. 33:37–41.

Gast, R.G. 1972. Alkali metal cation exchange on Chambers montmorillonite. Soil Sci. Soc. Am. Proc. 36:14–19.

Greathouse, J.A., and R.T. Cygan. 2005. Molecular dynamics simulation of uranyl(VI) adsorption equilibria onto an external montmorillonite surface. Phys. Chem. Chem. Phys. 7:3580–3586.

Greathouse, J.A., K. Refson, and G. Sposito. 2000. Molecular dynamics simulation of water mobility in magnesium-smectite hydrates. J. Am. Chem. Soc. 122:11459–11464.

Griffoen, J., and C.A.J. Appelo. 1993. Adsorption of calcium and its complexes by two sediments in calcium-hydrogen-chlorine-carbon dioxide systems. Soil Sci. Soc. Am. J. 57:716–722.

Gu, X., and L.J. Evans. 2008. Surface complexation modelling of Cd(II), Cu(II), Ni(II), Pb(II) and Zn(II) adsorption onto kaolinite. Geochim. Cosmochim. Acta 72:267–276.

Heidmann, I., I. Christl, C. Leu, and R. Kretzschmar. 2005. Competitive sorption of protons and metal cations onto kaolinite: Experiments and modeling. J. Colloid Interface Sci. 282:270–282.

Helling, C.S., G. Chesters, and R.B. Corey. 1964. Contribution of organic matter and clay to soil cation-exchange capacity as affected by the pH of the saturating solution. Soil Sci. Soc. Am. Proc. 28:517–520.

Hunter, R.J. 1993. Introduction to modern colloid science. Oxford University Press, Oxford, U.K.

Iwasaki, T., and T. Watanabe. 1988. Distribution of Ca and Na ions in dioctahedral smectites and interstratified dioctahedral mica/smectites. Clays Clay Miner. 36:73–82.

Janssen, R.P.T., M.G.M. Bruggenwert, and W.H. van Riemsdijk. 2003. Effect of Al hydroxide polymers on cation exchange of montmorillonite. Eur. J. Soil Sci. 54:335–345.

Jensen, H.E., and K.L. Babcock. 1973. Cation-exchange equilibria on a Yolo loam. Hilgardia 41:475–487.

Juo, A.S.R., and S.A. Barber. 1969. An explanation for the variability in Sr-Ca exchange selectivity of soils, clays and humic acid. Soil Sci. Soc. Am. Proc. 33:360–363.

Keren, R. 1979. The effect of hydroxy-aluminum precipitation on the exchange properties of montmorillonite. Clays Clay Miner. 27:303–304.

Kim, Y., and R.J. Kirkpatrick. 1998. NMR T_1 relaxation study of [133]Cs and [23]Na adsorbed on illite. Am. Mineral. 83:661–665.

Kopittke, P.M., H.B. So, and N.W. Menzies. 2006. Effect of ionic strength and clay mineralogy on Na–Ca exchange and the SAR-ESP relationship. Eur. J. Soil Sci. 57:626–633.

Kosakowski, G., S.V. Churakov, and T. Thoenen. 2008. Diffusion of Na and Cs in montmorillonite. Clays Clay Miner. 56:190–206.

Kozaki, T., A. Fujishima, S. Sato, and H. Ohashi. 1998. Self-diffusion of sodium ions in compacted sodium montmorillonite. Nucl. Technol. 121:63–69.

Kuyucak, S., O.S. Andersen, and S.-H. Chung. 2001. Models of permeation in ion channels. Rep. Prog. Phys. 64:1427–1472.

Laird, D.A., and C. Shang. 1997. Relationship between cation exchange selectivity and crystalline swelling in expanding 2:1 phyllosilicates. Clays Clay Miner. 45:681–689.

Laird, D.A., C. Shang, and M.L. Thompson. 1995. Hysteresis in crystalline swelling of smectites. J. Colloid Interface Sci. 171:240–245.

Laudelout, H., R. van Bladel, and J. Robeyns. 1972. Hydration of cations adsorbed on a clay surface from the effect of water activity on ion-exchange selectivity. Soil Sci. Soc. Am. Proc. 36:30–34.

Leote de Carvalho, R.J.F., and N.T. Skipper. 2001. Atomistic computer simulation of the clay-fluid interface in colloidal laponite. J. Chem. Phys. 114:3727–3733.

Levy, G.J., H.V.H. van der Watt, I. Shainberg, and H.M. du Plessis. 1988. Potassium-calcium and sodium-calcium exchange on kaolinite and kaolinitic soils. Soil Sci. Soc. Am. J. 52:1259–1264.

Lewis, R.J., and H.C. Thomas. 1963. Adsorption studies on clay minerals. VIII. A consistency test of exchange sorption in the systems sodium-cesium-barium montmorillonite. J. Phys. Chem. 67:1781–1783.

Liu, X., X. Lu, R. Wang, and H. Zhou. 2008. Effects of layer-charge distribution on the thermodynamic and microscopic properties of Cs-smectite. Geochim. Cosmochim. Acta 72:1837–1847.

Liu, C., J.M. Zachara, and S.C. Smith. 2004. A cation exchange model to describe Cs^+ sorption at high ionic strength in subsurface sediments at Hanford site, USA. J. Contam. Hydrol. 68:217–238.

Maes, A., and A. Cremers. 1977. Charge density effects in ion exchange. Part 1. Heterovalent exchange equilibria. J. Chem. Soc. Faraday Trans. I 73:1807–1814.

Maes, A., and A. Cremers. 1978. Charge density effects in ion exchange. Part 2. Homovalent exchange equilibria. J. Chem. Soc. Faraday Trans. I 74:1234–1241.

Maes, A., and A. Cremers. 1986. Highly selective ion exchange in clay minerals and zeolites, p. 254–295. *In* J.A. Davis, and K.F. Hayes (eds.) Geochemical processes at mineral surfaces. American Chemical Society, Washington, DC.

Marry, V., J.F. Dufreche, M. Jardat, G. Meriguet, P. Turq, and F. Grun. 2003. Dynamics and transport in charged porous media. Colloids Surf. A 222:147–153.

Marry, V., N. Malikova, A. Cadène, E. Dubois, S. Durand-Vidal, P. Turq, J. Breu, S. Longeville, and J.-M. Zanotti. 2008. Water diffusion in a synthetic hectorite by neutron scattering—Beyond the isotropic translational model. J. Phys. Condens. Matter 20:104205.

Marry, V., and P. Turq. 2003. Microscopic simulations of inter-layer structure and dynamics in bihydrated heteroionic montmorillonites. J. Phys. Chem. B 107:1832–1839.

McBride, M.B. 1979. An interpretation of cation selectivity variations in M^+–M^+ exchange on clays. Clays Clay Miner. 27:417–422.

McBride, M.B. 1980. Interpretation of the variability of selectivity coefficients for exchange between ions of unequal charge on smectites. Clays Clay Miner. 28:255–261.

McBride, M.B., T.J. Pinnavaia, and M.M. Mortland. 1975. Electron spin resonance studies of cation orientation in restricted water layers on phyllosilicate (smectite) surfaces. J. Phys. Chem. 79:2430–2435.

McKinley, J.P., J.M. Zachara, S.M. Heald, A. Dohnalkova, M.G. Newville, and S.R. Sutton. 2004. Microscale distribution of cesium sorbed to biotite and muscovite. Environ. Sci. Technol. 38:1017–1023.

McNab, W.W., Jr., M.J. Singleton, J.E. Moran, and B.K. Esser. 2009. Ion exchange and trace element surface complexation reactions associated with applied recharge of low-TDS water in the San Joaquin Valley, California. Appl. Geochem. 24:129–137.

Mizutani, T., T. Takano, and H. Ogoshi. 1995. Selectivity of adsorption of organic ammonium ions onto smectite clays. Langmuir 11:880–884.

Morel, J.-P., V. Marry, P. Turq, and N. Morel-Desrosiers. 2007. Effect of temperature on the retention of Cs^+ by Na-montmorillonite: Microcalorimetric investigation. J. Mater. Chem. 17:2812–2817.

Morton, J.D., J.D. Semrau, and K.F. Hayes. 2001. An x-ray absorp-tion spectroscopy study of the structure and reversibility of copper adsorbed to montmorillonite clay. Geochim. Cosmochim. Acta 65:2709–2722.

Nakano, M., K. Kawamura, and Y. Ichikawa. 2003. Local struc-tural information of Cs in smectite hydrates by means of an EXAFS study and molecular dynamics simulations. Appl. Clay Sci. 23:15–23.

Naylor, D.V., and R. Overstreet. 1969. Sodium-calcium exchange behavior in organic soils. Soil Sci. Soc. Am. Proc. 33:848–851.

Norrish, K. 1954. Manner of swelling of montmorillonite. Nature 173:256–257.

Onodera, Y., T. Iwasaki, T. Ebina, H. Hayashi, K. Torii, A. Chatterjee, and H. Mimura. 1998. Effect of layer charge on fixation of cesium ions in smectites. J. Contam. Hydrol. 35:131–140.

Pabalan, R.T., and F.P. Bertetti. 1999. Experimental and mod-eling study of ion exchange between aqueous solutions and the zeolite mineral clinoptilolite. J. Solution Chem. 28:367–393.

Papelis, C., and K.F. Hayes. 1996. Distinguishing between inter-layer and external sorption sites of clay minerals using x-ray absorption spectroscopy. Colloids Surf. A 107:89–96.

Park, C., P.A. Fenter, N.C. Sturchio, and K.L. Nagy. 2008. Thermodynamics, interfacial structure, and pH hysteresis of Rb^+ and Sr^{2+} adsorption at the muscovite (001)-solution interface. Langmuir 24:13993–14004.

Park, S.-H., and G. Sposito. 2002. Structure of water adsorbed on a mica surface. Phys. Rev. Lett. 89:085501.

Poinssot, C., B. Baeyens, and M.H. Bradbury. 1999. Experimental and modelling studies of caesium sorption on illite. Geochim. Cosmochim. Acta 63:3217–3227.

Powell, D.H., H.E. Fischer, and N.T. Skipper. 1998. The structure of interlayer water in Li-montmorillonite studied by neu-tron diffraction with isotopic substitution. J. Phys. Chem. B 102:10899–10905.

Rajec, P., V. Šucha, D.D. Eberl, J. Środón, and F. Elsass. 1999. Effect of illite particle shape on cesium sorption. Clays Clay Miner. 47:755–760.

Rhue, R.D., and R.S. Mansell. 1988. The effect of pH on sodium-calcium and potassium-calcium exchange selectivity for Cecil soil. Soil Sci. Soc. Am. J. 52:641–647.

Rotenberg, B., V. Marry, R. Vuilleumier, N. Malikova, C. Simon, and P. Turq. 2007. Water and ions in clays: Unraveling the interlayer/micropore exchange using molecular dynamics. Geochim. Cosmochim. Acta 71:5089–5101.

Rytwo, G., A. Banin, and S. Nir. 1996. Exchange reactions in the Ca-Mg-Na-montmorillonite system. Clays Clay Miner. 44:276–285.

Schlegel, M.L., K.L. Nagy, P. Fenter, L. Cheng, N.C. Sturchio, and S.D. Jacobsen. 2006. Cation sorption on the muscovite (001) surface in chloride solutions using high-resolution x-ray reflectivity. Geochim. Cosmochim. Acta 70:3549–3565.

Schramm, L.L., and J.C.T. Kwak. 1982. Influence of exchangeable cation composition on the size and shape of montmorillonite particles in dilute suspension. Clays Clay Miner. 30:40–48.

Serrano, S., P.A. O'Day, D. Vlassopoulos, M.T. García-González, and F. Garrido. 2009. A surface complexation and ion exchange model of Pb and Cd competitive sorption on natural soils. Geochim. Cosmochim. Acta 73:543–558.

Shainberg, I., N.I. Alperovitch, and R. Keren. 1987. Charge density and Na-K-Ca exchange on smectites. Clays Clay Miner. 35:68–73.

Shainberg, I., and W.D. Kemper. 1966. Hydration status of adsorbed cations. Soil Sci. Soc. Am. Proc. 30:707–713.

Shainberg, I., J.D. Oster, and J.D. Wood. 1980. Sodium/calcium exchange in montmorillonite and illite suspension. Soil Sci. Soc. Am. J. 44:960–964.

Shainberg, I., and H. Otoh. 1968. Size and shape of montmorillon-ite particles saturated with Na/Ca ions (inferred from viscos-ity and optical measurements). Isr. J. Chem. 6:251–259.

Shaviv, A., and S.V. Mattigod. 1985. Cation exchange equilibria in soils expressed as cation-ligand complex formation. Soil Sci. Soc. Am. J. 49:569–573.

Sheta, T.H., G.R. Gobran, J.E. Dufey, and H. Laudelout. 1981. Sodium-calcium exchange in Nile delta soils: Single values for Vanselow and Gaines–Thomas selectivity coefficients. Soil Sci. Soc. Am. J. 45:749–753.

Sinanis, C., V.Z. Keramidas, and S. Sakellariadis. 2003. Thermodynamics of potassium-magnesium exchange in two alfisols of northern Greece. Commun. Soil Sci. Plant Anal. 34:439–456.

Skipper, N.T., P.A. Lock, J.O. Titiloye, J. Swenson, Z.A. Mirza, W.S. Howells, and F. Fernandez-Alonso. 2006. The structure and dynamics of 2-dimensional fluids in swelling clays. Chem. Geol. 230:182–196.

Skipper, N.T., M.V. Smalley, G.D. Williams, A.K. Soper, and C.H. Thompson. 1995. Direct measurement of the electric double-layer structure in hydrated lithium vermiculite clays by neutron diffraction. J. Phys. Chem. 99:14201–14204.

Slade, P.G., J.P. Quirk, and K. Norrish. 1991. Crystalline swelling of smectite samples in concentrated NaCl solutions in relation to layer charge. Clays Clay Miner. 39:234–238.

Smith, D.E., Y. Wang, A. Chaturvedi, and H.D. Whitley. 2006. Molecular simulations of the pressure, temperature, and chemical potential dependencies of clay swelling. J. Phys. Chem. B 110:20046–20054.

Sposito, G. 1991. Effect of chloride ions on sodium-calcium and sodium-magnesium exchange on montmorillonite. Soil Sci. Soc. Am. J. 55:965–967.

Sposito, G. 1992. The diffuse-ion swarm near smectite particles suspended in 1:1 electrolyte solutions: Modified Gouy–Chapman theory and quasicrystal formation, p. 128–155. *In* N. Güven and R.M. Pollastro (eds.) Clay–water interface and its rheological implications. The Clay Minerals Society, Boulder, CO.

Sposito, G. 1993. Surface complexation of metals by natural colloids, p. 211–236. *In* J.A. Marinsky and Y. Marcus (eds.) Ion exchange and solvent extraction. Vol. 2. Marcel Dekker, New York.

Sposito, G. 1994. Chemical equilibria and kinetics in soils. Oxford University Press, Oxford, U.K.

Sposito, G. 1998. On points of zero charge. Environ. Sci. Technol. 32:2815–2819.

Sposito, G. 2004. The surface chemistry of natural particles. Oxford University Press, Oxford, U.K.

Sposito, G. 2008. The surface chemistry of soils. 2nd edn. Oxford University Press, Oxford, U.K.

Sposito, G., and P. Fletcher. 1985. Sodium-calcium-magnesium exchange reactions on a montmorillonitic soil: III. Calcium-magnesium exchange selectivity. Soil Sci. Soc. Am. J. 49:1160–1163.

Sposito, G., K.M. Holtzclaw, L. Charlet, C. Jouany, and A.L. Page. 1983a. Sodium-calcium and sodium-magnesium exchange on Wyoming bentonite in perchlorate and chloride background ionic media. Soil Sci. Soc. Am. J. 47:51–56.

Sposito, G., K.M. Holtzclaw, C.T. Johnston, and C.S. LeVesque-Madore. 1981. Thermodynamics of sodium-copper exchange on Wyoming bentonite at 298 K. Soil Sci. Soc. Am. J. 45:1079–1084.

Sposito, G., K.M. Holtzclaw, C. Jouany, and L. Charlet. 1983b. Cation selectivity in sodium-calcium, sodium-magnesium and calcium-magnesium exchange on Wyoming bentonite at 298 K. Soil Sci. Soc. Am. J. 47:917–921.

Sposito, G., C. Jouany, K.M. Holtzclaw, and C.S. LeVesque. 1983c. Calcium-magnesium exchange on Wyoming bentonite in the presence of adsorbed sodium. Soil Sci. Soc. Am. J. 47:1081–1085.

Sposito, G., and S.V. Mattigod. 1977. On the chemical foundation of the sodium adsorption ratio. Soil Sci. Soc. Am. J. 41:323–329.

Sposito, G., and S.V. Mattigod. 1979. Ideal behavior in Na$^+$-trace metal cation exchange on Camp Berteau montmorillonite. Clays Clay Miner. 27:125–128.

Sposito, G., and R. Prost. 1982. Structure of water adsorbed on smectites. Chem. Rev. 82:553–573.

Sposito, G., N.T. Skipper, R. Sutton, S.-H. Park, A.K. Soper, and J.A. Greathouse. 1999. Surface geochemistry of the clay minerals. Proc. Natl. Acad. Sci. U. S. A. 96:3358–3364.

Stadler, M., and P.W. Schindler. 1993. Modeling of H$^+$ and Cu^{2+} adsorption on calcium-montmorillonite. Clays Clay Miner. 41:288–296.

Strawn, D.G., and D.L. Sparks. 1999. The use of XAFS to distinguish between inner- and outer-sphere lead adsorption complexes on montmorillonite. J. Colloid Interface Sci. 216:257–269.

Sutton, R., and G. Sposito. 2005. Molecular structure in soil humic substances: The new view. Environ. Sci. Technol. 39:9009–9015.

Sutton, R., G. Sposito, M.S. Diallo, and H.-R. Schulten. 2005. Molecular simulation of a model of dissolved organic matter. Environ. Toxicol. Chem. 24:1902–1911.

Tambach, T.J., P.G. Bolhuis, E.J.M. Hensen, and B. Smit. 2006. Hysteresis in clay swelling induced by hydrogen bonding: Accurate prediction of swelling states. Langmuir 22:1223–1234.

Tambach, T.J., E.J.M. Hensen, and B.J. Smit. 2004. Molecular simulations of swelling clay minerals. J. Phys. Chem. B 108:7586–7596.

Tamura, K., H. Yamada, and H. Nakazawa. 2000. Stepwise hydration of high-quality synthetic smectite with various cations. Clays Clay Miner. 48:400–404.

Tang, L., and D.L. Sparks. 1993. Cation-exchange kinetics on montmorillonite using pressure-jump relaxation. Soil Sci. Soc. Am. J. 57:42–46.

Teppen, B.J., and V. Aggarwal. 2007. Thermodynamics of organic cation exchange selectivity in smectites. Clays Clay Miner. 55:119–130.

Teppen, B.J., and D.M. Miller. 2006. Hydration energy determines isovalent cation exchange selectivity by clay minerals. Soil Sci. Soc. Am. J. 70:31–40.

Thellier, C., and G. Sposito. 1989. Influence of electrolyte concentration on quaternary cation exchange by Silver Hill illite. Soil Sci. Soc. Am. J. 53:705–711.

Timms, W.A., and M.J. Hendry. 2007. Quantifying the impact of cation exchange on long-term solute transport in a clay-rich aquitard. J. Hydrol. 332:110–122.

Tournassat, C., E. Ferrage, C. Poinsignon, and L. Charlet. 2004. The titration of clay minerals. Part II. Structural-based model and implications for clay reactivity. J. Colloid Interface Sci. 273:234–246.

Tournassat, C., H. Gailhanou, C. Crouzet, G. Braibant, A. Gautier, A. Lassin, P. Blanc, and E.C. Gaucher. 2007. Two cation exchange models for direct and inverse modelling of solution major cation composition in equilibrium with illite surfaces. Geochim. Cosmochim. Acta 71:1098–1114.

Tournassat, C., C. Lerouge, P. Blanc, J. Brendlé, J.-M. Greneche, S. Touzelet, and E.C. Gaucher. 2008. Cation exchanged Fe(II) and Sr compared to other divalent cations (Ca, Mg) in the bure Callovian-Oxfordian formation: Implications for pore-water composition modelling. Appl. Geochem. 23:641–654.

Van Bladel, R., and H.R. Gheyi. 1980. Thermodynamic study of calcium-sodium and calcium-magnesium exchange in calcareous soils. Soil Sci. Soc. Am. J. 44:938–942.

Van Loon, L.R., and M.A. Glaus. 2008. Mechanical compaction of smectite clays increases ion exchange selectivity for cesium. Environ. Sci. Technol. 42:1600–1604.

van Oploo, P., I. White, P. Ford, M.D. Melville, and B.C.T. Macdonald. 2008. Pore water chemistry of acid sulfate soils: Chemical flux and oxidation rates. Geoderma 146:32–39.

Vanselow, A.P. 1932. Equilibria of the base-exchange reactions of bentonites, permutities, soil colloids, and zeolites. Soil Sci. 33:95–113.

Verburg, K., and P. Baveye. 1994. Hysteresis in the binary exchange of cations on 2:1 clay minerals: A critical review. Clays Clay Miner. 42:207–220.

Verburg, K., P. Baveye, and M.B. McBride. 1995. Cation-exchange hysteresis and dynamics of formation and breakdown of montmorillonite quasi-crystals. Soil Sci. Soc. Am. J. 59:1268–1273.

Voegelin, A., and R. Kretzschmar. 2003. Modelling sorption and mobility of cadmium and zinc in soils with scaled exchange coefficients. Eur. J. Soil Sci. 54:387–400.

Way, J.T. 1850. On the power of soils to absorb manure. J. R. Agric. Soc. Engl. 11:313–379.

Weiss, C.A., Jr., R.J. Kirkpatrick, and S.P. Altaner. 1990. The structural environments of cations adsorbed onto clays: [133]Cs variable-temperature MAS NMR spectroscopic study of hectorite. Geochim. Cosmochim. Acta 54:1655–1669.

White, A.F., M.S. Schulz, D.V. Vivit, A.E. Blum, D.A. Stonestrom, and J.W. Harden. 2005. Chemical weathering rates of a soil chronosequence on granitic alluvium: III. Hydrochemical evolution and contemporary solute fluxes and rates. Geochim. Cosmochim. Acta 69:1975–1996.

Whitley, H.D., and D.E. Smith. 2004. Free energy, energy, and entropy of swelling in Cs-, Na-, and Sr-montmorillonite clays. J. Chem. Phys. 120:5387–5395.

Wild, A., and J. Keay. 1964. Cation-exchange equilibria with vermiculite. J. Soil Sci. 15:135–144.

Wilson, J., J. Cuadros, and G. Cressey. 2004. An in situ time-resolved XRD-PSD investigation into Na-montmorillonite interlayer and particle rearrangement during dehydration. Clays Clay Miner. 52:180–191.

Xu, S., and J.B. Harsh. 1992. Alkali cation selectivity and surface charge of 2:1 clay minerals. Clays Clay Miner. 40:567–574.

Xu, X., A.G. Kalinichev, and R.J. Kirkpatrick. 2006. [133]Cs and [35]Cl NMR spectroscopy and molecular dynamics modeling of Cs^+ and Cl^- complexation with natural organic matter. Geochim. Cosmochim. Acta 70:4319–4331.

Zachara, J.M., and J.P. McKinley. 1993. Influence of hydrolysis on the sorption of metal cations by smectites: Importance of edge coordination reactions. Aquat. Sci. 55:250–261.

Zhang, Z.Z., and D.L. Sparks. 1996. Sodium-copper exchange on Wyoming montmorillonite in chloride, perchlorate, nitrate, and sulfate solutions. Soil Sci. Soc. Am. J. 60:1750–1757.

17

Chemisorption and Precipitation Reactions

Robert G. Ford
*United States Environmental
Protection Agency*

17.1 Introduction

The partitioning of dissolved components to a solid surface in soils is governed by the competition between sorption and desorption reactions. Chemisorption (or chemical adsorption) of a dissolved component (adsorbate) to a solid surface (adsorbent) has been defined as "adsorption that results from chemical bond formation (strong interaction) between the adsorbent and the adsorbate in a monolayer on the surface" (IUPAC, 1997). These partitioning reactions may be influenced by electrostatic forces between the adsorbate and adsorbent, similar to those controlling ion-exchange reactions (discussed in this volume). It has also been observed that sorption of solution species may result in more than monolayer coverage at the surface of solid surfaces. In these cases, the adsorbate may form polynuclear surface complexes or surface precipitates (e.g., Scheidegger and Sparks, 1996). Surface precipitation can be considered as a special form of homogeneous precipitation reaction where a solid component in the soil participates in the overall reaction process (Chapter 7 in Stumm and Morgan, 1996). In the following discussion, surface precipitation includes all processes that result in incorporation of the sorbate into a solid structure that could not form in the absence of the adsorbent. This includes (1) formation of a dilute solid solution with a structure that is distinct from or shared with the original adsorbent, (2) epitaxial growth of a new solid phase sharing a structural relationship with the adsorbent, and (3) nucleation of a new solid phase at the solid–water interface. Background information and discussion of the literature (dating up to 1998) addressing chemisorption and precipitation reactions in soils were provided by McBride (1999). The following discussion provides an update on the current understanding of these reactions in soil systems.

17.2 Conceptual Distinctions in Chemisorption and Precipitation

17.2.1 Chemisorption

Chemisorption reactions involving formation of mononuclear chemical bond between solutes in the solid solution and reactive surfaces in solid matrix are referred to as *surface complexation* reactions and are typically categorized by the relative binding strength of interaction between the adsorbate (species in solution) and the adsorbent (soil solid). One should note that the conceptual underpinning that guides our historical and contemporary understanding of chemisorption/surface complexation reactions is the notion that reactive surface sites can be viewed as a special type of ligand that coordinates to cations or exchanges with anions from solution (Schindler et al., 1976; Sigg and Stumm, 1981; Sposito, 1998, 1999). While our molecular-level understanding of the physicochemical forces that govern the characteristics of these surface reactions has improved over the years, the underlying conceptual basis remains technically sound. The driving force for molecules or ions in soil solution to form surface complexes with sites of adsorption in the soil solid matrix is governed by the physical and chemical properties of the adsorbate, sorption sites, and the soil solution (i.e., "solute," "sorbent," and "solvent" in the more general terminology in Stumm [1992]). As an example, the probability that a cation will form an inner-sphere versus an outer-sphere surface complex will, in part, be governed by the relative free-energy change between the cation maintaining a water molecule versus a surface hydroxyl site within its immediate coordination sphere. Cations for which coordination with a water molecule is more energetically favorable will tend to favor outer-sphere surface complexes.

However, the overall tendency for surface complex formation will be governed by other factors such as electrostatic forces between the adsorbate and sorption site. Therefore, projecting trends in surface complex formation must be predicated on knowledge of the physicochemical properties of the three primary components in the reaction (i.e., adsorbate, sorption sites, and solvent), in addition to how the surface complexation reaction(s) may be modified by interactions of any or all of these primary components with other solution or solid components in the reaction system.

There are several recent discussions that provide useful perspective on the fundamental chemical factors that govern the extent for and relative strength of surface complex bond formation (e.g., Sverjensky, 2006; Brown, 2009). It should be noted that current developments in our understanding of the physicochemical forces controlling chemisorption reactions have benefited from exploration of the molecular-level processes controlling ion partitioning onto soil solids, as reflected in discussions of bond formation by Sverjensky (2006) and Brown (2009). This emphasis is, to some extent, a departure from past efforts to relate apparent trends in ion partitioning based on periodic properties of the elements (i.e., adsorbate), for example, reference to apparent macroscopic trends with respect to qualitative concepts such as the "lyotropic" or "Hofmeister" series. However, as noted in several studies, deviations in observed ion partitioning trends from qualitative concepts specific to the sorbate support the need to further development of our molecular-level understanding of chemisorption reactions (e.g., Dove and Craven, 2005; Teppen and Miller, 2006; Rotenberg et al., 2009; Kowacz et al., 2010). (Another form of solid-solution partitioning, hydrophobic bonding, plays an important role in governing uptake of nonpolar organic compounds onto soil solids. This specific partitioning mechanism is not discussed further within this chapter, and the reader is referred to other comprehensive reviews [e.g., Karickhoff, 1981; Stumm, 1992; Schwarzenbach et al., 1993], along with recent developments in the concept of bonding mechanisms intermediate between nonspecific (hydrophobic) bonding and chemisorption reactions for organic molecules [e.g., Lambert, 2008; Keiluweit and Kleber, 2009].)

There is a range of binding strength for ion adsorption that depends on characteristics of the adsorbate, adsorbent, and chemistry of the soil solution. However, discussions of binding strength are generally couched in terms of "weak" or "strong" surface complexation reactions, albeit a common convention in chemistry would categorize both the covalent and electrostatic interactions involved in adsorption as "strong" intermolecular forces (Israelachvili, 1994). For description of surface complexation reactions, the microscopic distinction that the adsorbate possesses solvation properties is borrowed from the characterization of soluble ion pairs (Westall, 1986; Stumm, 1992). If solvating water molecules are interposed between the cation or anion and the surface, the surface complex is referred to as *outer sphere* and is considered to be "weak." Conversely, if upon surface complexation, the adsorbate loses waters of hydration such that there are no water molecules interposed between the cation

or anion and the surface, the surface complex is referred to as *inner sphere* and is considered to be "strong" (Sposito, 1984). The propensity of a cation or anion to form either an inner-sphere or outer-sphere surface complex is a function of the adsorbate, the surface functional groups of the adsorbent, and the aqueous-phase chemistry (e.g., pH and ionic strength).

For inner-sphere complexation of ions onto solid surfaces, the molecular rationale for the reaction is formation of covalent bonding within an ion's first coordination sphere. From this perspective, several authors have observed an apparent consistency between the approach to establish linear free-energy relationships (LFER) for aqueous coordination reactions and surface complexation reactions developed for the formation of surface complexes (e.g., Schindler et al., 1976; Sigg and Stumm, 1981; Tamura and Furuichi, 1997). The apparent systematic correlation captured in an LFER reflects the relationship between the free energies of complex formation and thermodynamic properties of ions. For surface complexation reactions, LFER plots of the stability constants of metal complexation with hydroxyl surface sites versus the hydrolysis constant for metals in aqueous solutions have been interpreted as providing a thermodynamic basis for the apparent selectivity trends for metal adsorption onto reference hydrous metal oxides. Recent compilations by Bradbury and Baeyens (2005, 2009) extend these relationships to metal adsorption onto the phyllosilicate minerals, montmorillonite, and illite. As stated by the authors, the utility of these relationships is their use as a tool to estimate surface complexation constants for ions in which no data or data of high uncertainty are available. In Figure 17.1, LFER plots of published data for cations and anions are provided to illustrate the correlations that one may observe with similarities in the energetics and molecular characteristics of homogeneous (aqueous) and heterogeneous (surface) coordination reactions. While these relationships are tied to our conceptual understanding of the molecular characteristics of surface-coordination reactions, their use to project surface complexation trends in soil systems should be limited to qualitative applications. The number of interacting components, both aqueous and solid, within a soil system necessitates a more comprehensive perspective.

Important adsorbent phases commonly found in the environment include phyllosilicate minerals, metal-oxohydroxide phases, sulfide phases, and natural organic matter (Dixon and Schulze, 2002). Many phyllosilicate minerals possess a permanent negative charge as a result of the substitution of lower-valence cations, that is, Mg(II), Fe(II), Li(I) for Al(III) in the octahedral layer and/or Al(III) for Si(IV) in the tetrahedral layer (referred to as isomorphic substitution). There are two main classes of phyllosilicate minerals based on layer structure. The 1:1 mineral layer type is comprised of one Si-tetrahedral layer and one Al-octahedral layer, which in soils is commonly represented by the mineral kaolinite having the general formula $[Si_4]$ $Al_4O_{10}(OH)_8 \cdot nH_2O$. Kaolinite and related minerals generally have insignificant degrees of cation substitution within their octahedral and tetrahedral layers, and, thus generally posses a very low permanent negative charge. The 2:1 mineral type is
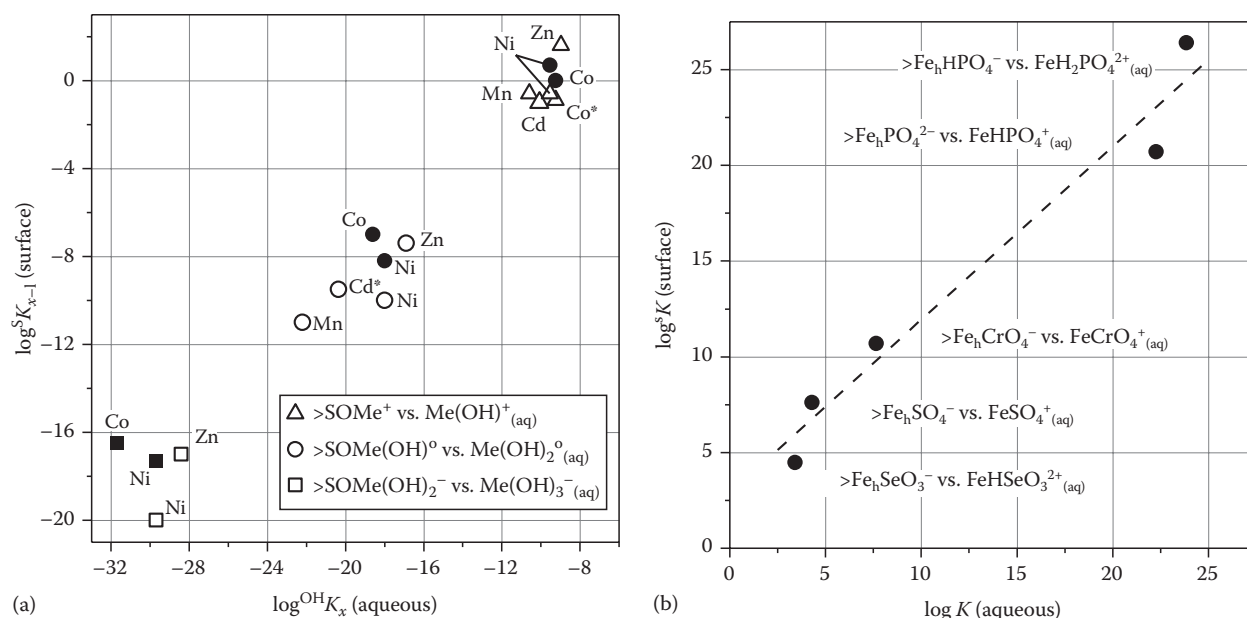
FIGURE 17.1 Comparison of related formation constants for surface and aqueous complexes for coordination reactions between a selection of ions and aqueous or surface functional groups. (a) Relationship between the stability constant for cation complexation with surface hydroxyls of Na-exchanged montmorillonite/illite versus the metal hydrolysis species formation constant in solution (Bradbury and Baeyens, 2005, 2009). Species stoichiometry noted in legend; starred metal labels indicate surface stability constant is for Ca-exchanged montmorillonite. (b) Relationship between the formation constant for anion ligand exchange with surface hydroxyls on hydrous ferric oxide versus formation of aqueous complexes with Fe^{3+} (Visual MINTEQ, 2009). Species stoichiometry shown next to data point.

comprised of one Al-octahedral layer interposed between two Si-tetrahedral layers comparable to the mica structures. The 2:1 layer class is represented by a variety of minerals, which are classified based on the location (tetrahedral versus octahedral layer) and relative amount of isomorphic substitution. The three major mineral classes within the 2:1 layer type are illite ($M_x[Si_{6.8}Al_{1.2}]$ $(Al_3Fe_{0.25}Mg_{0.75})O_{20}(OH)_4$), vermiculite ($M_x[Si_7Al](Al_3Fe_{0.5}Mg_{0.5})$ $O_{20}(OH)_4$), and smectite ($M_x[Si_8]Al_{3.2}Fe_{0.2}Mg_{0.6}O_{20}(OH)_4$), which display different levels of cation substitution in their tetrahedral and octahedral layers. The permanent negative charge imparted to 2:1 clay minerals by isomorphic substitution is typically balanced through exchange reactions involving major cations in soil solution (e.g., Na^+, K^+, Ca^{2+}, or Mg^{2+}; represented by "M_x" in the formulas listed above).

Chemisorption of ions to phyllosilicates may occur via electrostatic exchange reactions (on the basal plane or within clay particle interlayers) or surface complexation with surface functional groups, for example, terminal silanol or aluminol functional groups at the edges of clay particles, for example, montmorillonite (Bradbury and Baeyens, 2005) and illite (Bradbury and Baeyens, 2009). Due to differences in the levels of isomorphic substitution for the 1:1 and 2:1 clay-mineral classes, ion exchange is usually only significant for 2:1 phyllosilicates. In addition to siloxane oxygen atoms along the basal plane, phyllosilicates possess two types of terminal ionizable hydroxyl groups (>OH°), aluminol and silanol, protruding from the edge surface. These edge >OH° groups can form both inner- and outer-sphere surface complexes with metal cations and oxyanions depending on

the pH of the bathing solution and on the specific chemical characteristics of the adsorbate.

The most important inorganic phases for surface reactions with ions in many soil systems are the metal-oxohydroxide phases. These phases are characterized by hexagonal or cubic close-packed O or OH anions with Fe(II)/Fe(III), Al(III), and/or Mn(III)/Mn(IV) occupying octahedral sites. These oxides are present as discrete phases and as complex mineral assemblages, being associated with phyllosilicates and primary minerals as coatings or with humic macromolecules. In soils, the crystallinity of these phases typically varies from poorly ordered to well-crystalline forms and grain size from the nanometer to micrometer scale. Among the most common Fe-oxohydroxide phases found in soils are the poorly ordered phase ferrihydrite ($Fe_2O_3 \cdot nH_2O$), and the moderate to well-crystalline phases, goethite (α-FeOOH), and hematite (α-Fe_2O_3). The most common Al-oxohydroxide phase found in soils is gibbsite (γ-Al(OH)$_3$). Additionally, poorly ordered aluminosilicates can be important reactive phases in certain soils and these include the very poorly ordered allophanes (Si/Al ratios 1:2–1:1) and the paracrystalline phase, imogolite ($SiO_2 \cdot Al_2O_3 \cdot 2H_2O$). While Mn oxohydroxides are less prevalent than Fe and Al oxohydroxides in soils, they are very important phases in terms of surface-mediated redox reactions and because of their propensity for high metal sorption. The mineralogy of Mn is complicated by the range in Mn–O bond lengths resulting from extensive substitution of Mn(II) and Mn(III) for Mn(IV). Thus, there exists a continuous series of stable and metastable compositions from

MnO to MnO$_2$ forming a large variety of minerals. Among the more common Mn oxohydroxides are pyrolusite (β-MnO$_2$), the hollandite–cryptomelane family (α-MnO$_2$), todorokite, and birnessite (δ-MnO$_2$).

The reactive surface functional group for all of the metal-oxohydroxide phases is the inorganic >OH^o moiety exposed on the outer periphery of these minerals. The reactivity of a specific metal oxohydroxide is dependent on the surface area (S$_A$), surface-site density (N$_S$), the degree of coordination of the >OH^o group to the bulk structure, and the point of zero charge (PZC). The charge on the mineral surface may impose either attractive or repulsive contributions to the overall adsorption reaction, depending on the type of charge possessed by the adsorbate. The surface charge of oxohydroxide minerals and edge sites on phyllosilicates is derived from the protonation and ionization of exposed surface hydroxyl groups, represented by >OH^o, where ">" represents the bond to structural elements within the solid (e.g., Fe, Al, Mn). As a function of pH, the surface functional groups can be generally described with the following idealized nomenclature: >OH$_2^+$, >OH^o, and >O$^-$, which illustrates the pH-dependent charging behavior in the absence of other potential-determining ions in solution. The exact charge associated with the various surface functional groups is difficult to measure, so the main purpose of employing this nomenclature is to illustrate that surface charge varies as a function of soil solution chemistry. Discussions of surface-charging behavior and surface-coordination reactions for a range of common soil minerals is provided elsewhere (e.g., Stumm, 1992; Sparks, 2003). The properties of the adsorbent that impact surface complexation are controlled by both the grain size and specific chemical structure of the solid phase.

While the surfaces of oxohydroxide minerals constitute a significant fraction of sites that participate in chemisorption reactions, they are not the only type of surface sites present. Depending on soil type and mineralogy, other sources of reactive surfaces may include carbonates, sulfides, and soil organic matter. For these solids, the same conceptual convention can be employed to identify the reactive chemical surface moiety for the purpose of describing chemisorption reactions based on chemical and electrostatic properties. There are several published examples that illustrate application of the surface complexation model (SCM) concept to describe ion chemisorption to the surface of carbonates (Villegas-Jimenez et al., 2009), sulfides (Wolthers et al., 2005), and components of soil organic matter (Westall et al., 1995; Cabaniss, 2008; Cabaniss, 2009). Summarized in Table 17.1 is a listing of the types of chemical nomenclature used to represent the reactive surface sites for these solid components.

Soil organic matter, or humus, is derived from the physical, chemical, and biological breakdown of vegetation and fauna originating from above and below the soil surface (e.g., Wershaw, 1994; Zech et al., 1997; Ponge, 2003; Sutton and Sposito, 2005). The recalcitrant soil organic matter thus produced is comprised of complex polymers called humic substances, which constitute another very important reactive phase in soil solids

TABLE 17.1 Published Examples for Chemical Representations of Reactive Surface Sites Used in Development of Surface Complexation Models to Describe Protonation–Deprotonation Reactions for Variably Charged Solids

Chemical Representations of Reactive Surface Sites	
Chemical System	Protonation–Deprotonation Reactions
CaCO$_3$ (calcite)–H$_2$O–CO$_2$	>CaOH^o + H$^+$ = >CaOH$_2^+$
(Martin-Garin et al., 2003; Gaskova et al., 2009)	>CaOH^o = >CaO$^-$ + H$^+$
	>CaOH^o + CO$_3^{2-}$ + 2H$^+$ = >CaHCO$_3^o$ + H$_2$O
	>CaOH^o + CO$_3^{2-}$ + H$^+$ = >CaCO$_3^-$ + H$_2$O
	>CO$_3$H^o = >CO$_3^-$ + H$^+$
	>CO$_3$H^o + Ca^{2+} = >CO$_3$Ca$^+$ + H$^+$
FeS (mackinawite)–H$_2$O	>FeSH^o + H$^+$ = >FeSH$_2^+$
(Wolthers et al., 2005)	>FeSH^o = >FeS$^-$ + H$^+$
	>Fe$_3$SH^o + H$^+$ = >Fe$_3$SH$_2^+$
	>Fe$_3$SH^o = >Fe$_3$S$^-$ + H$^+$
Lignocellulosic substrate–H$_2$O	>S$_1$OH = >S$_1$O$^-$ + H$^+$ (carboxylic sites)
(Ravat et al., 2000)	>S$_2$OH = >S$_2$O$^-$ + H$^+$ (phenolic sites)
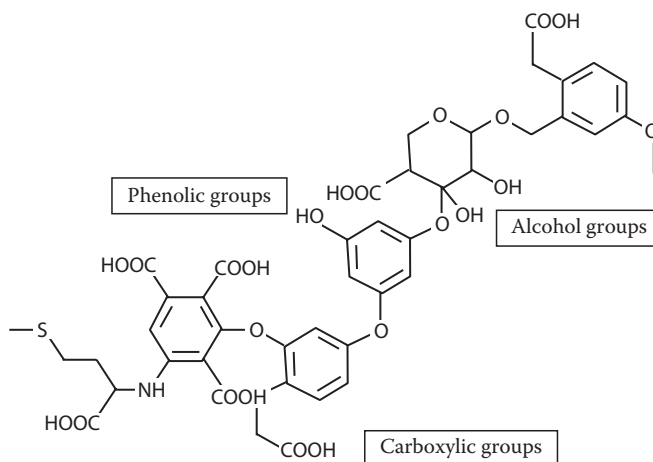
FIGURE 17.2 Hypothetical structural model of a soil organic matter structural fragment illustrating types of reactive functional groups. Protonation–deprotonation reactions occurring during acid titration are generally attributed to carboxylic and phenolic moieties.

(e.g., Kelleher and Simpson, 2006; Simpson et al., 2007a, 2007b; Smejkalova and Piccolo, 2008). A variety of functional groups are present in humic substances, and, like >OH^o functional groups of the inorganic metal oxohydroxides, these also are characterized by pH-dependent charging mechanisms. The primary functional groups associated with humic substances in terms of surface charge are typically represented as carboxyl and phenolic groups (Figure 17.2); however, less abundant amino, imidazole, sulfhydryl, or other potential structural groups may play an important role in protonation–deprotonation and chemisorption reactions of some ions when present at trace levels (e.g., Schaumann, 2006; Niederer et al., 2007; Matynia et al., 2010). Degradation of natural organic matter in soils leads to a change in the relative distribution of the chemical structures within the complex suite of macromolecular compounds that constitute

soil organic matter. As an example, Lorenz et al. (2007) illustrate a general scheme for the evolution of the chemical makeup of terrestrial biomass during weathering resulting in the depletion of oxygenated-alkyl carbon structures and the enhancement of alkyl carbon structures. As illustrated in the review by Chefetz and Xing (2009), changes in the relative proportion of structural components in soil organic matter may influence the reactivity of this adsorbent pool within soils, since these structural components comprise the reactive moieties in soil organic matter that are binding sites for dissolved inorganic and organic compounds in the soil solution.

As illustrated by the previous discussion, chemical bonding between solutes in the soil solution and the surfaces of soil solids will exert control over solute concentrations. Typically, these reactions dominate solid-water partitioning at low solute concentrations or in systems where there is minimal active precipitation of new soil minerals. However, some soil systems will be characterized by large fluxes of ions through the soil pore water either due to infiltration from external water sources or internal instability of the mineral matrix as a function of soil age (e.g., Huggett, 1998; Lilienfein et al., 2003) or microbial activity (e.g., Wiederhold et al., 2007). In these systems, active precipitation of new solid phases may dominate over chemisorption reactions to control solute concentrations in the soil solution.

17.2.2 Precipitation and Coprecipitation

Mineral–water reactions occur during movement of water through soil pores. These reactions may result in the release of structural components from soil solids due to mineral dissolution or result in the buildup to oversaturation and consequent precipitation of secondary minerals. As an outcome of mineral–water reactions along a flow path, fluid compositions and the mineralogical makeup of the soil matrix will continuously evolve toward a stable state or an equilibrium state. Mineral precipitation processes in soil systems may govern the concentrations of major and trace elements in the soil solution.

Full treatment of precipitation processes, including coverage of relevant thermodynamic and kinetic concepts, is outside the scope of this document. The reader is referred to numerous standard textbooks in geochemistry, soil science, and aquatic chemistry (e.g., Lindsay, 1979; Drever, 1982; Stumm, 1992; McBride, 1994; Stumm and Morgan, 1996; Langmuir, 1997; Lasaga, 1999; Sparks, 2003; Sposito, 2008). The purpose of this section is to introduce key concepts and issues regarding the potential impact precipitation reactions may exert on solute partitioning in soils. In general, mineral precipitation in relation to the solid-phase partitioning of soil solutes can be discussed in the context of four widely studied processes:

- *Precipitation from solution*: Nucleation and growth of a solid phase exhibiting a molecular unit that repeats itself in three dimensions. Homogeneous nucleation occurs from bulk solution and heterogeneous nucleation occurs on the surfaces of organic or mineral particles. Heterogeneous nucleation is thought to be more important in natural systems that are rich in reactive inorganic and biological surfaces. Precipitation may result in the formation of sparingly soluble oxohydroxides, carbonates, and, in anoxic systems, sulfides. Many precipitation reactions have a strong dependence on pH.

- *Coprecipitation*: Incorporation of an element as a trace or minor constituent within a precipitating phase. In this case, the ion substitutes for a more concentrated component in the crystal lattice (isomorphous substitution). This process is distinct from adsorption due to incorporation of the ion within the bulk structure of the major mineral phase. Examples of coprecipitation include Cr(III) in hydrous ferric oxide, Cd(II) in calcium carbonate, and As(III) in iron sulfide.

- *Surface precipitation*: A precipitation process intermediate between surface complexation and precipitation from bulk solution. Surface precipitation represents the continuous growth of particles formed via heterogeneous nucleation. Macroscopic studies of adsorption of some solutes, particularly divalent and trivalent cations, suggest that precipitation occurs at surfaces under conditions where the solid is apparently undersaturated based on solution concentrations (Dzombak and Morel, 1990).

- *Mineral transformation*: Adsorbed ions can become incorporated into minerals that form as a result of recrystallization or mineral transformation processes in soils. Transformation reactions may be accelerated or retarded by the ion, and in some cases mineral transformation may result in the exclusion of the solute from the solid phase. Examples include incorporation of anions, such as As(V) and U(VI), into metastable hydrous ferric oxide (poorly crystalline ferrihydrite) and transformation to more crystalline Fe oxohydroxides (e.g., Ford, 2002; Nico et al., 2009), coprecipitation of metals with iron monosulfide and transformation to iron disulfide (e.g., Lowers et al., 2007), and incorporation of metals into layered double hydroxides (LDHs; typically with Al) as intermediates between adsorbed/surface precipitated metal ions like Ni and Zn and metal-ion-containing phyllosilicates.

The relative importance of these processes will be determined by solute characteristics as well as characteristics of soil solution and solids within a given system. These individual processes are discussed in more detail in the following sections.

17.2.2.1 Precipitation from Solution

Solution precipitation or crystallization can be divided into two main processes: nucleation and crystal growth. Nucleation occurs prior to growth of a mineral crystal. Both nucleation and growth processes require a system to be oversaturated in the new phase. The probability that nucleation will occur increases exponentially as a function of the degree of oversaturation. Nucleation of a new phase is often facilitated in the presence of

a surface (heterogeneous nucleation) rather than in bulk solution (homogeneous nucleation). Because nucleation and growth are processes that compete for dissolved solutes, at high degrees of oversaturation, the rate of nucleation may be so fast that all excess solute is partitioned into crystal nuclei. In contrast, lower levels of oversaturation can result in the growth of existing crystals without nucleation. Well-formed or euhedral crystals typically develop slowly via growth from solution at low degrees of oversaturation. During crystal growth, various chemical reactions can occur at the surface of the growing mineral, such as adsorption, ion exchange, diffusion, and formation of surface precipitates. In general, the rate of crystal growth is controlled either by transport of solutes to the growing surface (i.e., transport controlled), by reactions at the surface (i.e., surface controlled), or a combination of these factors.

For the most abundant cations present in soils, such as Al, Si, Fe, Mn, Ca, and Mg, precipitation of mineral forms is common and will in many cases control concentrations observed in solution (e.g., see Chapter 5 in Sposito, 2008). Concentrations of trace solutes are typically several orders of magnitude below the concentrations of the major ions in soil solutions. At low concentrations, adsorption, surface precipitation, or formation of a dilute solid solution (coprecipitation; e.g., Shao et al., 2009) may be the more probable removal processes for trace solutes (McBride, 1994).

17.2.2.2 Continuum from Surface Complexation to Precipitation

Experimental studies examining chemisorption at high solute concentrations have been used to illustrate that surface loadings can exceed monolayer coverages (Dzombak and Morel, 1990). Surface precipitation may result when adsorption leads to high sorbate coverage at the mineral–water interface. Surface precipitation can be thought of as an intermediate stage between surface complexation and bulk precipitation of the sorbing ion in solution (Farley et al., 1985; Katz and Hayes, 1995; Lützenkirchen and Behra, 1995). At low concentrations of the sorbing metal at the mineral surface, surface complexation is the dominant process. As the concentration of the sorbate increases, the surface complexation concentration increases to the point where nucleation and growth of a surface precipitate occurs. Surface precipitation can be viewed as a special case of coprecipitation, where the mineral interface is a mixing zone for ions incorporated into the surface precipitate and those from the underlying substrate (e.g., Thompson et al., 1999; Schlegel et al., 2001; Román-Ross et al., 2006; Schlegel and Manceau, 2006). It is generally believed that surface precipitation can occur from solutions that would appear to be undersaturated relative to precipitate formation based on considering solution saturation indices. The reasons for this may be due, in part, to incomplete consideration of all possible precipitate phases with lower solubility that could form under system conditions or to the way component activities at the mineral–water interface are modeled (Kulik, 2002a, 2002b; Sverjensky, 2003).

Characterization of metal partitioning in soils developed under a variety of conditions have confirmed the importance of the interfacial precipitation products identified in simple experimental systems (e.g., Manceau et al., 2000; Jacquat et al., 2008). However, it is unclear if these phases were formed via homogeneous or heterogeneous reactions in the observed soil systems. With improvements in the availability of solubility and thermochemical data for solids that were previously unidentified in soil systems, it may become apparent that their formation can be assumed as a more common feature of solute partitioning. It should be noted, however, that solid surfaces do play significant roles as collectors of solutes into a more confined reaction volume and they often serve as the primary source of soluble components that combine with solutes to form precipitation products.

17.3 Influence of Abiotic and Biotic Processes

The chemical characteristics of the soil solution and properties of soil mineral components are, in part, influenced by microbial reactions (e.g., Brown et al., 1999; Chadwick and Chorover, 2001; Birkham et al., 2007). Microbial activity within the soil system may also play a more direct role in controlling ion speciation and mobility via direct respiration of solutes in soil pore water or adsorbed to mineral surfaces (e.g., Lloyd et al., 2003). The influence of microbial reactions may be more pronounced in soils with large pools of degradable organic carbon and supporting nutrient fluxes. For microbially productive soil systems, soil-solution chemical parameters such as pH, alkalinity, and the concentrations of iron, manganese, and sulfur species may be regulated, in part, by microbial respiration. As an example, microbial respiration within the rhizosphere impacts the level of CO_2 in the soil solution, with resultant CO_2 partial pressure 10- to 100-fold higher than that of the atmosphere (Hinsinger et al., 2009). Ultimately, the soil solution chemistry will be governed by the dynamic interaction between these biotic processes and concurrent chemical weathering of solid components within the soil. In order to provide perspective on the potential influence of the subsurface microbiology, the following discussion provides an overview of general characteristics of subsurface microbiology and the influence of microbial activity on the redox state within soil pore water.

Since microorganisms chemically transform soil constituents such as dissolved oxygen, iron (aqueous and solid forms of Fe(III) and Fe(II)), and sulfur (aqueous species such as sulfate and dissolved sulfide), their metabolic reactions may exert significant influence on the redox chemistry within soil pore water. As previously discussed, redox conditions within the soil may govern precipitation–dissolution reactions that control precipitation or coprecipitation of solutes, as well as the types and concentration of soil solids that may serve as adsorbents. From this perspective, some knowledge about the microbial populations that a function within the soil system may be necessary

to understand the existing redox status and make projections about how it may evolve. For heterotrophic microorganisms, the electrons or reducing equivalents (hydrogen or electron-transferring molecules) produced during degradation of organic compounds must be transferred to a terminal electron acceptor (TEA). Observations of microbial systems have led to the development of a classification system that groups microorganisms into three categories, according to predominant TEAs (Chapelle et al., 2002; Salminen et al., 2006):

- *Aerobic bacteria*: Bacteria that can only utilize molecular oxygen as a TEA. Without molecular oxygen, these bacteria are not capable of degradation.
- *Facultative aerobes/anaerobes*: Bacteria that can utilize molecular oxygen or when oxygen concentrations are low or nonexistent, may switch to nitrate, manganese oxides, or iron oxides as electron acceptors.
- *Anaerobes*: Bacteria that cannot utilize oxygen as an electron acceptor and for which oxygen is toxic. Though members may utilize nitrate or other electron acceptors, it can be said that they generally utilize sulfate or carbon dioxide as electron acceptors.

In any environment in which microbial activity occurs, there is a progression from aerobic to anaerobic conditions (ultimately methanogenic) with an associated change in the redox status of the system. There is generally a definite sequence of electron acceptors used in this progression through distinctly different redox states. The rate, type of active microbial population, and level of activity under each of these environments are controlled by several factors. These include the concentration of the electron acceptors, substrates that can be utilized by the bacteria, and specific microbial populations active within the soil system. This results in a loss of organic carbon and various electron acceptors from the system as well as a progression in the types and physiological activity of the indigenous bacteria.

If microbial activity is high, the soil system would be expected to progress rapidly through these conditions. The following scenario outlines a general sequence of events in which aerobic metabolism of preferential carbon sources would occur first. The carbon source may be from anthropogenic organic sources or other more readily degradable forms of natural organic carbon, which has entered the system previously or simultaneously with surface infiltration.

- *Oxygen-reducing to nitrate-reducing conditions*: Once available oxygen is consumed, active aerobic populations begin to shift to nitrate respiration. Denitrification will continue until available nitrate is depleted or usable carbon sources become limiting.
- *Nitrate-reducing to manganese-reducing conditions*: Once nitrate is depleted, populations that reduce manganese may dominate. Bacterial metabolism of substrates utilized by manganese-reducing populations will continue until the concentration of manganese oxide becomes limiting.

- *Manganese-reducing to iron-reducing conditions*: When manganese oxide becomes limiting, iron reduction becomes the predominant reaction mechanism. Available evidence suggests that iron reduction does not occur until all Mn(IV) oxides are depleted. In addition, bacterial Mn(IV) respiration appears to be restricted to areas where sulfate is nearly or completely absent.
- *Iron-reducing to sulfate-reducing conditions*: Iron reduction continues until substrate or carbon limitations allow sulfate-reducing bacteria (SRB) to become active. SRB then dominate until usable carbon or sulfate limitations impede their activity.
- *Sulfate-reducing to methanogenic conditions*: Once usable carbon or sulfate limitations occur, methanogenic bacteria are able to dominate.

Ultimately, these processes may govern both the chemical speciation of solutes in soil solutions and the solid-phase partitioning reactions that control aqueous solute concentrations. For the latter, biotic controls on solution chemistry could potentially dictate the types of adsorbent surfaces or the saturation state of the soil solution relative to solute precipitation reactions.

As a unique example of the biogeochemical complexity of soil systems, it is worth examining the volume of soil immediately adjacent to living roots that project down from plants located at the soil surface. This zone within the soil system is referred to as the rhizosphere, which encompasses a unique setting relative to the fluxes of chemical constituents that are governed by interactions among plant physiological processes, soil microorganisms, soil mineral components, and the transport of soil water and gases (e.g., Belnap et al., 2003; Hinsinger et al., 2009). Within the rhizosphere, plants mediate the chemical composition of the soil solution via release of low-molecular-weight organic acids such as malate, citrate, and oxalate (Jones, 1998; Bais et al., 2006). These organic acids may form complexes with ions in solution, form surface complexes with soil minerals, or may be oxidized to forms of inorganic carbon via microbial degradation (along with new biomass). Depending on the relative concentration of these low-molecular-weight organic acids to ions and sorption sites, complexation reactions that occur in these ternary systems may lead to decreased (Kraemer et al., 1999; Neubauer et al., 2000) or increased (Neubauer et al., 2000) ion adsorption onto mineral surfaces. More complex organic molecules, such as siderophores and other hydroxamic acids that are secreted by plants and microorganisms (Kraemer et al., 2006; Crumbliss and Harrington, 2009), present a more specialized class of organic molecules that can form polydentate complexes with ions in solution. These organic molecules may possess specific functional roles such as increasing the availability of nutritional forms of iron (e.g., Crumbliss and Harrington, 2009) or detoxifying metal ions present in the soil solution (e.g., Gilis et al., 1998; Dimkpa et al., 2008). Interactions of ions with these organic molecules in the soil solution exert direct influence on the potential for chemisorption or precipitation reactions to occur.

17.4 Quantitative Descriptions of Chemisorption and Precipitation Reactions

Reviews of mathematical models for describing chemisorption reactions and their practical application for depicting solute complexation to solid surfaces in natural systems is provided in Zachara and Westall (1999) and Goldberg et al. (2007 and references therein). In general, these models can be grouped into two categories that project solute partitioning using semiempirical relationships or through the development of chemical reaction expressions that depict chemisorption as the formation of chemical bonds between the solute and surface sites representing the termination of the adsorbent structure. The second category of models is commonly referred to as SCMs. Their use for describing chemisorption reactions has become standard since they provide the following: (1) the means to rationalize observed patterns in solute binding to solid surfaces due to electrostatic and chemical properties of the reactants and (2) a mathematical framework consistent with thermodynamic descriptions of aqueous complexation and precipitation reactions. The utility and chemical sensibility of this modeling framework has resulted in the development of computer software applications and supporting data compilations of chemisorption reactions for cations and anions on soil mineral phases (e.g., Dzombak and Morel, 1990; Tonkin et al., 2004). Detailed descriptions of the chemical basis and development of SCMs is available in a number of sources (e.g., Sposito, 1984; Davis and Kent, 1990; Stumm and Morgan, 1996). Brown (2009) presents an alternative framework in which to evaluate thermodynamic properties of chemisorption reactions. In this approach, the bond valence model is invoked to assess constraints on the types of surface bonds that can form. A primary constraint that is applied in this framework is the valence sum rule, which states that the sum of bond valences around any atom should be equal to the atomic valence. In contrast to the common formalism within most SCMs of providing explicit descriptions and accounting of electrostatic interactions across the surface, the bond valence model incorporates all electrostatic effects into the bonds that are formed. The extent of adhesion between atoms in the solid and solution (i.e., surface complex formation) is evaluated by ensuring that the valence sum rule is obeyed around each atom in the system. Although the bond valence model provides a way to evaluate the chemical suitability of proposed bonding structures, this approach is not currently implemented in a mathematical framework that can be merged with existing thermodynamic formalisms for calculating solution complexation and precipitation reactions. Until this occurs, the formalisms employed for describing surface bond formation in current SCMs will likely continue to be employed.

Goldberg et al. (2007) and Goldberg and Criscenti (2008) have highlighted potential limitations for the application of SCMs to project chemisorption reactions in soil systems. Reaction expressions depicting chemisorption reactions presume knowledge of the types and availability of solid surfaces in the system being modeled. The heterogeneity of soil systems hampers selection of appropriate reactive surfaces and adsorption site densities to be explicitly represented in an SCM. Two proposed modeling strategies to address this difficulty include the component additivity (CA) and general composite (GC) approaches. The CA approach is based on the summed contribution of individual adsorbent phases, and, thus, makes use of individual reaction expressions developed for each solute–adsorbent pair from model experimental systems. The predictive capability of the CA approach will be highly dependent on knowledge of the specific adsorbents active in solute chemisorption for a particular soil. The GC approach is based on the assumption that the assemblage of adsorbent phases within the soil can be represented by "generic" surface functional groups with reaction stoichiometries and formation constants derived from experimental adsorption data for individual soil samples. The utility of the GC approach over purely empirical partitioning coefficients is that the chemisorption model can be coupled to equilibrium models of aqueous solute speciation and solubility. Thus, chemisorption model projections will be sensitive to changes in aqueous solute speciation and/or solution saturation state relative to homogeneous precipitation of the solute. In addition, the GC approach allows for explicit accounting of electrostatic interactions across the surface (based on laboratory measurements of surface-charging behavior) or incorporation of these interactions implicitly within the optimized surface complexation constant determined for a given system.

Another limitation of SCMs has been the standard state used to express the activities of surface sites and surface species. Kulik (2002a, 2002b) and Sverjensky (2003) have illustrated that the hypothetical standard state employed in current model formulations results in equilibrium constants that are directly dependent on properties of the adsorbent such as the site density and surface area. Thus, surface protonation and stability constants for solute adsorption developed for a particular mineral specimen, for example, goethite, at a given solid:solution ratio may not be directly transferable to modeling adsorption of another goethite specimen with different specific surface area and/or solid:solution ratio. Both authors have proposed an alternative standard state that can be used to develop surface protonation and complex stability constants that are independent of individual adsorbent properties and experimental conditions. The new standard-state convention can be used to convert conditional stability constants derived using law-of-mass action (LMA) algorithms (Sverjensky, 2003) or the Gibbs energy minimization (GEM) approach (Kulik, 2002b). As illustrated by Kulik (2002a), surface complexation constants derived using the LMA convention can be converted to the GEM convention using the newly proposed standard state for surface species. In addition, Kulik (2002b) illustrates that the GEM approach also allows one to factor in which adsorbents are stable or metastable, and how the site types are distributed between the adsorbent surfaces. Given our understanding that the suite of mineral components, and their inherent solubility and surface complexation properties, undergoes dynamic transitions in soil systems

(e.g., Chadwick and Chorover, 2001), the GEM approach presents an attractive alternative for modeling low-temperature aqueous chemistry reactions.

A final issue of concern for modeling solute adsorption onto mineral surfaces is development of approaches to represent the transition from surface loadings less than to greater than monolayer coverage. As discussed previously, the formation of a mononuclear surface complex may represent an intermediate step to the development of a polynuclear species with 2D or 3D order, commonly referred to as surface precipitates. The composition of surface precipitates will be governed by the participating elements from the solution and the host mineral surface. Modeling strategies have been developed within the convention of the LMA-SCM (Farley et al., 1985; Katz and Hayes, 1995), and they have been employed to replicate sorption data for wide ranges of surface coverage assuming either the formation of a pure solid or solid solution incorporating the solute species (e.g., Charlet and Manceau, 1992; Román-Ross et al., 2006). For these applications, assumptions are made about the fraction of the adsorbent that participates in formation of binary or higher-order solid solutions, as well as the ideality of the modeled solid solutions relative to the endmember structures. Curti et al. (2005) and Kulik (2006) have proposed alternative strategies to represent and select the solid solutions that may realistically form based on GEM. Significant model development and testing remains before much confidence can be placed on projections of surface precipitation or surface solid–solution formation in soil systems.

17.5 Observations of Cation and Anion Solid-Phase Partitioning in Soils

The distribution of cations and anions between the solid–water interfaces present in soils is governed by chemical characteristics of the aqueous and solid species and the chemistry of the soil solution. The distribution between adsorbed and precipitated forms of these constituents will also be influenced by the total concentrations of reactive species, as well as the influence of biotic processes that govern the distributions of redox-sensitive reactants within the soil. As stated earlier, properties of the soil solution, adsorbate, and sorption sites will govern the types of chemisorption reactions that may take place. The solution speciation of cations/anions, as controlled by solution complexation reactions and the distribution of ion oxidation states will govern the saturation state of the soil solution with respect to potential precipitates. As one would anticipate, it is difficult to project what reactions may dominate solid-solution partitioning for ions without full knowledge of the solution chemistry and solid surface properties. For some ions, such as the alkali metals or halide anions, this task is simplified due to the predominance of a single oxidation state in natural systems and the high solubility of potential precipitate phases. However, this simplicity rapidly diminishes with departure from these bracketing groups within the periodic table of elements. Thus, the remaining discussion

will focus on illustrative examples of the types of chemical factors that influence chemisorption–precipitation versus an exhaustive review of the periodic properties of the elements. In order to illustrate the disparity in potential reaction paths for ions in soil solutions, the controls on solid-water partitioning for nickel (Ni) and selenium (Se) are reviewed in the following sections. These elements illustrate both the range of chemisorption–precipitation reactions that influence ion partitioning in soils and the degree of complexity inherent in monitoring and projecting these reactions in dynamic soil systems.

17.5.1 Nickel

Nickel is not commonly seen as a critical element in agricultural or environmental systems. However, as noted in a recent review by Hansch and Mendel (2009), it is considered to be an important micronutrient, and, for specific plant species, deficiencies in nickel availability may limit plant growth for agricultural (Bais et al., 2006) or phytoremediation applications (Wood et al., 2006). At higher concentrations in soil solutions, nickel may also exert a toxic effect on plant species (e.g., Khan et al., 2006). In ambient aqueous systems, nickel exists in the divalent oxidation state and is not subject to oxidation-state transformations under typical conditions. Nickel predominantly exists as a cationic species (Ni^{2+}) or various hydrolysis species (e.g., $NiOH^+$) at near-neutral pH (Baes and Mesmer, 1986). However, nickel may also form dissolved complexes in the presence of high concentrations of inorganic ions such as carbonate/bicarbonate and sulfate (Hummel and Curti, 2003; Chen et al., 2005) or organic ligands such as natural/synthetic carboxylic acids and dissolved humic compounds (Bryce and Clark, 1996; Baeyens et al., 2003; Strathmann and Myneni, 2004). It is anticipated that nickel may form complexes with dissolved sulfide under sulfate-reducing conditions, although the current state of knowledge is insufficient to ascertain the relative importance of these species in aqueous systems (Thoenen, 1999). The formation of solution complexes, especially with organic ligands (e.g., Strobel, 2001; Fröberg et al., 2006), may limit sorption of nickel to mineral surfaces in soils (see Section 17.5.1.2).

The solid-solution partitioning of Ni is chosen as a point of reference for ions that are not subject to a change in valence state due to shifts in redox characteristics in the soil solution. However, redox chemistry may exert an indirect control on Ni partitioning by governing the distribution of solid surfaces that may influence chemisorption reactions, as well as the distribution and concentrations of solution species that limit the solubility of Ni-bearing precipitates that may form. Nickel also provides an example of an ion that demonstrates a wide range of solid-solution partitioning behavior including adsorption, surface precipitation, and homogeneous precipitation.

17.5.1.1 Chemisorption

Nickel displays the commonly observed chemisorption behavior for cation adsorption onto soil minerals with uptake onto solid surface increasing with pH. Typically, the increase in

adsorption occurs over a narrow pH range, and the resultant plot of adsorption data has been given the name "adsorption edge" (e.g., Stumm, 1992; Sparks, 2003). Adsorption of nickel in soil environments is dependent on pH, temperature, and type of adsorbent (minerals or organic matter), as well as the concentration of aqueous complexing agents, competition from other adsorbing cations, and the ionic strength in the soil solution. Nickel has been shown to adsorb onto many solid components encountered in soils, including iron/manganese oxides, clay minerals (Dähn et al., 2003; Bradbury and Baeyens, 2005), and solid organic matter (Nachtegaal and Sparks, 2003). Sorption to iron/manganese oxides and clay minerals has been shown to be of particular importance for controlling nickel mobility in subsurface systems. The relative affinity of these individual minerals for nickel uptake will depend on the mass distribution of the adsorbent minerals as well as the predominant geochemical conditions (e.g., pH and Ni aqueous speciation). For example, the pH-dependent distribution of nickel between iron and manganese oxides (hydrous ferric oxide [HFO] and a birnessite-like mineral [nominally MnO_2]) for a representative soil solution composition is shown in Figure 17.3. Based on the available compilations for surface complexation constants onto these two solid phases (Dzombak and Morel, 1990; Tonkin et al., 2004), one would project the predominance of nickel sorption to MnO_2 at more acidic pH and the predominance of HFO (or ferrihydrite) at more basic pH. With increasing mass of MnO_2, the solid-phase speciation of nickel will be progressively dominated by sorption to this phase. There are examples of the relative preference of nickel sorption to manganese oxides over iron oxides for natural systems (e.g., Manceau et al., 2002; Kjoller et al., 2004; Manceau et al., 2007). As shown in Figure 17.3b, the presence of

humic/fulvic compounds along with simple organic acids may not significantly influence the extent of nickel sorption to oxide surfaces. This behavior is predicated on the availability of sorption sites at concentrations significantly greater than competing ligands in solution.

In contrast, the presence of ligands that form highly stable or multidentate solution complexes may exert greater influence on the degree of nickel sorption to oxide surfaces. As shown in Figure 17.3c, nickel sorption may be inhibited (or nickel desorption enhanced) through the formation of solution complexes with some classes of organic ligands such as hydroxamate siderophores (Dimkpa et al., 2008). Published stability constants are available for a commonly occurring hydroxamate siderophore, desferrioxamine B (DFB; Farkas et al., 1999). In order to model the nickel speciation in Figure 17.3c, equilibrium expressions for the protonation of DFB and its complexation with nickel were derived from published conditional stability constants (Evers et al., 1989; Farkas et al., 1999; see Table 17.2). As discussed previously, the production of siderophores provides a mechanism for plants/organisms to overcome the low solubility of metals such as nickel in soil systems where the abundance of solid sorption sites is high. As shown in Figure 17.3, the stability of solution complexes with carboxylic acids (natural or synthetic) and the relative availability of sites for chemisorption affect that degree of competition. For the hypothetical conditions modeled in Figure 17.3, only strong complexation with DFB significantly inhibits Ni adsorption to surface of hydrous Fe/Mn oxides. This behavior is similar to that observed in experimental systems for the synthetic chelating agent, EDTA (e.g., Bryce and Clark, 1996; Nowack et al., 1997). The model calculations presented in Figure 17.3 do not include the possible formation of ternary surface
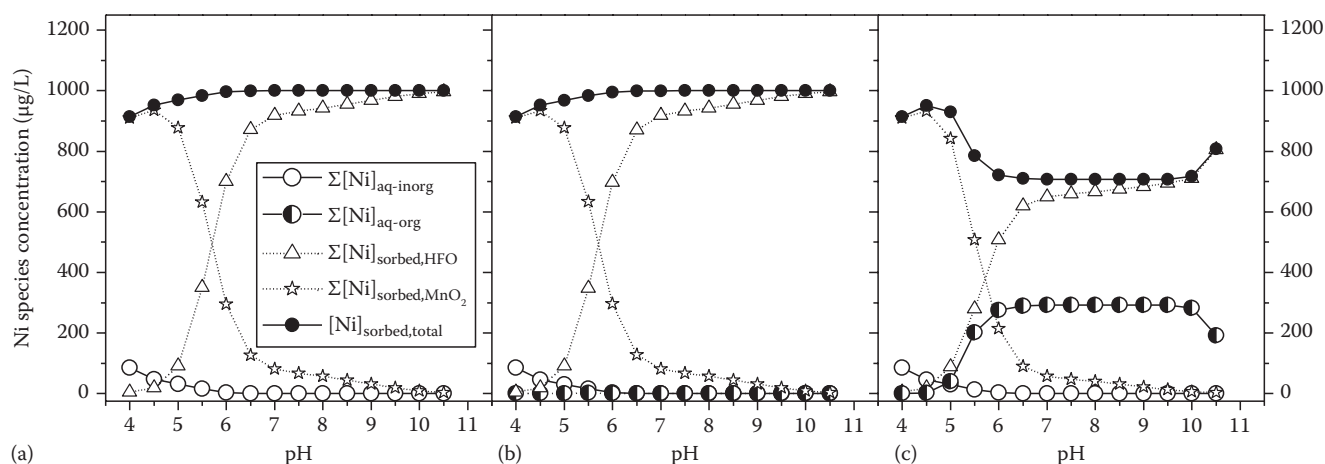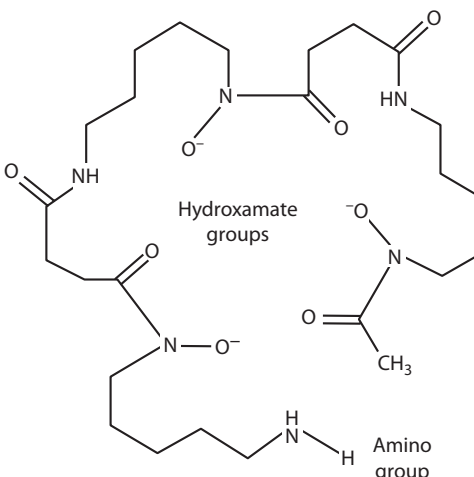
FIGURE 17.3 (a) Nickel sorption as a function of pH in the presence of a hypothetical soil with iron and manganese oxides reflective of the crustal abundance of these elements (Schulze, 2002; assumed 30% porosity with 185.0 g HFO L^{-1} and 1.66 g MnO_2 L^{-1}); no organic ligands. (b) Same conditions as in (a) but with 20 mg L^{-1} dissolved organic carbon (DOC; humics and fulvics) and 25 μM each of citrate, malonate, and oxalate. (c) Same conditions as in (a), but with 5 μM DFB added as the sole organic ligand. Nominal water composition: 0.005 mol L^{-1} NaCl, 0.001 mol L^{-1} K_2SO_4, 0.001 mol L^{-1} $MgNO_3$, 0.001 mol L^{-1} $CaCO_3$, and 17 μM Ni (1 mg Ni L^{-1}). Model predictions using Visual MINTEQ (2009) Version 2.61 with available surface complexation parameters derived from Dzombak and Morel (1990) and Tonkin et al. (2004), and metal-DOC speciation using default settings in the DOC (SHM) module (Gustafsson, 2001); kaolinite set as an "infinite" solid for pH titration. This model does not make use of the alternate hypothetical reference states proposed by Kulik (2002a, 2002b) or Sverjensky (2003).

TABLE 17.2 Complexation Reactions between Ni(II) and DFB, a Microbially Produced Trihydroxamate Siderophore in the Soil Rhizosphere

Acid–Base Reactions		
Stoichiometry	Log K (25°C)	Structure of DFB
$H_3DFB + H^+ = H_4DFB^+$	8.32	
$H_3DFB - H^+ = H_2DFB^-$	−8.74	
$H_3DFB - 2H^+ = HDFB^{2-}$	−17.84	
$H_3DFB - 3H^+ = DFB^{3-}$	−27.95	
Complexation Reactions		
$Ni^{2+} + H_3DFB - 3H^+ = NiDFB^-$	−20.41	
$Ni^{2+} + H_3DFB - 2H^+ = NiHDFB$	−0.95	
$Ni^{2+} + H_3DFB - H^+ = NiH_2DFB^+$	−1.89	
$Ni^{2+} + H_3DFB = NiH_3DFB^{2+}$	3.90	

Equilibrium reaction expressions and associated constants for protonation and complexation with nickel were derived from Evers et al. (1989) and Farkas et al. (1999), respectively. All constants have been recalculated to zero ionic strength using the Davies equation for the estimation of single-ion activities (Stumm and Morgan, 1981, p. 135). Reaction expressions from the referenced sources were converted to the reaction format required within the Visual MINTEQ (2009) database using H^+, Ni^{2+}, and H_3DFB as components (MINTEQA2/PRODEFA2, 1999).

complexes with DFB or other multidentate ligands that might serve as a bridging ligand between Ni and an oxyhydroxide surface (e.g., Nowack and Sigg, 1996; Nowack, 2002).

As previously noted, adsorption of nickel onto mineral surfaces may serve as a precursory step to the formation of trace precipitates that reduce the potential for desorption with changes in soil solution chemistry. This may be realized through the nucleation and growth of surface precipitates on clay mineral surfaces due to continued uptake of nickel. Three categories of surface precipitation have been identified in experimental systems: (1) heterogeneous nucleation of pure Ni-bearing phases such as hydroxides (e.g., Scheinost et al., 1999), (2) coherent growth of Ni-bearing solid phases that mimic the structure of the adsorbent (e.g., Dähn et al., 2002), and (3) nucleation of mixed-metal precipitates within the solid–water interface (e.g., Scheckel and Sparks, 2000, 2001; Scheckel et al., 2000; Dähn et al., 2006). This type of process may compete with other adsorption processes, such as ion exchange, depending on the prevailing soil solution chemistry and characteristics of the clay mineral (e.g., Elzinga and Sparks, 2001).

17.5.1.2 Precipitation

Nickel may be immobilized within soils through formation of pure nickel precipitates such as hydroxides, silicates, or sulfides (Merlen et al., 1995; Mattigod et al., 1997; Scheidegger et al., 1998; Thoenen, 1999; Scheinost and Sparks, 2000; Peltier et al., 2006) or through coprecipitation with other soil-forming minerals such as silicates, iron oxides/sulfides, or carbonates (Manceau et al., 1985; Manceau and Calas, 1986; Huerta-Diaz

and Morse, 1992; Ford et al., 1999a; Hoffmann and Stipp, 2001). Predicted nickel concentrations in the presence of sulfide for several potential pure nickel precipitates are shown in Figure 17.4a. These data suggest that phyllosilicate and LDH precipitates (incorporating aluminum) may limit Ni solubility in soils. These data also point to the limited capability of pure nickel carbonates and hydroxides in controlling dissolved nickel concentrations to sufficiently low values except under very alkaline conditions. For LDH precipitates, calorimetric measurements have demonstrated that the interlayer anion exerts significant influence on the precipitate stability (Peltier et al., 2006). In the presence of dissolved sulfide, the precipitation of a nickel sulfide may plausibly control the concentration of dissolved nickel. The influence of dissolved silicate on the stability of Ni-bearing precipitates is shown in Figure 17.4b. According to these data, nickel-bearing phyllosilicate and/or LDH precipitates possess large stability fields indicating their relative importance to controlling nickel solubility under a range of conditions. These calculations point to the importance of dissolved aluminum and silicon concentrations in the soil solution relative to the potential sequestration of nickel via precipitation (Ford et al., 1999b; Scheinost et al., 1999). As discussed in Section 17.5.1.1, the formation of these nickel-bearing precipitates may be facilitated through initial adsorption onto clay minerals within the soil.

Attenuation of nickel may also occur via coprecipitation during the formation of (hydr)oxides or sulfides of iron. These minerals have been observed to form at the boundaries between oxidizing and reducing zones within soil systems. There are numerous laboratory and field observations that demonstrate
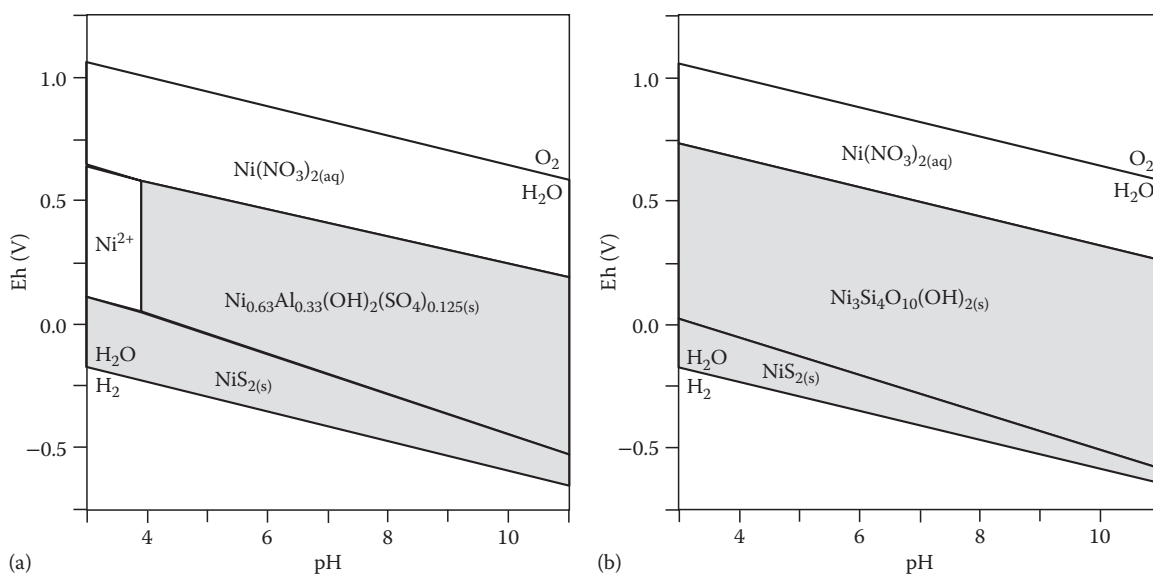
FIGURE 17.4 Eh–pH diagrams for nickel at 25°C. (a) System Ni–H_2O–Ca–Al–NO_3–HCO_3–SO_4 (2 mg Ni L^{-1}; 40 mg Ca L^{-1}; 3 mg Al L^{-1}; 6 mg NO_3 L^{-1}; 60 mg HCO_3 L^{-1}; 100 mg SO_4 L^{-1}). Stability fields for solids are shaded gray (Vaesite = NiS_2). (b) Same system plus 3 mg Si L^{-1}. Thermodynamic data for $Ni_3Si_4O_{10}(OH)_2$ and $Ni_{0.63}Al_{0.33}(OH)_2(SO_4)_{0.125}$ are from Peltier et al. (2006). (Note that the solubility of the Ni–Al–SO_4 LDH was adjusted to correct for charge imbalance for the chemical structure published in Peltier et al. [2006].)

the capacity of these precipitates for nickel uptake (Schultz et al., 1987; Huerta-Diaz and Morse, 1992; Coughlin and Stone, 1995; Ford et al., 1997, 1999a). Under these circumstances, the solubility of nickel will depend on the stability of the host precipitate phase. For example, iron oxide precipitates may alternatively transform to more stable forms (Ford et al., 1997), stabilizing coprecipitated nickel over the long term, or these precipitates may dissolve concurrently with changes in soil solution redox chemistry (e.g., Zachara et al., 2001).

17.5.2 Selenium

Selenium provides an example of an element that is subject to significant changes in solution and solid speciation as influenced by changes in system redox. Selenium is a metalloid exhibiting physical and chemical properties between that of metals and nonmetals. It chemically resembles sulfur and exists in organic and inorganic chemical forms. Inorganic species include selenide [Se(–II)], elemental selenium [Se(0)], selenite [Se(IV)], and selenate [Se(VI)] (e.g., Fernandez-Martınez and Charlet, 2009). Organic species include methylated compounds, selenoamino acids, and selenoproteins and their derivatives. The speciation of selenium is greatly influenced by the pH and redox conditions of the environment (Figure 17.5). For example, Se(–II) exists in a reducing environment as hydrogen selenide (H_2Se) and as metal selenides. Reduction of selenate to selenite and Se(0) has been shown to decrease its mobility in saline, mildly alkaline water (White et al., 1991). When dissolved in water, H_2Se can oxidize to elemental selenium. Elemental selenium is stable in a reduced environment, but it can be oxidized to selenite and to selenate by a variety of microorganisms (Sarathchandra and Watkinson, 1981; Dowdle and Oremland, 1998).
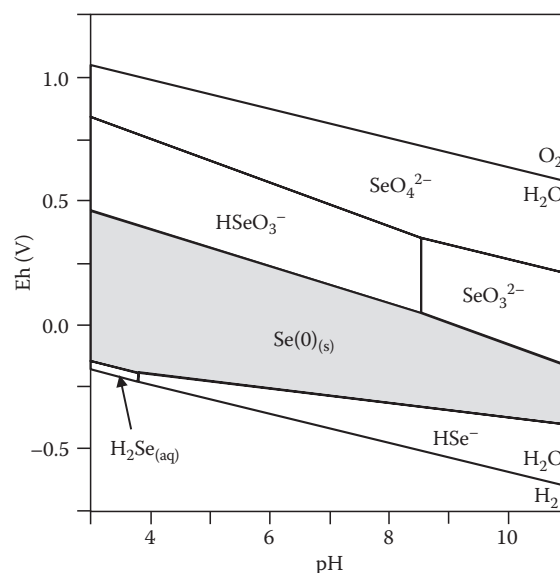
FIGURE 17.5 Eh–pH diagram for selenium at 25°C using thermodynamic data from Seby et al. (2001). $\Sigma Se = 10^{-5}$ M (790 μg L^{-1}). Solid stability field for elemental selenium is shaded gray.

Various strains of bacteria have been identified to facilitate selenate reduction in soil systems. Two microbial processes, namely, methylation of selenium and reduction of both selenate and selenite to Se(0), have a major influence on the fate and mobility of this element in the environment (Dungan and Frankenberger, 1999; Dungan et al., 2003). Methylation of selenium, and subsequent selenium volatilization, leads to dissipation of soil selenium to the atmosphere. Environmental factors such as the existing microbial community, pH,

temperature, moisture, and organic amendments control the rate of selenium volatilization from seleniferous soils (Frankenberger and Karlson, 1989; Zhang and Frankenburger, 1999). Under flooding conditions, part of the methylated selenium may be transported in water, thus decreasing selenium volatilization to the atmosphere (Zhang and Frankenburger, 1999). The addition of organic amendments to soils has been reported to stimulate indigenous microbes to methylate selenium (Abu-Erreish et al., 1968; Frankenberger and Karlson, 1989); whereas, organic substrates added to ponded sediments (submerged soils) have been found to accelerate the reduction of selenate and selenite to Se(0) (Tokunaga et al., 1996) and similar effects were reported in laboratory batch experiments (Zhang et al., 2003).

Mechanisms for selenium reduction by microbes are complex as it occurs under both aerobic (Lortie et al., 1992) and anaerobic conditions (Oremland et al., 1989, 1990; Tomei et al., 1992). Both dissimilatory and detoxification mechanisms are possible (Oremland, 1994). The occurrence of sequential reduction of selenate to selenite and then to Se(0) is suggested after amendment of contaminated soils with barley straw under field capacity moisture conditions (Camps Arbestain, 1998). The rate of selenate to selenite reduction in wastewaters is proportional to their respective concentrations in solution and also to the amount of the microbial biomass (Rege et al., 1999). A recent study has documented the occurrence of both intracellular and extracellular Se(0) granules in three phylogenetically and physiologically distinct bacteria that are able to respire selenium oxyanions, suggesting that this phenomenon appears to be widespread among such bacteria (Oremland et al., 2004). The metal sites of selenate reductase from *Thauera selenatis* have been characterized (Maher et al., 2004); the enzyme was found to contain Se in a reduced form (probably organic) and the Se is coordinated to both a metal (probably Fe) and carbon. Assessment of Se(IV) and Se(0) reduction in anaerobic microcosms demonstrated the sequential formation of Se(0) and ultimately Se(–II) during incubation with elemental selenium and/or lactate as electron donors for microbial reduction (Herbel et al., 2003). Dissolved Se(–II) did not accumulate in pore water during incubation due to precipitation with ferrous iron to form FeSe(s), which was determined via solid-phase characterization using x-ray absorption spectroscopy.

17.5.2.1 Chemisorption

As noted above, the most common species of selenium in soil solution are in the form of oxyanions. Thus, selenium adsorption onto soil minerals tends to be greatest under acidic pH with a gradual decrease as pH increases. In contrast to cation adsorption, the decrease in anion adsorption with increasing pH is more gradual, and the resultant plot of these adsorption data has been given the name "adsorption envelope" (e.g., Stumm, 1992; Sparks, 2003). Selenate and selenite adsorption behavior on individual soil minerals (Fe, Al, and Mn oxides, kaolinite, and calcite) and whole soils throughout the United States has been documented (Neal et al., 1987; Neal and Sposito, 1989; Zachara et al., 1994). This research demonstrated the importance of Fe- and Al-oxide surfaces for selenium adsorption onto soils and highlighted the dependence of the extent of adsorption on the pH of soil solution and the presence of anions that compete for adsorption sites. Additional review of the published literature is provided in the following.

Selenate has been shown to behave like sulfate with minimal adsorption and high mobility (Goldberg and Glaubig, 1988; Neal and Sposito, 1989); whereas, selenite behaves analogously to phosphate, with greater adsorption than selenate (Neal et al., 1987; Barrow and Whelan, 1989a; Zhang and Sparks, 1990). Strawn et al. (2002) have observed selenite and selenate associated with iron-bearing minerals in acid-sulfate soils. In these soils, selenite was predominantly associated with iron oxyhydroxides and selenate was associated with jarosite, presumably in a coprecipitated form. Adsorption of selenite on goethite decreases with increasing pH, with decreasing selenite concentration, and with competing anions such as phosphate, silicate, citrate, molybdate, carbonate, oxalate, and fluoride (Balistrieri and Chao, 1987). More selenite is adsorbed onto montmorillonite than on kaolinite (Frost and Griffin, 1977). Selenite adsorption in seleniferous soils is decreased in the presence of sulfate, nitrate, and phosphate (Pareek et al., 2000). Selenite adsorption by aluminum hydroxides is adversely affected by organic acids (Dynes and Huang, 1997). Desorption of selenate is faster and more nearly complete than selenite after adsorption and incubation in soil (Barrow and Whelan, 1989b). Differences in the stability of selenate and selenite chemisorption to ferrihydrite are illustrated in Figure 17.6. Selenite adsorption predominates for most of the modeled pH range, but there is a small stability field for selenate adsorption in the pH range 7 < pH < 9, in part, due to the pH-dependent distribution of selenium oxyanion oxidation state. As shown, model predictions indicate limited competition from competing oxyanions common to soil solutions (e.g., carbonate, silicate, and sulfate) when available surface-site densities are far from saturation.

Selenite selectively adsorbs at the carbonate (CO_3^{2-}) site on calcite ($CaCO_3$) via ionic exchange, forming a two-dimensional solid solution of the form $Ca(SeO_3)_x(CO_3)_{1-x}$ at the interface; under identical chemical conditions, selenate adsorption is inhibited (Cheng et al., 1997). An earlier study showed selenate substitution in calcite may also be supported under appropriate conditions (Reeder et al., 1994).

Mechanisms of selenium adsorption have been studied from both macroscale batch and microscale spectroscopic approaches. The presence of either selenate or selenite lowers the electrophoretic mobility and decreases the PZC of ferrihydrite and goethite, suggesting inner-sphere complexation for both selenate and selenite species (Su and Suarez, 2000). Both in situ attenuated total reflectance–Fourier transform infrared (ATR–FTIR) and diffuse reflectance infrared Fourier transform (DRIFT) spectra indicated formation of bidentate complexes of selenate with ferrihydrite, and the DRIFT spectra of selenite on goethite indicated formation of bridging bidentate complex of selenite.
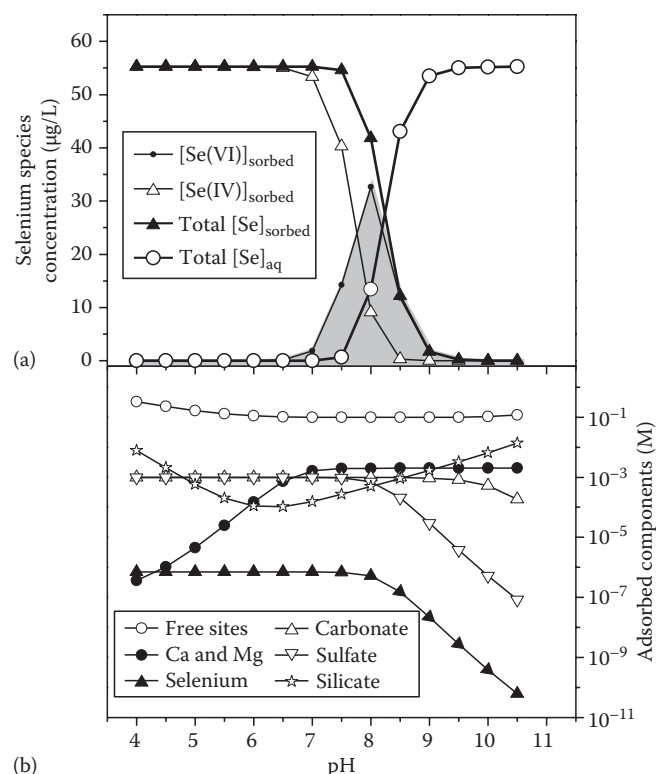
FIGURE 17.6 (a) Selenium species adsorption as a function of pH in the presence of a hypothetical soil with iron oxide content reflective of the crustal abundance (Schulze, 2002; assumed 30% porosity with 185.0 g HFO L⁻¹). (b) Total component adsorption for modeled water composition. Nominal water composition: 0.005 M NaCl, 0.001 M K₂SO₄, 0.001 M MgNO₃, 0.001 M CaCO₃, and 0.7 μM Se (55.2 μg Se L⁻¹). Model predictions using Visual MINTEQ (2009) Version 2.61 with available surface complexation parameters derived from Dzombak and Morel (1990) at a fixed Eh = 500 mV; kaolinite set as an "infinite" solid for pH titration. This model does not make use of the alternate hypothetical reference states proposed by Kulik (2002a, 2002b) or Sverjensky (2003).

These results are consistent with an earlier in situ extended x-ray absorption fine structure (EXAFS) spectroscopic study (Manceau and Charlet, 1994) that shows that selenate forms an inner-sphere binuclear bridging surface complex on hydrous ferric oxide and goethite. On the contrary, an earlier EXAFS study (Hayes et al., 1987) concluded that selenate forms an outer-sphere surface complex on goethite. A recent combined data set of Raman and ATR–FTIR spectra indicate that both inner- and outer-sphere surface complexes of selenate occur on goethite, as predominantly monodentate inner-sphere surface complexes at pH < 6, and as predominantly outer-sphere surface complexes at pH > 6 (Wijnja and Schulthess, 2000).

17.5.2.2 Precipitation

Selenium has the potential of forming precipitates for all of its oxidation states (Seby et al., 2001; see Figure 17.7). For selenate and selenite, this includes precipitates with common major cations in soil solutions (Ca, Mg) as well as transition metals (Fe, Mn) or heavy metals that may be anticipated in contaminated
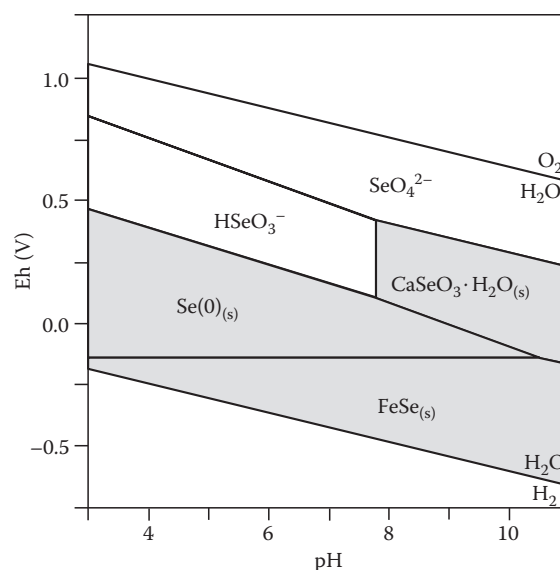
FIGURE 17.7 Eh–pH diagram for selenium at 25°C using thermodynamic data from Seby et al. (2001). System Se–H₂O–Fe–Ca, with ΣSe = 10⁻⁵ M (790 μg L⁻¹), ΣFe = 10⁻⁴ M (5.6 mg L⁻¹), and ΣCa = 10⁻² M (400 mg L⁻¹). Solid stability fields for elemental selenium, hydrous calcium selenite, and ferrous selenide are shaded gray. FeSe₂ was suppressed (data not available in the Seby et al., 2001); however, the stability field of the diselenide would be intermediate between elemental selenium and FeSe.

soils (Rai et al., 1995; Sharmasarkar et al., 1996). These phases are anticipated to primarily be significant in situations where selenium concentrations are highly elevated. Reduction to form elemental selenium [Se(0)] can result in very low concentrations of dissolved selenium. In general, it has been observed that selenium reduction to insoluble Se(0) results in immobilization and stabilization of this element in the soil matrix, since the reoxidation reaction of Se(0) to soluble selenate and selenite is relatively slow (Tokunaga et al., 1994; Zawislanski and Zavarin, 1996; Dowdle and Oremland, 1998; Losi and Frankenberger, 1998). Abiotic reduction of selenite to Se(0) was also suggested in SRB biofilms (Hockin and Gadd, 2003). Further reduction to selenide [Se(–II)] can lead to precipitation of metal selenides, including ferrous iron and manganese selenides, similar to the formation of metal sulfides under sulfate-reducing conditions. As demonstrated by Scheinost and Charlet (2008a) and Scheinost et al. (2008b), selenite may be reduced and precipitated via abiotic reactions with Fe(II)-bearing minerals in soils. The propensity to form elemental selenium versus ferrous-selenide precipitates will likely be governed by pH, major ion composition of the soil solution, and the solubility of ferrous iron.

While not commonly observed, it is anticipated that many suboxic soil environments (e.g., redoximorphic soils) contain green rust, which is a mixed ferrous, ferric hydroxide that also contains interlayer anions such as sulfate and carbonate in its structure (Feder et al., 2005). Identification of green rusts in soils

is hampered by the rapid oxidation of green rusts by atmosphere oxygen, and for this reason, they have not been commonly reported. However, recent thermodynamic and spectroscopic studies give direct evidence for the existence of green rusts in soils (Hansen et al., 1994; Trolard et al., 1997, 2007; Feder et al., 2005). Due to high reactivity, green rust minerals are envisioned as potential reducing agents of a number of anions such as nitrate, chromate, and selenate (Johnson and Bullen, 2003). Direct evidence for the formation of reduced selenium species under anoxic conditions via abiotic redox reactions with sulfate green rust was provided using x-ray absorption near-edge spectroscopy (XANES) and Fourier transform EXAFS spectroscopy (Myneni et al., 1997). The mechanism of selenate reduction was described by the following equation:

$$HSeO_4^- + 4Fe^{(II)}_4Fe^{(III)}_2(OH)_{12}SO_4 \cdot 3H_2O$$

$$\rightarrow HSe^- + 8Fe_3O_4 + 4SO_4^{2-} + 8H^+ + 32H_2O \qquad (17.1)$$
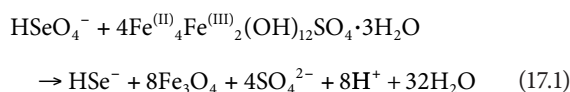
in which sulfate green rust was oxidized to form magnetite, whereas selenate was reduced to Se(0) and subsequently to selenide. In addition, a laboratory study has demonstrated that a significant fraction of dissolved selenate can be coprecipitated with Fe(II) and Fe(III) ions to form Fe(II)–Fe(III) hydroxyselenate green rust with simultaneous reduction of an equal amount of selenate anions to selenite anions (Refait et al., 2000; similar to partitioning to LDH phases, You et al., 2001). In the subsurface environment, selenate reduction by coprecipitation and adsorption pathways can occur when reducing conditions develop (Pickering et al., 1995). In the coprecipitation–reduction pathway, reductive dissolution of Fe(III) oxides precipitates green rust with selenate followed by selenate reduction to Se(0) and selenide. Reduction of selenium oxyanions to Se(0) has also been observed in the presence of iron sulfides (Bruggeman et al., 2005) and ferrous hydroxide (Zingaro et al., 1997).

In summary, the chemical behavior of nickel and selenium point to the importance of identifying aqueous and solid speciation in confirming the processes controlling element partitioning for a given soil system. Chemical equilibrium models provide a tool for assessing potential endpoints for these partitioning reactions, but the accuracy of these projections will depend on the adequacy of the supporting thermochemical database for relevant solution and solid species. Advances have been made in modeling chemisorption and precipitation reactions, in part, due to developments in analytical approaches to identify aqueous and solid speciation for trace elements in soils. For trace-element partitioning, it is increasingly understood that the initial form of the chemisorbed solute may represent a metastable state relative to its long-term mobility. Due to the internal influence of biotic respiration in concert with external fluxes of aqueous and gaseous components physically transported through the soil matrix, the composition of the soil solution and associated solids will change in time. Ultimately, knowledge of the rates and extents of these aqueous and mineralogic changes will prove critical for understanding the dynamics of solute partitioning for soil systems.

References

Abu-Erreish, G.M., E.I. Whitehead, and O.E. Olson. 1968. Evolution of volatile selenium from soils. Soil Sci. 106:415–420.

Baes, C.F., and R.E. Mesmer. 1986. The hydrolysis of cations. Krieger Publishing Company, Malabar, FL.

Baeyens, B., M.H. Bradbury, and W. Hummel. 2003. Determination of aqueous nickel-carbonate and nickel-oxalate complexation constants. J. Solution Chem. 32:319–339.

Bais, H.P., T.L. Weir, L.G. Perry, S. Gilroy, and J.M. Vivanco. 2006. The role of root exudates in rhizosphere interactions with plants and other organisms. Annu. Rev. Plant Biol. 57:233–266.

Balistrieri, L.S., and T.T. Chao. 1987. Selenium adsorption by goethite. Soil Sci. Soc. Am. J. 51:1145–1151.

Barrow, N.J., and B.R. Whelan. 1989a. Testing a mechanistic model. VII. The effects of pH and of electrolyte on the reaction of selenite and selenate with a soil. J. Soil Sci. 40:17–28.

Barrow, N.J., and B.R. Whelan. 1989b. Testing a mechanistic model. VIII. The effects of time and temperature of incubation on the sorption and subsequent desorption of selenite and selenate by a soil. J. Soil Sci. 40:29–37.

Belnap, J., C.V. Hawkes, and M.K. Firestone. 2003. Boundaries in miniature: Two examples from soil. Bioscience 53:739–749.

Birkham, T.K., M.J. Hendry, L.I. Wassenaar, and C.A. Mendoza. 2007. A transient model of vadose zone reaction rates using oxygen isotopes and carbon dioxide. Vadose Zone J. 6:67–76.

Bradbury, M.H., and B. Baeyens. 2005. Modelling the sorption of MnII, CoII, NiII, ZnII, CdII, EuIII, AmIII, SnIV, ThIV, NpV and UVI on montmorillonite: Linear free energy relationships and estimates of surface binding constants for some selected heavy metals and actinides. Geochim. Cosmochim. Acta 69:875–892.

Bradbury, M.H., and B. Baeyens. 2009. Sorption modelling on illite. Part II: Actinide sorption and linear free energy relationships. Geochim. Cosmochim. Acta 73:1004–1013.

Brown, I.D. 2009. Recent developments in the methods and applications of the bond valence model. Chem. Rev. 109:6858–6919.

Brown, G.E., V.E. Henrich, W.H. Casey, D.L. Clark, C. Eggleston, A. Felmy, W. Goodman, M. Grätzel, G. Maciel, M.I. McCarthy, K.H. Nealson, D.A. Sverjensky, M.F. Toney, and J.M. Zachara. 1999. Metal oxide surfaces and their interactions with aqueous solutions and microbial organisms. Chem. Rev. 99:77–174.

Bruggeman, C., A. Maes, J. Vancluysen, and P. Vandemussele. 2005. Selenite reduction in Boom clay: Effect of FeS$_2$, clay minerals and dissolved organic matter. Environ. Pollut. 137:209–221.

Bryce, A.L., and S.B. Clark. 1996. Nickel desorption kinetics from hydrous ferric oxide in the presence of EDTA. Colloids Surf. A 107:123–130.

Cabaniss, S.E. 2008. Quantitative structure-property relationships for predicting metal binding by organic ligands. Environ. Sci. Technol. 42:5210–5216.

Cabaniss, S.E. 2009. Forward modeling of metal complexation by NOM: I. A priori prediction of conditional constants and speciation. Environ. Sci. Technol. 43:2838–2844.

Camps Arbestain, M. 1998. Effect of straw amendment and plant growth on selenium transfer in a laboratory soil-plant system. Can. J. Soil Sci. 78:187–195.

Chadwick, O.A., and J. Chorover. 2001. The Chemistry of pedogenic thresholds. Geoderma 100:321–353.

Chapelle, F.H., P.M. Bradley, D.R. Lovley, K. O'Neill, and J.E. Landmeyer. 2002. Rapid evolution of redox processes in a petroleum hydrocarbon-contaminated aquifer. Ground Water 40:353–360.

Charlet, L., and A. Manceau. 1992. X-ray absorption spectroscopic study of the sorption of CrIII at the oxide-water interface. II. Adsorption, coprecipitation, and surface precipitation on hydrous ferric oxide. J. Colloid Interface Sci. 148:443–458.

Chefetz, B., and B.S. Xing. 2009. Relative role of aliphatic and aromatic moieties as sorption domains for organic compounds: A review. Environ. Sci. Technol. 43:1680–1688.

Chen, T., G. Hefter, and R. Buchner. 2005. Ion association and hydration in aqueous solutions of nickelII and cobaltII sulfate. J. Solution Chem. 34:1045–1066.

Cheng, L., P.F. Lyman, N.C. Sturchio, and M.J. Bedzyk. 1997. X-ray standing wave investigation of the surface structure of selenite anions adsorbed on calcite. Surf. Sci. 382:L690–L695.

Coughlin, B.R., and A.T. Stone. 1995. Nonreversible adsorption of divalent ions MnII, CoII, NiII, CuII, and PbII onto goethite: Effects of acidification, FeII addition and picolinic acid addition. Environ. Sci. Technol. 29:2445–2455.

Crumbliss, A.L., and J.M. Harrington. 2009. Iron sequestration by small molecules: Thermodynamic and kinetic studies of natural siderophores and synthetic model compounds. Adv. Inorg. Chem. 61:179–250.

Curti, E., D.A. Kulik, and J. Tits. 2005. Solid solutions of trace EuIII in calcite: Thermodynamic evaluation of experimental data over a wide range of pH and pCO_2. Geochim. Cosmochim. Acta 69:1721–1737.

Dähn, R., M. Jullien, A.M. Scheidegger, C. Poinssot, B. Baeyens, and M.H. Bradbury. 2006. Identification of neoformed Ni-phyllosilicates upon Ni uptake in montmorillonite: A transmission electron microscopy and extended X-ray absorption fine structure study. Clay. Clay Miner. 54:209–219.

Dähn, R., A.M. Scheidegger, A. Manceau, M.L. Schlegel, B. Baeyens, M.H. Bradbury et al. 2003. Structural evidence for the sorption of NiII atoms on the edges of montmorillonite clay minerals: A polarized X-ray absorption fine structure study. Geochim. Cosmochim. Acta 67:1–15.

Dähn, R., A.M. Scheidegger, A. Manceau, M.L. Schlegel, B. Baeyens, M.H. Bradbury et al. 2002. Neoformation of Ni phyllosilicate upon Ni uptake on montmorillonite: A kinetics study by powder and polarized extended X-ray absorption fine structure spectroscopy. Geochim. Cosmochim. Acta 66:2335–2347.

Davis, J.A., and D.B. Kent. 1990. Surface complexation modeling in aqueous geochemistry. Rev. Mineral. Geochem. 23:177–260.

Dimkpa, C., A. Svatoš, D. Merten, G. Büchel, and E. Kothe. 2008. Hydroxamate siderophores produced by *Streptomyces acidiscabies* E13 bind nickel and promote growth in cowpea *Vigna unguiculata* L. under nickel stress. Can. J. Microbiol. 54:163–172.

Dixon, J.B., and D.G. Schulze (eds.). 2002. Soil mineralogy with environmental applications. SSSA, Madison, WI.

Dove, P.M., and C.M. Craven. 2005. Surface charge density on silica in alkali and alkaline earth chloride electrolyte solutions. Geochim. Cosmochim. Acta 69:4963–4970.

Dowdle, P.R., and R.S. Oremland. 1998. Microbial oxidation of elemental selenium in soil slurries and bacterial cultures. Environ. Sci. Technol. 32:3749–3755.

Drever, J.I. 1982. The geochemistry of natural waters. Prentice-Hall, Englewood Cliffs, NJ.

Dungan, R.S., and W.T. Frankenberger Jr. 1999. Microbial transformations of selenium and the bioremediation of seleniferous environments. Biorem. J. 3:171–188.

Dungan, R.S., S.R. Yates, and W.T. Frankenberger Jr. 2003. Transformations of selenate and selenite by *Stenotrophomonas maltophilia* isolated from a seleniferous agricultural drainage pond sediment. Environ. Microbiol. 5:287–295.

Dynes, J.J., and P.M. Huang. 1997. Influence of organic acids on selenite sorption by poorly ordered aluminum hydroxides. Soil Sci. Soc. Am. J. 61:772–783.

Dzombak, D.A., and F.M.M. Morel. 1990. Surface complexation modeling: Hydrous ferric oxide. John Wiley & Sons, New York.

Elzinga, E.J., and D.L. Sparks. 2001. Reaction condition effects on nickel sorption mechanisms in illite-water suspensions. Soil Sci. Soc. Am. J. 65:94–101.

Evers, A., R.D. Hancock, A.E. Martell, and R.J. Motekaitis. 1989. Metal ion recognition in ligands with negatively charged oxygen donor groups. Complexation of FeIII, GaIII, InIII, and other highly charged metal ions. Inorg. Chem. 28:2189–2195.

Farkas, E., E.A. Enyedy, and H. Csoka. 1999. A comparison between the chelating properties of some dihydroxamic acids, desferrioxamine B and acetohydraxamic acid. Polyhedron 18:2391–2398.

Farley, K.J., D.A. Dzombak, and F.M.M. Morel. 1985. A surface precipitation model for the sorption of cations on metal oxides. J. Colloid Interface Sci. 106:226–242.

Feder, F., F. Trolard, G. Klingelhofer, and G. Bourrie. 2005. In situ Mossbauer spectroscopy: Evidence for green rust fougerite in a gleysol and its mineralogical transformations with time and depth. Geochim. Cosmochim. Acta 69:4463–4483.

Fernandez-Martinez, A., and L. Charlet. 2009. Selenium environmental cycling and bioavailability: A structural chemist point of view. Rev. Environ. Sci. Biotechnol. 8:81–110.

Ford, R.G. 2002. Rates of hydrous ferric oxide crystallization and the influence on coprecipitated arsenate. Environ. Sci. Technol. 36:2459–2463.

Ford, R.G., P.M. Bertsch, and K.J. Farley. 1997. Changes in transition and heavy metal partitioning during hydrous iron oxide aging. Environ. Sci. Technol. 31:2028–2033.

Ford, R.G., K.M. Kemner, and P.M. Bertsch. 1999a. Influence of sorbate-sorbent interactions on the crystallization kinetics of nickel- and lead-ferrihydrite coprecipitates. Geochim. Cosmochim. Acta 63:39–48.

Ford, R.G., A.C. Scheinost, K.G. Scheckel, and D.L. Sparks. 1999b. The link between clay mineral weathering and the stabilization of Ni surface precipitates. Environ. Sci. Technol. 33:3140–3144.

Frankenberger, W.T. Jr., and U. Karlson. 1989. Environmental factors affecting microbial production of dimethylselenide in a selenium-contaminated sediment. Soil Sci. Soc. Am. J. 53:1435–1442.

Fröberg, M., D. Berggren, B. Bergkvist, C. Bryant, and J. Mulder. 2006. Concentration and fluxes of dissolved organic carbon DOC in three Norway spruce stands along a climatic gradient in Sweden. Biogeochemistry 77:1–23.

Frost, R.R., and R.A. Griffin. 1977. Effect of pH on adsorption of arsenic and selenium from landfill leachate by clay minerals. Soil Sci. Soc. Am. J. 41:53–56.

Gaskova, O.L., M.B. Bukaty, G.P. Shironosova, and V.G. Kabannik. 2009. Thermodynamic model for sorption of bivalent heavy metals on calcite in natural-technogenic environments. Russ. Geol. Geophys. 50:87–95.

Gilis, A., P. Corbisier, W. Baeyens, S. Taghavi, M. Mergeay, and D. van der Lelie. 1998. Effect of the siderophore alcaligin E on the bioavailability of Cd to *Alcaligenes eutrophus* CH34. J. Ind. Microbiol. Biotechnol. 20:61–68.

Goldberg, S., and L.J. Criscenti. 2008. Modeling adsorption of metals and metalloids by soil components, p. 215–264. *In* A. Violante, P.M. Huang, and G.M. Gadd (eds.) Biophysicochemical processes of heavy metals and metalloids in soil environments. Wiley-Interscience, Hoboken, NJ.

Goldberg, S., L.J. Criscenti, D.R. Turner, J.A. Davis, and K.J. Cantrell. 2007. Adsorption–desorption processes in subsurface reactive transport modeling. Vadose Zone J. 6:407–435.

Goldberg, S., and R.A. Glaubig. 1988. Anion sorption on a calcareous, montmorillonitic soil-selenium. Soil Sci. Soc. Am. J. 52:954–958.

Gustafsson, J.P. 2001. Modeling the acid–base properties and metal complexation of humic substances with the Stockholm Humic Model. J. Colloid Interface Sci. 244:102–112.

Hansch, R., and R.R. Mendel. 2009. Physiological functions of mineral micronutrients Cu, Zn, Mn, Fe, Ni, Mo, B, Cl. Curr. Opin. Plant Biol. 12:259–266.

Hansen, H.C.B., O.K. Borggaard, and J. Sorensen. 1994. Evaluation of the free energy of formation of FeII-FeIII hydroxidesulphate green rust and its reduction of nitrite. Geochim. Cosmochim. Acta 58:2599–2608.

Hayes, K.F., A.L. Roe, G.E. Brown Jr., K.O. Hodgson, J.O. Leckie, and G.A. Parks. 1987. In situ X-ray absorption study of surface complexes: Selenium oxyanions on a-FeOOH. Science 238:783–786.

Herbel, M.J., J.S. Blum, R.S. Oremland, and S.E. Borglin. 2003. Reduction of elemental selenium to selenide: Experiments with anoxic sediments and bacteria that respire Se-oxyanions. Geomicrobiol. J. 20:587–602.

Hinsinger, P., A.G. Bengough, D. Vetterlein, and I.M. Young. 2009. Rhizosphere: Biophysics, biogeochemistry and ecological relevance. Plant Soil 321:117–152.

Hockin, S.L., and G.M. Gadd. 2003. Linked redox precipitation of sulfur and selenium under anaerobic conditions by sulfate-reducing bacterial biofilms. Appl. Environ. Microb. 69:7063–7072.

Hoffmann, U., and S.L.S. Stipp. 2001. The behavior of Ni^{2+} on calcite surfaces. Geochim. Cosmochim. Acta 65:4131–4139.

Huerta-Diaz, M.A., and J.W. Morse. 1992. Pyritization of trace metals in anoxic marine sediments. Geochim. Cosmochim. Acta 56:2681–2702.

Huggett, R.J. 1998. Soil chronosequences, soil development, and soil evolution: A critical review. Catena 32:155–172.

Hummel, W., and E. Curti. 2003. Nickel aqueous speciation and solubility at ambient conditions: A thermodynamic elegy. Monatsh. Chem. Chem. Mon. 134:941–973.

Israelachvili, J.N. 1994. Intermolecular and surface forces. Academic Press Inc., San Diego, CA.

IUPAC. 1997. Compendium of chemical terminology, 2nd Ed. The "gold book". Compiled by A.D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford, UK. XML online corrected version: http://goldbook.iupac.org 2006—created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. ISBN 0-9678550-9-8.

Jacquat, O., A. Voegelin, A. Villard, M.A. Marcus, R. Kretzschmar. 2008. Formation of Zn-rich phyllosilicate, Zn-layered double hydroxide and hydrozincite in contaminated calcareous soils. Geochim. Cosmochim. Acta 72:5037–5054.

Johnson, T.M., and T.D. Bullen. 2003. Selenium isotope fractionation during reduction by FeII-FeIII hydroxide-sulfate green rust. Geochim. Cosmochim. Acta 67:413–419.

Jones, D.L. 1998. Organic acids in the rhizosphere—A critical review. Plant Soil 205:25–44.

Karickhoff, S.W. 1981. Semi-empirical estimation of sorption of hydrophobic pollutants on natural sediments and soils. Chemosphere 108:833–846.

Katz, L.E., and K.F. Hayes. 1995. Surface complexation modeling. 2. Strategy for modeling polymer and precipitation reactions at high surface coverage. J. Colloid Interface Sci. 170:491–501.

Keiluweit, M., and M. Kleber. 2009. Molecular-level interactions in soils and sediments: The role of aromatic π-systems. Environ. Sci. Technol. 43:3421–3429.

Kelleher, B.P., and A.J. Simpson. 2006. Humic substances in soils: Are they really chemically distinct? Environ. Sci. Technol. 40:4605–4611.

Khan, M.R., S.M. Khan, F.A. Mohiddin, and T.H. Askary. 2006. Effects of high nickel soil on root-knot Nematode disease of tomato. Nematropica 36:79–87.

Kjoller, C., D. Postma, and F. Larsen. 2004. Groundwater acidification and the mobilization of trace metals in a sandy aquifer. Environ. Sci. Technol. 38:2829–2835.

Kowacz, M., M. Prieto, and A. Putnis. 2010. Kinetics of crystal nucleation in ionic solutions: Electrostatics and hydration forces. Geochim. Cosmochim. Acta 74:469–481.

Kraemer, S.M., S.-F. Cheah, R. Zapf, J. Xu, K.N. Raymond, and G. Sposito. 1999. Effect of hydroxamate siderophores on Fe release and PbII adsorption by goethite. Geochim. Cosmochim. Acta 63:3003–3008.

Kraemer, S.M., D.E. Crowley, and R. Kretzschmar. 2006. Geochemical aspects of phytosiderophore-promoted iron acquisition by plants. Adv. Agron. 91:1–46.

Kulik, D.A. 2002a. Sorption modelling by Gibbs energy minimisation: Towards a uniform thermodynamic database for surface complexes of radionuclides. Radiochim. Acta 90:815–832.

Kulik, D.A. 2002b. Gibbs energy minimization approach to modeling sorption equilibria at the mineral–water interface: Thermodynamic relations for multi-site-surface complexation. Am. J. Sci. 302:227–279.

Kulik, D.A. 2006. Dual-thermodynamic estimation of stoichiometry and stability of solid solution end members in aqueous–solid solution systems. Chem. Geol. 225:189–212.

Lambert, J.-L. 2008. Adsorption and polymerization of amino acids on mineral surfaces: A review. Orig. Life Evol. Biosph. 38:211–242.

Langmuir, D. 1997. Aqueous environmental geochemistry. Prentice-Hall, Upper Saddle River, NJ.

Lasaga, A.C. 1999. Kinetic theory in earth sciences. Princeton University Press, Princeton, NJ.

Lilienfein, J., R.G. Qualls, S.M. Uselman, and S.D. Bridgham. 2003. Soil formation and organic matter accretion in a young andesitic chronosequence at Mt. Shasta, California. Geoderma 116:249–264.

Lindsay, W.L. 1979. Chemical equilibria in soils. John Wiley & Sons, New York.

Lloyd, J.R., D.R. Lovley, and L.E. Macaskie. 2003. Biotechnological application of metal-reducing microorganisms. Adv. Appl. Microbiol. 53:85–128.

Lorenz, K., R. Lal, C.M. Preston, and K.G.J. Nierop. 2007. Strengthening the soil organic carbon pool by increasing contributions from recalcitrant aliphatic biomacromolecules. Geoderma 142:1–10.

Lortie, L., W.D. Gould, S. Rajan, R.G.L. McCready, and K.J. Cheng. 1992. Reduction of selenate and selenite to elemental selenium by a *Pseudomonas stutzeri* isolate. Appl. Environ. Microbiol. 58:4042–4044.

Losi, M.E., and W.T. Frankenberger. 1998. Microbial oxidation and solubilization of precipitated elemental selenium in soil. J. Environ. Qual. 27:836–843.

Lowers, H.A., G.N. Breit, A.L. Foster, J. Whitney, J. Yount, Md. N. Uddin, and Ad. A. Muneem. 2007. Arsenic incorporation into authigenic pyrite, Bengal Basin sediment, Bangladesh. Geochim. Cosmochim. Acta 71:2699–2717.

Lützenkirchen, J., and P.H. Behra. 1995. On the surface precipitation model for cation sorption at the hydroxide water interface. Aquat. Geochem. 1:375–397.

Maher, M.J., J. Santini, I.J. Pickering, R.C. Prince, J.M. Macy, and G.N. George. 2004. X-ray absorption spectroscopy of selenate reductase. Inorg. Chem. 43:402–404.

Manceau, G., and G. Calas. 1986. Nickel-bearing clay minerals: II. Intracrystalline distribution of nickel: An X-ray absorption study. Clay Miner. 21:341–360.

Manceau, G., G. Calas, and A. Decarreau. 1985. Nickel-bearing clay minerals: I. Optical spectroscopic study of nickel crystal chemistry. Clay Miner. 20:367–387.

Manceau, A., and L. Charlet. 1994. The mechanism of selenate adsorption on goethite and hydrous ferric oxide. J. Colloid Interface Sci. 168:87–93.

Manceau, A., M. Lanson, and N. Geoffroy. 2007. Natural speciation of Ni, Zn, Ba, and As in ferromanganese coatings on quartz using X-ray fluorescence, absorption, and diffraction. Geochim. Cosmochim. Acta 71:95–128.

Manceau, A., B. Lanson, M.L. Schlegel, J.C. Hargé, M. Musso, L. Eybert-Bérard, J.-L. Hazemann, D. Chateigner, and G.M. Lamble. 2000. Quantitative Zn speciation in smelter-contaminated soils by EXAFS spectroscopy. Am. J. Sci. 300:289–343.

Manceau, A., N. Tamura, M.A. Marcus, A.A. MacDowell, R.S. Celestre, R.E. Sublett, G. Sposito, and H.A. Padmore. 2002. Deciphering Ni sequestration in soil ferromanganese nodules by combining X-ray fluorescence, absorption, and diffraction at micrometer scales of resolution. Am. Mineral. 87:1494–1499.

Martin-Garin, A., P. Van Cappellen, and L. Charlet. 2003. Aqueous cadmium uptake by calcite: A stirred flow-through reactor study. Geochim. Cosmochim. Acta 67:2763–2774.

Mattigod, S.V., D. Rai, A.R. Felmy, and L. Rao. 1997. Solubility and solubility product of crystalline $NiOH_2$. J. Solution Chem. 26:391–403.

Matynia, A., T. Lenoir, B. Causse, L. Spadini, T. Jacquet, and A. Manceau. 2010. Semi-empirical proton binding constants for natural organic matter. Geochim. Cosmochim. Acta 74:1836–1851.

McBride, M.B. 1994. Environmental chemistry of soils. Oxford University Press, New York.

McBride, M.B. 1999. Chemisorption and precipitation reactions, p. B265–B302. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Merlen, E., P. Guéroult, J.-B. d'Espinose de la Caillerie, B. Rebours, C. Bobin, and O. Clause. 1995. Hydrotalcite formation at the alumina/water interface during impregnation with Ni II aqueous solutions at neutral pH. Appl. Clay Sci. 10:45–56.

MINTEQA2/PRODEFA2. 1999. A geochemical assessment model for environmental systems: User manual supplement for version 4.0. USEPA, Washington, DC. Available at http://www.epa.gov/ceampubl/mmedia/minteq/SUPPLE1.PDF (accessed on March 17, 2010)

Myneni, S.C.B., T.K. Tokunaga, and G.E. Brown Jr. 1997. Abiotic selenium redox transformations in the presence of Fe(II,III) oxides. Science 278:1106–1109.

Nachtegaal, M., and D.L. Sparks. 2003. Nickel sequestration in a kaolinite-humic acid complex. Environ. Sci. Technol. 37:529–534.

Neal, R.H., and G. Sposito. 1989. Selenate adsorption on alluvial soils. Soil Sci. Soc. Am. J. 53:70–74.

Neal, R.H., G. Sposito, K.M. Holtzclaw, and S.J. Traina. 1987. Selenite adsorption on alluvial soils: I. Soil composition and pH effects. Soil Sci. Soc. Am. J. 51:1161–1165.

Neubauer, U., B. Nowack, G. Furrer, and R. Schulin. 2000. Heavy metal sorption on clay minerals affected by the siderophore Desferrioxamine B. Environ. Sci. Technol. 34:2749–2755.

Nico, P.S., B.D. Stewart, and S. Fendorf. 2009. Incorporation of oxidized uranium into Fe hydroxides during FeII catalyzed remineralization. Environ. Sci. Technol. 43:7391–7396.

Niederer, C., R.P. Schwarzenbach, and K.-U. Goss. 2007. Elucidating differences in the sorption properties of 10 humic and fulvic acids for polar and nonpolar organic chemicals. Environ. Sci. Technol. 41:6711–6717.

Nowack, B. 2002. Environmental chemistry of aminopolycarboxylate chelating agents. Environ. Sci. Technol. 36:4009–4016.

Nowack, B., and L.J. Sigg. 1996. Adsorption of EDTA and metal-EDTA complexes onto goethite. J. Colloid Interface Sci. 177:106–121.

Nowack, B., H. Xue, and L. Sigg. 1997. Influence of natural and anthropogenic ligands on metal transport during infiltration of river water to groundwater. Environ. Sci. Technol. 31:866–872.

Oremland, R.S. 1994. Biogeochemical transformations of selenium in anoxic environments, p. 389–419. *In* W.T. Frankenberger Jr. and S. Benson (eds.) Selenium in the environment. Marcel Dekker, New York.

Oremland, R.S., M.J. Herbel, J.S. Blum, S. Langley, T.J. Beveridge, P.M. Ajayan et al. 2004. Structural and spectral features of selenium nanospheres produced by Se-respiring bacteria. Appl. Environ. Microbiol. 70:52–60.

Oremland, R.S., J.T. Hollibaugh, A.S. Maest, T.S. Presser, L.G. Miller, and W.C. Culbertson. 1989. Selenate reduction to elemental selenium by anaerobic bacteria in sediments and culture: Biogeochemical significance of a novel, sulfate-independent respiration. Appl. Environ. Microbiol. 55:2333–2343.

Oremland, R.S., N.A. Steinberg, A.S. Maest, L.G. Miller, and J.T. Hollibaugh. 1990. Measurement of in situ rates of selenate removal by dissimilatory bacteria reduction in sediments. Environ. Sci. Technol. 24:1157–1163.

Pareek, N., K.S. Dhillon, and S.K. Dhillon. 2000. Effect of sulphate, nitrate and phosphate ions on adsorption of selenium in seleniferous soils of Punjab. J. Nucl. Agric. Biol. 29:167–174.

Peltier, E., R. Allada, A. Navrotsky, and D.L. Sparks. 2006. Nickel solubility and precipitation in soils: A thermodynamic study. Clay. Clay Miner. 54:153–164.

Pickering, I.J., G.E. Brown Jr., and T.K. Tokunaga. 1995. Quantitative speciation of selenium in soils using X-ray absorption spectroscopy. Environ. Sci. Technol. 29:2456–2459.

Ponge, J.-F. 2003. Humus forms in terrestrial ecosystems: A framework to biodiversity. Soil Biol. Biochem. 35:935–945.

Rai, D., A.R. Felmy, and D.A. Moore. 1995. The solubility product of crystalline ferric selenite hexahydrate and the complexation constant of $FeSeO_3^+$. J. Solution Chem. 24:735–752.

Ravat, C., J. Dumonceau, and F. Monteil-Rivera. 2000. Acid/base and CuII binding properties of natural organic matter extracted from wheat bran: Modeling by the surface complexation model. Water Res. 34:1327–1339.

Reeder, R.J., G.M. Lamble, J.F. Lee, and W.J. Staudt. 1994. Mechanism of SeO_4^{2-} substitution in calcite: An XAFS study. Geochim. Cosmochim. Acta 58:5639–5646.

Refait, P.H., L. Simon, C. Louis, and J.M.R. Genin. 2000. Reduction of SeO_4^{2-} anions and anoxic formation of ironII–ironIII hydroxyselenate green rust. Environ. Sci. Technol. 34:819–825.

Rege, M.A., D.R. Yonge, D.P. Mendoza, J.N. Petersen, Y. Bereded-Samuel, D.L. Johnstone, W.A. Apel, and J.M. Barnes. 1999. Selenium reduction by a denitrifying consortium. Biotechnol. Bioeng. 62:479–484.

Román-Ross, G., G.J. Cuello, X. Turrillas, A. Fernández-Martínez, and L. Charlet. 2006. Arsenite sorption and co-precipitation with calcite. Chem. Geol. 233:328–336.

Rotenberg, B., J.-P. Morel, V. Marry, P. Turq, and N. Morel-Desrosiers. 2009. On the driving force of cation exchange in clays: Insights from combined microcalorimetry experiments and molecular simulation. Geochim. Cosmochim. Acta 73:4034–4044.

Salminen, J.M., P.J. Hanninen, J. Leveinen, P.T.J. Lintinen, and K.S. Jørgensen. 2006. Occurrence and rates of terminal electron-accepting processes and recharge processes in petroleum hydrocarbon-contaminated subsurface. J. Environ. Qual. 35:2273–2282.

Schaumann, G.E. 2006. Soil organic matter beyond molecular structure Part I: Macromolecular and supramolecular characteristics. J. Plant Nutr. Soil Sci. 169:145–156.

Scheckel, K.G., A.C. Scheinost, R.G. Ford, and D.L. Sparks. 2000. Stability of layered Ni hydroxide surface precipitates—A dissolution kinetics study. Geochim. Cosmochim. Acta 64:2727–2735.

Scheckel, K.G., and D.L. Sparks. 2000. Kinetics of the formation and dissolution of Ni precipitates in a gibbsite/amorphous silica mixture. J. Colloid Interface Sci. 229:222–229.

Scheckel, K.G., and D.L. Sparks. 2001. Temperature effects on nickel sorption kinetics at the mineral–water interface. Soil Sci. Soc. Am. J. 65:719–728.

Scheidegger, A.M., and D.L. Sparks. 1996. A critical assessment of sorption-desorption mechanisms at the soil mineral/water interface. Soil Sci. 161:813–831.

Scheidegger, A.M., D.G. Strawn, G.M. Lamble, and D.L. Sparks. 1998. The kinetics of mixed Ni-Al hydroxide formation on clays and aluminum oxides: A time-resolved XAFS study. Geochim. Cosmochim. Acta 62:2233–2245.

Scheinost, A.C., and L. Charlet. 2008a. Selenite reduction by mackinawite, magnetite and siderite: XAS characterization of nanosized redox products. Environ. Sci. Technol. 42:1984–1989.

Scheinost, A.C., R.G. Ford, and D.L. Sparks. 1999. The role of Al in the formation of secondary Ni precipitates on pyrophyllite, gibbsite, talc, and amorphous silica: A DRS study. Geochim. Cosmochim. Acta 63:3193–3203.

Scheinost, A.C., R. Kirsch, D. Banerjee, A. Fernandez-Martinez, H. Zaenker, H. Funke, and L. Charlet. 2008b. X-ray absorption and photoelectron spectroscopy investigation of selenite reduction by Fe^{II}-bearing minerals. J. Contam. Hydrol. 102:228–245.

Scheinost, A.C., and D.L. Sparks. 2000. Formation of layered single and double metal hydroxide precipitates at the mineral/water interface: A multiple-scattering XAFS analysis. J. Colloid Interface Sci. 223:167–178.

Schindler, P.W., B. Furst, R. Dick, and P.U. Wolf. 1976. Ligand properties of surface silanol groups. I. Surface complex formation with Fe^{3+}, Cu^{2+}, Cd^{2+} and Pb^{2+}. J. Colloid Interface Sci. 55:469–475.

Schlegel, M.L., and A. Manceau. 2006. Evidence for the nucleation and epitaxial growth of Zn phyllosilicate on montmorillonite. Geochim. Cosmochim. Acta 70:901–917.

Schlegel, M.L., A. Manceau, L. Charlet, D. Chateigner, and J.-L. Hazemann. 2001. Sorption of metal ions on clay minerals. III. Nucleation and epitaxial growth of Zn phyllosilicate on the edges of hectorite. Geochim. Cosmochim. Acta 65:4155–4170.

Schultz, M.F., M.M. Benjamin, and J.F. Ferguson. 1987. Adsorption and desorption of metals on ferrihydrite: Reversibility of the reaction and sorption properties of the regenerated solid. Environ. Sci. Technol. 21:863–869.

Schulze, D.G. 2002. An introduction to soil mineralogy, p. 1–35. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Schwarzenbach, R.P., P.M. Gschwend, and D.M. Imboden. 1993. Environmental organic chemistry, 1st Ed. Wiley-Interscience, New York.

Seby, F., M. Potin-Gautier, E. Giffaut, G. Borge, and O.F.X. Donard. 2001. A critical review of thermodynamic data for selenium species at 25°C. Chem. Geol. 171:173–194.

Shao, H., S.V. Dmytrieva, O. Kolditz, D.A. Kulik. W. Pfingsten, and G. Kosakowski. 2009. Modeling reactive transport in non-ideal aqueous–solid solution system. Appl. Geochem. 24:1287–1300.

Sharmasarkar, S., K.J. Reddy, and G.F. Vance. 1996. Preliminary quantification of metal selenite solubility in aqueous solutions. Chem. Geol. 132:165–170.

Sigg, L., and W. Stumm. 1981. The interactions of anions and weak acids with the hydrous goethite α-FeOOH surface. Colloids Surf. 2:101–117.

Simpson, A.J., M.J. Simpson, E. Smith, and B.P. Kelleher. 2007a. Microbially derived inputs to soil organic matter: Are current estimates too low? Environ. Sci. Technol. 41:8070–8076.

Simpson, A.J., G. Song, E. Smith, B. Lam, E. Novotny, and M.H.B. Hayes. 2007b. Unraveling the structural components of soil humin by use of solution-state nuclear magnetic resonance spectroscopy. Environ. Sci. Technol. 41:876–883.

Smejkalova, D., and A. Piccolo. 2008. Aggregation and disaggregation of humic supramolecular assemblies by NMR diffusion ordered spectroscopy DOSY-NMR. Environ. Sci. Technol. 42:699–706.

Sparks, D.L. 2003. Environmental soil chemistry, 2nd Ed. Academic Press, New York.

Sposito, G. 1984. The surface chemistry of soils. Oxford University Press, New York.

Sposito, G. 1998. On points of zero charge. Environ. Sci. Technol. 32:2815–2819.

Sposito, G. 1999. On points of zero charge. Environ. Sci. Technol. 33:208–208.

Sposito, G. 2008. The chemistry of soils, 2nd Ed. Oxford University Press, New York.

Strathmann, T.J., and S.C.B. Myneni. 2004. Speciation of aqueous NiII-carboxylate and NiII-fulvic acid solutions: Combined ATR-FTIR and XAFS analysis. Geochim. Cosmochim. Acta 68:3441–3458.

Strawn, D., H. Doner, M. Zavarin, and S. McHugo. 2002. Microscale investigation into the geochemistry of arsenic, selenium, and iron in soil developed in pyritic shale materials. Geoderma 108:237–257.

Strobel, B.W. 2001. Influence of vegetation on low-molecular-weight carboxylic acids in soil solution—A review. Geoderma 99:169–198.

Stumm, W. 1992. Chemistry of the solid-water interface. John Wiley & Sons, New York.

Stumm, W., and J.J. Morgan. 1981. Aquatic chemistry: An introduction emphasizing chemical equilibria in natural waters, 2nd Ed. John Wiley & Sons, New York.

Stumm, W., and J.J. Morgan. 1996. Aquatic chemistry: Chemical equilibria and rates in natural waters, 3rd Ed. John Wiley & Sons, New York.

Su, C., and D.L. Suarez. 2000. Selenate and selenite sorption on iron oxides: An infrared and electrophoretic study. Soil Sci. Soc. Am. J. 64:101–111.

Sutton, R., and G. Sposito. 2005. Molecular structure in soil humic substances: The new view. Environ. Sci. Technol. 39:9009–9015.

Sverjensky, D.A. 2003. Standard states for the activities of mineral surface sites and species. Geochim. Cosmochim. Acta 67:17–28.

Sverjensky, D.A. 2006. Prediction of the speciation of alkaline earths adsorbed on mineral surfaces in salt solutions. Geochim. Cosmochim. Acta 70:2427–2453.

Tamura, H., and R. Furuichi. 1997. Adsorption affinity of divalent heavy metal ions for metal oxides evaluated by modeling with the Frumkin isotherm. J. Colloid Interface Sci. 195:241–249.

Teppen, B.J., and D.M. Miller. 2006. Hydration energy determines isovalent cation exchange selectivity by clay minerals. Soil Sci. Soc. Am. J. 70:31–40.

Thoenen, T. 1999. Pitfalls in the use of solubility limits for radioactive waste disposal: The case of nickel in sulfidic groundwaters. Nucl. Technol. 126:75–87.

Thompson, H.A., G.A. Parks, and G.E. Brown. 1999. Dynamic interactions of dissolution, surface adsorption, and precipitation in an aging cobaltII-clay-water system. Geochim. Cosmochim. Acta 63:1767–1779.

Tokunaga, T.K., I.J. Pickering, and G.E. Brown Jr. 1996. Selenium transformation in ponded sediments. Soil Sci. Soc. Am. J. 60:781–790.

Tokunaga, T.K., S.R. Sutton, and S. Bajt. 1994. Mapping of selenium concentrations in soil aggregates with synchrotron X-ray fluorescence microprobe. Soil Sci. 158:421–434.

Tomei, F.A., L.L. Barton, C.L. Lemanski, and T.G. Zocco. 1992. Reduction of selenate and selenite to elemental selenium by *Wolinella succinogenes*. Can. J. Microbiol. 38:1328–1333.

Tonkin, J.W., L.S. Balistrieri, and J.W. Murray. 2004. Modeling sorption of divalent metal cations on hydrous manganese oxide using the diffuse double layer model. Appl. Geochem. 19:29–53.

Trolard, F., G. Bourrie, M. Abdelmoula, P. Refait, and F. Feder. 2007. Fougerite, a new mineral of the pyroaurite-iowaite group: Description and crystal structure. Clay. Clay Miner. 55:323–334.

Trolard, F., J.M.R. Genin, M. Abdelmoula, G. Bourrie, B. Humbert, and A. Herbillon. 1997. Identification of a green rust mineral in a reductomorphic soil by Mossbauer and Raman spectroscopies. Geochim. Cosmochim. Acta 61:1107–1111.

Villegas-Jimenez, A., A. Mucci, O.S. Pokrovsky, and J. Schott. 2009. Defining reactive sites on hydrated mineral surfaces: Rhombohedral carbonate minerals. Geochim. Cosmochim. Acta 73:4326–4345.

Visual MINTEQ. 2009. Version 2.61. Stockholm, Sweden. Available online at http:/www.lwr.kth.se/English/OurSoftware/vminteq/ (accessed on March 17, 2010)

Wershaw, R.L. 1994. Membrane-micelle model for humus in soils and sediments and its relation to humification. Water-Supply Paper 2410. U.S. Geological Survey, Denver, CO.

Westall, J. 1986. Reactions at the oxide-solution interface: Chemical and electrostatic models, p. 54–78. *In* J.A. Davis and K.F. Hayes (eds.) Geochemical processes and mineral surfaces. ACS Symp. Ser. 323. American Chemical Society, Washington, DC.

Westall, J.C., J.D. Jones, G.D. Turner, and J.M. Zachara. 1995. Models for association of metal ions with heterogeneous environmental sorbents. 1. Complexation of CoII by Leonardite humic acid as a function of pH and NaClO$_4$ concentration. Environ. Sci. Technol. 29:951–959.

White, A.F., S.M. Benson, A.W. Yee, H.A. Wollenberg, and S. Flexer. 1991. Groundwater contamination at the Kesterson reservoir, California, 2. Geochemical parameters influencing selenium mobility. Water Resour. Res. 27:1085–1098.

Wiederhold, J.G., N. Teutsch, S.M. Kraemer, A.N. Halliday, and R. Kretzschmar. 2007. Iron isotope fractionation during pedogenesis in redoximorphic soils. Soil Sci. Soc. Am. J. 71:1840–1850.

Wijnja, H., and C.P. Schulthess. 2000. Vibrational spectroscopy study of selenate and sulfate adsorption mechanisms on Fe and Al hydroxide surfaces. J. Colloid Interface Sci. 229:286–297.

Wolthers, M., L. Charlet, P.R. van Der Linde, D. Rickard, and C.H. van Der Weijden. 2005. Surface chemistry of disordered mackinawite FeS. Geochim. Cosmochim. Acta 69:3469–3481.

Wood, B.W., R. Chaney, and M. Crawford. 2006. Correcting micronutrient deficiency using metal hyperaccumulators: Alyssum biomass as a natural product for nickel deficiency correction. Hortscience 41:1231–1234.

You, Y., G.F. Vance, and H. Zhao. 2001. Selenium adsorption on Mg-Al and Zn-Al layered double hydroxides. Appl. Clay Sci. 20:13–25.

Zachara, J.M., J.K. Fredrickson, S.C. Smith, and P.L. Gassman. 2001. Solubilization of FeIII oxide-bound trace metals by a dissimilatory FeIII reducing bacterium. Geochim. Cosmochim. Acta 65:75–93.

Zachara, J.M., D. Rai, D.A. Moore, G.D. Turner, and A.R. Felmy. 1994. Chemical attenuation reactions of selenium. TR-103535. Electric Power Research Institute, Palo Alto, CA.

Zachara, J.M., and J.C. Westall. 1999. Chemical modeling of ion adsorption in soils, p. 47–96. *In* D.L. Sparks (ed.) Soil physical chemistry, 2nd Ed. CRC Press, Boca Raton, FL.

Zawislanski, P.T., and M. Zavarin. 1996. Nature and rates of selenium transformations: A laboratory study of Kesterson Reservoir soils. Soil Sci. Soc. Am. J. 60:791–800.

Zech, W., N. Senesi, G. Guggenberger, K. Kaiser, J. Lehmann, T.M. Miano, A. Miltner, and G. Schroth. 1997. Factors controlling humification and mineralization of soil organic matter in the tropics. Geoderma 79:117–161.

Zhang, Y.Q., and W.T. Frankenburger Jr. 1999. Effects of soil moisture, depth, and organic amendments on selenium volatilization. J. Environ. Qual. 28:1321–1326.

Zhang, P.C., and D.L. Sparks. 1990. Kinetics of selenate and selenite adsorption/desorption at the goethite/water interface. Environ. Sci. Technol. 24:1848–1856.

Zhang, Y.Q., Z.A. Zahir, and W.T. Frankenburger Jr. 2003. Factors affecting reduction of selenate to elemental selenium in agricultural drainage water by *Enterobacter taylorae*. J. Agr. Food Chem. 51:7073–7078.

Zingaro, R.A., D.C. Dufner, A.P. Murphy, and C.D. Moody. 1997. Reduction of oxoselenium anions by ironII hydroxide. Environ. Int. 23:299–304.

Role of Abiotic Catalysis in the Transformation of Organics, Metals, Metalloids, and Other Inorganics

Pan Ming Huang
(Deceased)
University of Saskatchewan

A.G. Hardie
Stellenbosch University

18.1 Introduction

Abiotic catalysis plays a vital role in many physicochemical processes in soil and related environments. Clay minerals, metal (oxy)hydroxides and oxides, and dissolved metals often demonstrate their ability to catalyze the transformations of natural organic and anthropogenic organic compounds, metals, metalloids, and other inorganics.

Metal (oxy)hydroxides and oxides and clay minerals have the ability to catalyze the transformation of biomolecules and the resultant formation of humic substances (Huang and Hardie, 2009). Iron (oxy)hydroxides and oxides (Scheffer et al., 1959; Shindo and Huang, 1984a; Wang and Huang, 2000a, 2000b; Gonzalez and Laird, 2004) and especially Mn oxides (Shindo and Huang, 1982, 1984a, 1984b; Kung and McBride, 1988; Wang and Huang, 1992, 2000a,b; Jokic et al., 2001a, 2001b, 2004a, 2004b; Hardie et al., 2007) are most reactive in mediating the transformations of phenolic compounds, amino acids, and sugars. Other organic compounds such as aromatic amines and organic acids can also be oxidatively transformed by abiotic

catalysis (Furukawa and Brindley, 1973; McBride, 1979; Jauregui and Reisenauer, 1982; Stone and Morgan, 1984b).

Abiotic catalytic reactions are also important in the transformations of anthropogenic organic compounds such as pesticides, antibiotics, explosives, and dyes (Theng, 1974, 1979; Cheng, 1991; Stone and Torrents, 1995; Smolen and Stone, 1998; Wang and Arnold, 2003; Szecsody et al., 2004; Barrett and McBride, 2005; Baldrian et al., 2006; Hofstetter et al., 2006; Kang et al., 2006; Rubert and Pedersen, 2006; Fimmen et al., 2007; Cheng et al., 2008; Zhang et al., 2008). Soils and sediments contain a series of solid surfaces and dissolved constituents. These components can catalyze the transformation reactions through Brønsted and Lewis acidity, hydrolysis, and oxidative or reductive processes, which can result in the alteration of transformation pathways and kinetics, thereby influencing the toxicity and environmental fate of these organic compounds.

Manganese oxides and Fe-bearing minerals have the ability to catalyze the transformations of metals, metalloids, and/or other inorganics. Manganese oxides are effective catalysts in promoting many reactions such as the transformation of As(III)

to As(V) (Oscarson et al., 1981a; Chiu and Hering, 2000; Power et al., 2005; Feng et al., 2006b), Fe(II) to Fe(III) (Krishnamurti and Huang, 1987, 1988), Pu(III/IV) (Cleveland, 1970; Amacher and Baker, 1982; Morgenstern and Choppin, 2002) and Pu(V) (Keeneykennicutt and Morse, 1985; Duff et al., 1999) to Pu(VI), as well as Cr(III) to the more toxic and mobile Cr(VI) (Bartlett and James, 1979; Amacher and Baker, 1982; Stepniewska et al., 2004; Negra et al., 2005; Feng et al., 2007). Manganese(IV) oxides also catalyze the oxidation of nitrite to nitrate (Bartlett, 1981; Luther and Popp, 2002), and NH_3 and organic N to N_2 (Luther et al., 1997). Heterogeneous oxidation/reduction reactions involving electron transfer between metals/metalloids and Fe-bearing minerals have been demonstrated (Wehrli and Stumm, 1989; Ilton and Veblen, 1994; Peterson et al., 1996; White and Peterson, 1996; Myneni et al., 1997; Powell et al., 2004; Jeon et al., 2005; Jung et al., 2007; Jang and Dempsey, 2008; Scheinost and Charlet, 2008; Su and Puls, 2008). Further, it has been suggested that reduced Fe and Mn minerals are responsible for catalyzing nitrate immobilization into dissolved and humic N fractions in soils and sediments (Davidson et al., 2003; Huygens et al., 2008).

Therefore, abiotic catalysis plays an important role in the transformation of organics, metals, metalloids, and other inorganics in soil and related environments. The impact of abiotic catalysis on environmental quality and ecosystem integrity deserves increasing attention.

18.2 Fundamentals of Catalysis

18.2.1 Definition of Catalysis

The process of changing the rate of a chemical reaction by the use of a catalyst is termed catalysis, coined by Berzelius in 1836 to describe some enhanced chemical reactions (Williams, 1965). The accepted definition of a catalyst, due to Oswald, is that it is a substance that changes the speed of a chemical reaction without itself undergoing any permanent chemical change. Since a reactant or a product may also be a catalyst, Bell (1941) suggests the definition, "A catalyst is a substance which appears in the rate expression to a power higher than that to which it appears in the stoichiometric equation." Actually many substances classified as catalysts are destroyed either as a result of the process that gives them their catalytic activity or because of subsequent combination with the products (Moore and Pearson, 1981). From a practical point of view, a catalyst is a substance that changes the rate of a desired reaction, regardless of the fate of the catalyst itself.

An important criterion of a catalyst is that it changes the mechanism of the parent reaction (Moore and Pearson, 1981). Without this change in mechanism, the observed change in rate could not occur. Since catalysts increase the rate of reaction, the mechanism must change to one that is easier for the system to follow, involving, in general, a lower energy barrier. Therefore, the catalyst provides an alternative pathway by which the reaction comes to equilibrium, although it does not alter the position of the equilibrium (Daintith, 1990). The catalyst itself takes part

in the reaction. In certain circumstances, very small quantities of catalyst can speed up very large reactions. Some catalysts are also highly specific in the type of reaction they catalyze, particularly in biochemical reactions.

18.2.2 Homogeneous and Heterogeneous Catalysis

The process of changing the rate of a chemical reaction by the use of a catalyst that has the same phase as the reactant is homogeneous catalysis (e.g., dissolved metals in catalyzing organic reactions or enzymes in biochemical reactions) (Daintith, 1990). A process driven by a catalyst that has a phase different from the reactant is heterogeneous catalysis (e.g., metal oxides in catalyzing organic and inorganic reactions). In heterogeneous catalysis, the catalyzed reaction steps take place at or very close to the solid surface. These steps may be between molecules adsorbed on the catalyst surface or may involve the topmost atomic layers of the catalysts (Twigg, 1989). The sequence of stages for a heterogeneous catalytic reaction is shown in Table 18.1. Any of these stages, if slow, may limit the overall rate of a catalytic reaction. Distinctions are often drawn between catalysts that are film-diffusion controlled (i.e., limited by stages 1 and/or 7), pore-diffusion controlled (i.e., limited by stages 2 and/or 6), and reaction controlled (i.e., limited by stages 3, 4, and/or 5). There is a complex interaction between the relative importance of these different stages and the resulting catalytic effect on organic and inorganic reactions.

Advances in surface science and catalysis are presented in a treatise edited by Hightower et al. (1996). Both homogeneous and heterogeneous catalytic reactions are significant in soil and environmental sciences (Bartlett, 1986; Huang, 1990, 1991b; Stumm, 1992; Stone and Torrents, 1995; Smolen and Stone, 1998; Davidson et al., 2003; Livens et al., 2004; Stepniewska et al., 2004; Barrett and McBride, 2005; Hofstetter et al., 2006; Cheng et al., 2008; Scheinost and Charlet, 2008; Zhang et al., 2008).

TABLE 18.1 Sequence of Stages in the Catalysis of a Reaction by a Heterogeneous Catalyst[a]

1. Transport of reactants through the liquid or gas phase to the exterior of the catalyst
2. Transport of reactants through the pore system of the catalyst to a catalytically active site
3. Adsorption of reactants at the catalytically active site
4. Chemical reactions between reactants at the catalytically active site (frequently several steps)
5. Desorption of products from the catalytically active site
6. Transport of products through the catalyst pore system from the catalytically active site to the exterior of the catalyst
7. Transport of products into the liquid or gas phase from the exterior of the catalyst

Source: Modified from Twigg, M.V. 1989. Catalyst handbook. Wolfe Publishing Ltd., London, U.K. Copyright Wolfe Publishing Ltd., with permission.

[a] Several different catalytically active sites may be involved. Adsorption, possibly followed by reaction, may occur at one site, followed by transport of an intermediate product to a different site for further reactions.

18.2.3 Proton and Electron Transfer Catalysis

Brønsted acid–base catalysis is effective because proton transfers are generally rapid compared with the making and breaking of other chemical bonds. Therefore, reactions involving proton transfer in a typical acid or base catalysis are rapid compared with similar reactions of comparable free energy. The rate of a chemical reaction is related to the steric hindrance involved. Steric hindrance, which is the repulsion of nonbonded atoms in an activated complex, is the most important factor in determining the activation energy of a reaction (Moore and Pearson, 1981). Since a proton lacks the filled inner electron shells usually responsible for repulsion and is not surrounded by other groups, it is quite free from steric hindrance effects. Proton transfers involving oxygen–hydrogen bonds are generally rapid but they are not instantaneous. For instance, in the ionization of water, the activation energy is at least 57 kJ mol^{-1} (the heat of the reverse reaction) and the entropy of activation is negative; therefore, the rate constant (5×10^{-7} mol L^{-1} s^{-1}) is small (Moore and Pearson, 1981). This example demonstrates that an unfavorable equilibrium constant must necessarily make a reaction slow, even if other factors are quite favorable. Given a favorable equilibrium, proton transfers involving O and N bonds to H are almost always very fast, approaching diffusion control in many cases. Exceptions can occur if the proton is in a well-shielded position (Kresge, 1975).

Catalysis by proton transfer is by far the most common in homogeneous reactions. For those reactions that are subject to proton transfer catalysis, an expected relationship exists between the strength of the acid or base, as determined by its ionization constant, and its efficiency as a catalyst, determined by the observed rate constant. This relationship is best shown by the Brønsted catalysis law (Brønsted, 1928):

$$k_a = C_A K_a^\alpha \quad k_b = C_B K_b^\beta, \tag{18.1}$$

where
  - k_a and k_b are the rate constants (also termed the catalytic constants) for acid and base catalytic reactions, respectively
  - K_a and K_b are the acid and base ionization constants
  - C_A, C_B, α, and β are constants characteristic of the reaction, the solvent, and the temperature

Normally, α and β are positive and have values between 0 and 1. In the Brønsted equation, a low value of α and β signifies a low sensitivity of the catalytic constant to the strength of the catalyzing acid or base, and vice versa. Proton transfer catalysis is of significance in soils and associated environments (Theng, 1974, 1979; Cheng, 1991; Nannipieri and Gianfreda, 1998) as discussed later.

In acid catalysis of reactions involving negatively charged organic molecules, adsorption of a proton reduces their negative charge and, thus, facilitates the transfer of electrons (Steinberger and Westheimer, 1949, 1951). In agreement with this explanation, a number of multiply charged cations act as catalysts in the transformation of organic compounds (Stone and Torrents, 1995).

Presumably a metal-organic complex is formed that reduces the negative charge and increases the electron transfer. The catalytic efficiency of a metal ion depends both on its positive charge and on its ability to form a stable complex and chelate the reactant. In such reactions, the metal ions are acting as generalized acids. However, metal ions have some significant advantages over protons. They can have greater charges, which lead to greater polarization of the reactant molecules. Unlike the proton, metal ions can be stabilized by other ligands and, thus, can exist in neutral or even basic solutions. The high coordination numbers of metal ions permit the binding of a substrate at more than one site. This advantage helps to make metal ions very efficient catalysts for the hydrolysis of many organic compounds (Kroll, 1952). Further, some metal ions have the ability to simultaneously bind both a substrate and a reagent. This can have the effect of a template, in which the two reactants are assembled (brought into close proximity) prior to combination (Basolo and Pearson, 1968).

Many metal ions, especially of the transition series, have several stable oxidation states, which enables them to act as catalysts in certain redox reactions. Transition metals are the best catalysts, in most cases, for catalyzing reactions that are slow for symmetry reasons (Pearson, 1976). In addition to the slow three-body reactions, a second class of slow reactions is forbidden by orbital symmetry (Moore and Pearson, 1981). Even when the reaction is thermodynamically favorable, a large energy barrier can exist. Such reactions are prime candidates for catalysis. The reason why transition metals are often the best catalysts is due to their partly filled d orbitals, which have symmetry properties that are different from those of s and p orbitals. Because of the special properties of metal ions, particularly transition metal ions, they can catalyze a wide variety of organic and inorganic reactions in soil and related ecosystems.

18.3 Abiotic Catalysis of Natural and Anthropogenic Organic Compounds

18.3.1 Oxidative Transformation of Phenolic and Other Organic Compounds

18.3.1.1 Phenolics

Phenolics are the most widely distributed class of plant secondary metabolites and play an important role in regulating nutrient cycling in the terrestrial environment (Hättenschwiler and Vitousek, 2000). Phenolics may be released from natural sources, industrial activities, or agricultural practices; and they can harm the environment, especially in the form of pesticides (e.g., chlorophenols), nitrophenols, hormones, and azo dyes (Gianfreda et al., 2006). The oxidative transformation of phenolics can be accelerated enzymatically and nonenzymatically (Huang, 1990; Bollag et al., 1995; Naidja et al., 2000; Gianfreda et al., 2006). Soil minerals, in particular (SRO) Fe(III) and Mn(IV) oxides, play an important role in catalyzing the abiotic oxidative polymerization of phenolic compounds and the subsequent formation of humic

substances (Wang et al., 1986; Pal et al., 1994; Bollag et al., 1998; Huang, 2004; Huang et al., 2005). The surfaces of soil mineral colloids promote oxidative polymerization reactions by acting as electron acceptors. The rate-determining step in the formation of humic acids (HAs) from polyphenols is apparently the formation of a semiquinone-free radical involving a single electron transfer reaction (Schnitzer, 1982). Semiquinones couple with each other to form a stable HA polymer. The coupling of free radicals requires little activation energy, in contrast to electron transfer reactions (Chang and Allen, 1971). Therefore, coupling of semiquinones rather than the formation of quinones should be kinetically the preferred reaction pathway in the transformation of polyphenols to humic macromolecules.

Manganese(IV) oxides, such as birnessite (δ-MnO_2), cryptomelane (α-MnO_2), and pyrolusite (β-MnO_2), which commonly occur in soils and sediments (McKenzie, 1989), are powerful catalysts of the abiotic oxidation of polyphenols, such as catechol (1,2-dihydroxybenzene), compared with Fe, Al, and Si oxides (Shindo and Huang, 1982, 1984a; Shindo, 1992; Liu and Huang, 2001; Colarieti et al., 2006). This is partially attributable to the lower electronegativity of Mn (Liu and Huang, 2001). The electronegativity values of Mn, Fe, Al, Si, H, and O are, respectively, 1.55, 1.83, 1.61, 1.90, 2.20, and 3.44 (Porterfield, 1983). Catechol acts as a hard Lewis base, and Mn, Fe, Al, and Si are hard Lewis acids. When Mn, Fe, Al, or Si replace H in catechol to form metal–catechol complexes, the electron cloud delocalizes from the phenolic oxygen into the π-orbital formed from overlap of the 2p orbitals of the aromatic C atoms, thus accelerating the formation of semiquinone-free radicals and their polymerization. The electron cloud around the Mn–O bond in the Mn oxide–phenolic complex should be more delocalized than that around the Al–O bond in the Al–catechol complex and especially the Si–O or Fe–O bond in the Si or Fe oxide–phenolic complex due to the lower electronegativity of Mn than those of Al, Fe, and Si. This provides a partial explanation for the greater accelerating effect of Mn oxide on the humification of catechol than Fe, Al, or Si oxides. Naturally, redox reactions also play an important role in many abiotic catalytic reactions. Aluminum and silicon oxides are not subject to redox reaction. The standard electrode potential ($E°$) values of the overall redox reaction of the Fe(III) oxide–catechol and Mn(IV) oxide–catechol systems are +0.071 V and +0.509 V, respectively, as indicated by the following reactions (Shindo and Huang, 1984a):

$$Fe^{3+} + e^- = Fe^{2+} \quad E° = 0.770\,V, \tag{18.2}$$

$$MnO_2 + 4H^+ + 2e^- = Mn^{2+} + 2H_2O \quad E° = 1.208\,V, \tag{18.3}$$

$$C_6H_4(OH)_2 = C_6H_4O_2 + 2H^+ + 2e^- \quad E° = 0.6992V. \tag{18.4}$$
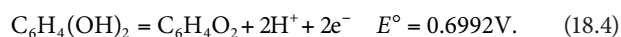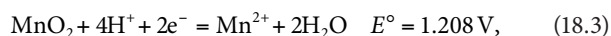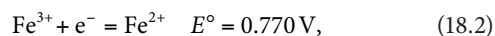
The positive $E°$ values of the overall redox reactions indicate that the reactions are thermodynamically feasible, and catechol oxidation can thus be accelerated by Fe oxide and especially Mn

oxide. This also explains the stronger catalytic ability of the Mn oxide than Fe, Al, and Si oxides in accelerating catechol oxidation. In addition, the lower point of zero salt effect (PZSE) and more negative charges of the Mn oxide than the Fe and Al oxides could also enhance the oxidation of catechol (Liu and Huang, 2001). A more negatively charged mineral surface may favor the binding of protons released from catechol and subsequently increase the catalytic reaction rate. Therefore, the catalytic ability of a metal oxide in polyphenol transformation depends on the $E°$ value of the overall redox reaction and the ability of the metal ions to complex with ligands, to shift electron density and molecular confrontation in the way conducive to the reaction, and to favor the binding of protons to the metal oxide.

Birnessite strongly promotes the formation of humic macromolecules from pyrogallol (1,2,3-trihydroxybenzene) under environmentally relevant conditions; total yields of humic substances were 10.5-fold higher than in pyrogallol reacted in the absence of birnessite (Wang and Huang, 1992). Manganese(IV) oxide has also been effectively used to treat polyphenol-polluted olive mill wastewater by enhancing humification (polymerization) (Brunetti et al., 2007). The Mn oxide-treated wastewater was found to enhance the humified C content of amended soils and increased the overall fertility of the soil (Brunetti et al., 2007).

During the abiotic catalytic transformation of pyrogallol to HA, Mn(IV), Fe(III), Al, and Si oxides also promote the abiotic generation of CO_2 through their ability to cleave the ring structure of pyrogallol (Wang and Huang, 1992, 2000b). The order of CO_2 release from the oxide-catalyzed pyrogallol reaction systems was Mn(IV) oxide >>> Fe(III) oxide > Al oxide > Si oxide (Wang and Huang, 2000b). The release of CO_2 was related to the development of carboxylic group contents in the HA fraction. Wang and Huang (2000a) showed that the infrared spectrum of the fulvic acid (FA) fraction from the Mn oxide–pyrogallol system closely resembles that of the FA fraction extracted from a natural Borosaprist soil. The abiotic ring cleavage of polyphenols by soil inorganic components may partially account for the findings of the high aliphaticity of natural humic substances (Wilson and Goh, 1977; Hatcher et al., 1981). Lee and Huang (1995) showed that the abiotic release of CO_2 in the birnessite–polyphenol and polyphenol systems increases with light intensity, a consequence of a photofragmentation of polyphenolics catalyzed by birnessite. These findings imply that the pathways of C turnover in the photic zones of soils and aquatic environments may differ from those in their subsurface and submerged layers.

Shindo and Huang (1992) compared the catalytic effects of Mn oxide and the oxidoreductase enzyme, tyrosinase, on the oxidative polymerization of diphenols over the pH range common in soil environments. Manganese oxide influences the oxidative polymerization of hydroquinone and resorcinol to a larger extent than does tyrosinase, whereas the reverse is true for catechol. The yields of HAs are significantly influenced by the kind of catalyst and polyphenol. In the Mn oxide system, the yield of HAs is in the order hydroquinone > catechol > resorcinol. In the tyrosinase system, catechol produces the highest yield of HA, followed by hydroquinone and resorcinol. These findings indicate that the

relative catalytic effects of Mn(IV) oxides and enzymes such as tyrosinase would vary with the type of polyphenols in soils. HAs formed by mineral catalysis have a better defined chemical structure than those formed by enzymatic oxidative polymerization, favoring the formation of components with lower degrees of aromatic ring condensation and lower molecular weights compared with those generated in the presence of tyrosinase (Naidja et al., 1998). Ahn et al. (2006) showed that the presence of the abiotic catalyst birnessite actually inhibits the catechol oxidative polymerization activity of the enzyme laccase, which was attributed to the Mn^{2+} released during the reduction of birnessite by catechol, which binds to the enzyme and alters its active site.

Phenolic acids have been shown to be oxidized rapidly in the presence of MnO_2 to form a number of soluble products (Lehmann and Cheng, 1988). Mass spectrometric data show that some of the soluble products of the reaction have somewhat higher molecular weights than the parent compounds. However, the soluble products of the reaction of ferulic acid and MnO_2 do not contain any ferulic acid hexamers (Liu et al., 1981; Bollag et al., 1982). The oxidized products of ferulic acids are apparently rapidly sorbed on the surfaces of MnO_2. Polyhydroxyphenolic acids with *p*- and *o*-OH groups are rapidly oxidized by Mn oxides (Pohlman and McColl, 1989) (Table 18.2) to polymeric humic products in both soil and Mn suspensions. On the other hand, *m*-polyhydroxyphenolic acids are not readily oxidized.

Monophenolic compounds, such as phenol (Jung et al., 2008b), and particularly those containing electron-donating substituent groups on the aromatic ring, can be oxidatively transformed by Mn(IV) oxides (Lehmann et al., 1987; Stone, 1987; Ulrich and Stone, 1989; Zhao et al., 2006). Manganese(IV) oxide has been found to be an effective abiotic catalyst of the oxidative transformation of highly toxic and recalcitrant compound, pentachlorophenol (PCP), under aquifer (Petrie et al., 2002) and near-dry

TABLE 18.2 Kinetic Constants for Polyhydroxyphenolic Acid Oxidation by Soil and Manganese Oxide Suspension[a]

Compound	Rate Constants (L mol^{-1} s^{-1})	
	Challenge A Horizon	MnO$_2$
2,3-Dihydroxybenzoic acid	ND[b]	0.03
2,5-Dihydroxybenzoic acid	0.06	0.04
2,6-Dihydroxybenzoic acid	0.00[c]	ND
3,4-Dihydroxybenzoic acid	0.10	0.04
3,5-Dihydroxybenzoic acid	0.00	0.00
Gallic acid	0.25	0.05
Syringic acid	ND	0.01
Vanillic acid	ND	0.00

Source: From Pohlman, A.A., and J.G. McColl. 1989. Organic oxidation and manganese and aluminum mobilization in forest soils. Soil Sci. Soc. Am. J. 53:686–690. With permission of the Soil Science Society of America.

[a] Oxidations were run using 10 g L^{-1} soil and 2.5 × 10^{-4} mol L^{-1} phenolic acid or 0.19 g L^{-1} MnO$_2$ and 5.0 × 10^{-4} mol L^{-1} phenolic acid at pH 4.5 and 30°C.

[b] ND, not determined.

[c] No oxidation of phenolic acid within 120 min of reaction.

conditions (Pizzigallo et al., 2004). Pyrolusite is able to oxidatively degrade the carcinogenic pesticide and antiseptic agent TCP (2,4,6-trichlorophenol) under near-dry conditions; however, the transformation products may actually be more toxic and persistent than the parent compound (Smith et al., 2006). Methoxylyated phenols have been used to mediate the oxidative polymerization of the nonphenolic fungicide, cyprodinil (4-cyclopropyl-6-methyl-*N*-phenyl-2-pyrimidinamine), in the presence of birnessite (Kang et al., 2004). Birnessite-induced oxidative coupling also results in the decarboxylation, demethoxylation, and dehalogenation of substituted phenolic subtrates (Dec et al., 2001, 2003). Electron-withdrawing substituents, such as –COOH and –Cl, are more susceptible to release than electron-donating ones, such as –OCH$_3$ and –CH$_3$ (Dec et al., 2003).

The catalytic ability of Fe oxides in the rapid oxidative polymerization of polyphenols (Scheffer et al., 1959) increases in the following order: ferrihydrite > goethite > maghemite > lepidocrocite > hematite. Ferrihydrite, which is SRO Fe oxide with high surface area, is most reactive in catalyzing the oxidative polymerization reaction. Besides the nature of Fe oxides, the catalytic ability of Fe is related to structure and functionality of phenolic compounds (Shindo and Huang, 1984a; Shindo, 1992; Pracht et al., 2001).

The oxidative polymerization of polyphenols is substantially influenced by the catalysis of Al hydroxides (Wang et al., 1983). Soluble silicic acid in aqueous solution and precipitated short-range order silica can catalyze the oxidation of polyphenols (Ziechmann, 1959). Liu and Huang (2000) reported that silicic acid and especially hydroxyl Al ions substantially enhance oxidative polymerization of catechol. Liu and Huang (2002) showed that hydroxy-aluminosilicate ions, which are precursors to noncrystalline aluminosilicates, are also effective in promoting the oxidative polymerization of catechol. The surface of ground quartz has a disturbed layer, which is SRO in nature (Iler, 1979). Similar disturbed surface layers are present on quartz grains in soils (Ribault, 1971). Oxidative polymerization of polyphenols may be catalyzed by the disturbed surface of quartz in soils.

Besides metal oxides, clay size layer silicates have the ability to catalyze the oxidative transformation of polyphenols. Before the pioneering work on the catalytic role of clay size layer silicates in oxidative polymerization of phenolic compounds and the subsequent formation of humic substances (Kumada and Kato, 1970; Filip et al., 1977; Wang and Li, 1977), the conversion of many aromatic amines into their color derivatives by clay minerals and clays had been investigated (Faust, 1940; Hauser and Legget, 1940). Solomon (1968) reported that, except for talc, a large number of representative minerals produce a blue color of varying intensity when brought into contact with a saturated solution of benzidine hydrochloride. The active sites for the oxidation of benzidine are located on the crystal edges and on transition metal atoms in the higher oxidation state that occupy octahedral sites in the silicate layers. Thompson and Moll (1973) measured the oxidative power of smectites by oxidation of hydroquinone to *p*-benzoquinone in a clay slurry. Oxidation occurs

in the presence of O_2 (air), but not N_2 unless Fe^{3+} or Cu^{2+} are the exchangeable cations. Adsorbed O_2 molecules or radicals on the clay surface are apparently partially responsible for the oxidation.

Pinnavais et al. (1974) reported aromatic radical cation formation on the intracrystal surfaces of transition metal-saturated layer silicates. Aromatic molecules, when reacted under very moderated conditions with Cu(II) or Fe(II) ions, may donate electrons to the metal cations, leading to the formation of polymers (Mortland and Halloran, 1976). Montmorillonite, vermiculite, illite, and kaolinite accelerate the formation of HAs to varying degrees (Shindo and Huang, 1985a). The promoting effect of 2:1 layer silicates is higher than that of 1:1 layer silicates because of the larger specific surface and lattice imperfections, which favor adsorption of O_2 molecules or radicals.

One of the well-identified precursors (Flaig et al., 1975; Hayes and Swift, 1978) for the formation of humic substances, hydroquinone, can be transformed in aqueous solution at near-neutral pH (6.5) to humic macromolecules and deposited in the interlayers of nontronite saturated with Ca, which is the most common and most abundant exchangeable cation in soils and sediments (Wang and Huang, 1986). Most of the interlayer humic macromolecules are highly resistant to alkali extraction and may, thus, be humin-type materials. Therefore, besides Al interlayers in 2:1 expansible layer silicates, the formation of humic substance interlayers in 2:1 expansible layer silicates, through polymerization of phenol monomers and the associated reactions in soils and sediments, deserves close attention.

The ability of nontronite to promote the oxidation of polyphenols is related to the structure and functionality of the polyphenols, and part of the reaction process may proceed as shown below (Wang and Huang, 1994):

$$(18.5)$$

$$(18.6)$$

$$(18.7)$$
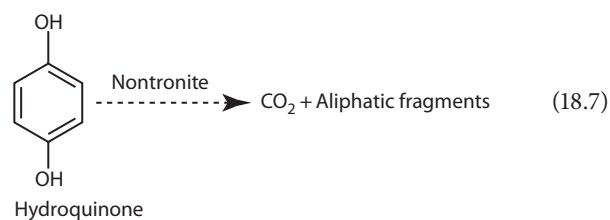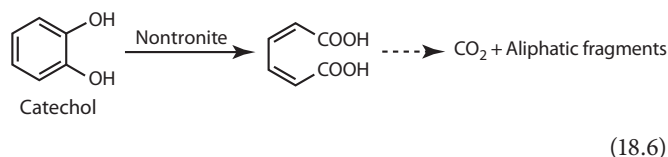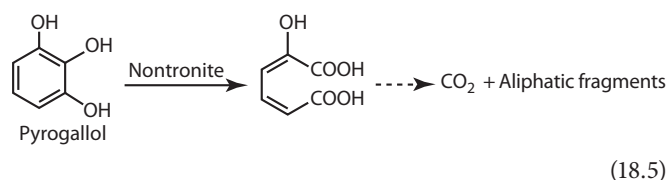
Catechol with two *o*-OHs is evidently more easily cleaved than hydroquinone with two *p*-OHs while pyrogallol with three hydroxyls in adjacent positions is even more easily cleaved than

catechol. The resultant carboxyl group-containing intermediates are further oxidized to form CO_2 and aliphatic fragments. In the reaction systems, intermediate products and aliphatic fragments may form polycondensates. The structure and functionality of polyphenols thus have an important role in influencing the extent of catalytic transformations by nontronite.

Primary minerals, which are commonly present in soil environments (Dixon and Weed, 1989), can catalyze the oxidative polymerization of polyphenols (Shindo and Huang, 1985b). The degree of acceleration of the oxidative polymerization of hydroquinone is greatest in the tephroite system, which increases the total HA yields more than ninefold because (1) tephroite (ideal chemical formula, $MnSiO_4$) is a Mn-bearing silicate, (2) part of the Mn in tephroite is present in the higher valence states, and (3) the oxidation of diphenols $[C_6H_4(OH)_2]$ by Mn(III) and Mn(IV) is thermodynamically favorable (Lide, 2008).

The hydroquinone-derived polymers with molecular weights of approximately 3500 and higher formed in the presence of the tephroite system (Shindo and Huang, 1985b) have similar IR absorption bands to those of humic substances (Schnitzer, 1978). The surface features of these polymers (Figure 18.1) are similar to those of soil HA and FA (Stevenson and Schnitzer, 1982) with the smallest discrete particles being spheroids with diameters of 0.1–0.2 µm (Figure 18.1a) and some aggregation of individual spheroids (Figure 18.1b and c). Small aggregates (Figure 18.1c) resemble moss while the large aggregates are nodule like (1–5 µm diameter) and doughnut like (6–8 µm diameter) (Figure 18.1a and b). The polymers do not appear to be associated with the surfaces of tephroite particles (Figure 18.1d). Consequently, the role of primary minerals in the oxidative polymerization of polyphenols and the subsequent formation of humic substances in soils should not be overlooked (Shindo and Huang, 1985b).

SRO aluminosilicates, such as allophone, are known to act as catalysts in the oxidative degradation of polyphenols (Kyuma and Kawaguchi, 1964; Kumada and Kato, 1970). Hydroxy-aluminosilicate ions (proto-imogolite sol) also have the ability to catalyze polyphenol humification (Liu and Huang, 2002). However, the role of other SRO aluminosilicates remains obscure. On the other hand, the formation of SRO aluminosilicates is significantly affected by inorganic ligands, low-molecular-weight organic acids, humic substances, metallic cations, and expansible layer silicates (Huang, 1991a), resulting in the formation of ill-defined aluminosilicate complexes and hydroxyaluminosilicate-intercalated layer silicates. The catalytic ability of these SRO mineral colloids in the transformation of polyphenols and other organic compounds in soils has yet to be investigated.

18.3.1.2 Other Organic Compounds

Many naturally occurring organic acids, including salicylic, pyruvic, oxalic, and malic acids, are degraded by mineral colloids such as Mn oxides by electron transfer reactions (Stone and Morgan, 1984a, 1984b). A wide range of nonphenolic xenobiotic compounds, including pesticides, antibiotics, dyes, and
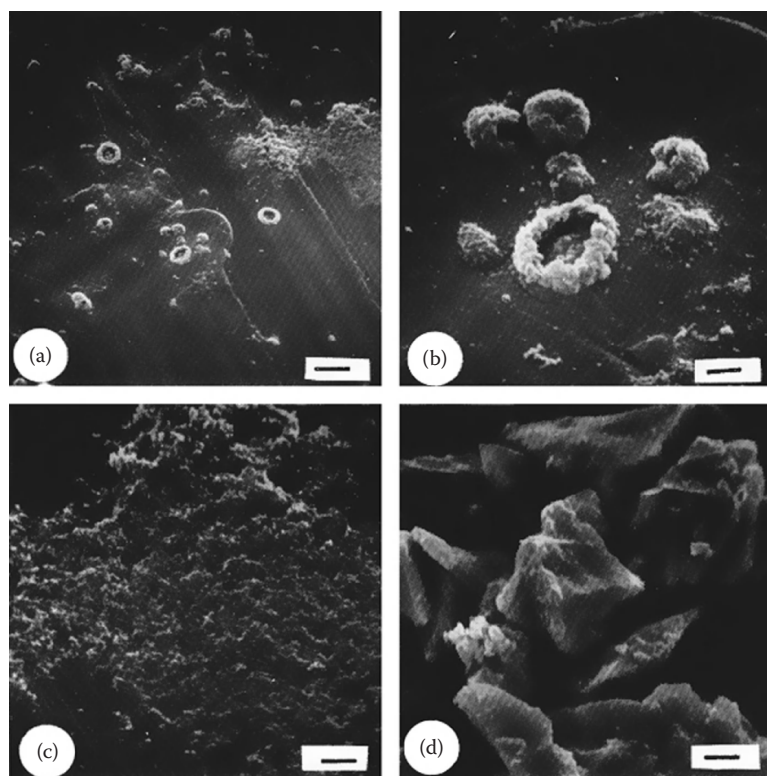
FIGURE 18.1 SEM micrographs of hydroquinone polymers in the supernatant and mineral particles settled in the tephroite system at the ratio of mineral to hydroquinone solution of 0.01 at the initial pH of 6.0 at the end of 7 days. (a, b, c) Hydroquinone polymers; (d) tephroite particles after reaction with hydroquinone. Bar in (a) equals $10\,\mu m$; bars in (b)–(d) equals $2\,\mu m$. (Reprinted from Shindo, H., and P.M. Huang. 1985b. Catalytic polymerization of hydroquinone by primary minerals. Soil Sci. 139:505–511.)

explosives, are subject to abiotic soil mineral-catalyzed oxidative transformation. Abiotic degradation pathways of biocides (pesticides and antibiotics) are particularly important in the environment as the toxicity of these substances can limit their biotic degradation.

18.3.1.2.1 Pesticides

Manganese(IV) oxides are able to catalyze the oxidative polymerization of toxic anilines, chloroanilines, and other aromatic amines, which originate from pesticides, as well as from chemical manufacturing residues (Laha and Luthy, 1990; Pizzigallo et al., 1998; Li et al., 2003). Birnessite is an effective solid-state catalyst of the breakdown of organic pollutants such as the herbicide 2,4-D (2,4-dichlorophenoxyacetic acid) and diethyl ether, which are adsorbed on birnessite and rapidly oxidized (Cheney et al., 1996), both producing CO_2 as a major product, but by somewhat different mechanisms. Nasser et al. (2000) demonstrated a dry mechanochemical technique for quickly and completely degrading 2,4-D by lightly grinding it with birnessite, thus eliminating the need for organic solvents such as diethyl ether. The widely used herbicide atrazine (2-chloro-4-ethylamino-6-isopropylamino-*s*-triazine) is also partially degraded by mechanochemical contact with Mn(IV) oxides (birnessite, cryptomelane, and pyrolusite) via N-dealkylation and subsequent

decarboxylation mechanisms (Shin et al., 2000). The removal of 4-chloroaniline and PCP by a mechanochemical procedure was far more effective than by batch contact (solution) in the presence of birnessite and ferrihydrite (Pizzigallo et al., 2004). The mechanochemical contact of polychlorinated biphenyls (PCBs) and birnessite produced a removal of pollutant that was a function of the number of chlorine atoms, that is, the more the chlorine atoms the less effective the oxidative degradation (Pizzigallo et al., 2004). Possible contributions of solid-state degradation to herbicide breakdown by abiotic catalysis should be considered in modeling herbicide breakdown or when designing experiments for soil remediation.

Birnessite is able to oxidatively degrade glyphosphate (*N*-phosphonomethyl-glycine), which is the most commonly used pesticide worldwide (Barrett and McBride, 2005). The abiotic degradation mechanism involves C–P and C–N bond cleavage of glyphosphate and its degradation product, sarcosine, due to electron transfer reactions at the Mn oxide surface (Figure 18.2; Barrett and McBride, 2005). Pyrolusite is able to effectively oxidatively degrade the widely found organic pollutant 2-mercaptobenzothiazole (a biocide and compound used in manufacture of rubber) into SO_4^{2-} and NO_3^-; however, the presence of organic acids, carboxylic acids (oxalic, citric, tartaric, and malic acids), or metal ions (Ni^{2+}, Ca^{2+}, Mn^{2+}, and Cr^{3+}) has an inhibiting effect on the rate of degradation (Li et al., 2008).
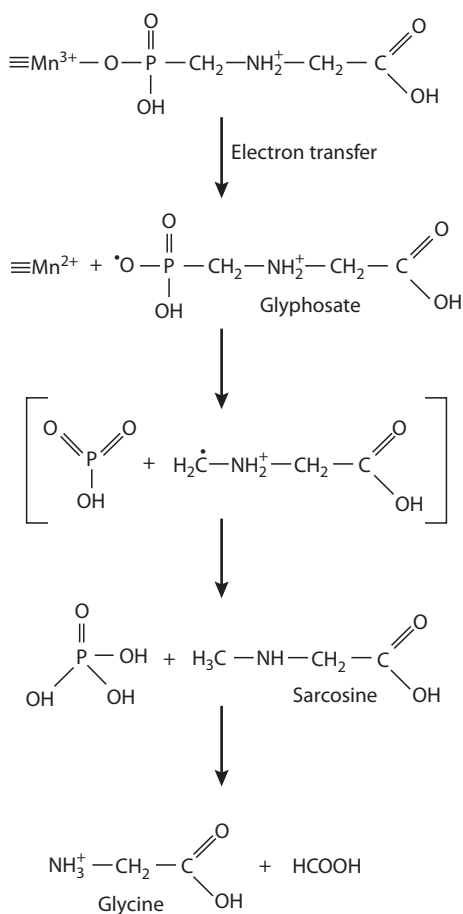
FIGURE 18.2 A possible degradation reaction scheme for glyphosphate adsorbed on Mn oxide. (Reprinted with permission from Barrett, K.A., and M.B. McBride. 2005. Oxidative degradation of glyphosate and aminomethylphosphonate by manganese oxide. Environ. Sci. Technol. 39:9223–9228. Copyright 2005 American Chemical Society.)
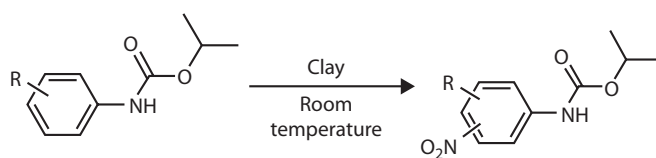
FIGURE 18.3 Chemical structure of carbamates and their nitro derivatives in clay-catalyzed nitration. (Reprinted with permission from Kodaka, R., T. Sugano, T. Katagi, and Y. Takimoto. 2003. Clay-catalyzed nitration of a carbamate fungicide diethofencarb. J. Agric. Food Chem. 51:7730–7737. Copyright 2003 American Chemical Society.)

Clay minerals, kaolinite and montmorillonite, are able to catalyze the nitration of the carbamate fungicide, diethofencarb (isopropyl 3,4-diethoxycarbanilate), which most likely proceeds through the formation of $NO_2^\bullet$ radicals on the surface of the clay (Figure 18.3). The similar nitration of the phenyl ring of the pesticide, famoxadone (3-anilino-5-methyl-5- (4-phenoxyphenyl)-1,3-oxazolidine-2,4-dione), has been reported (Jernberg and Philip, 1999). Nitration of pesticides is only known as a clay-catalyzed reaction (Nikalje et al., 2000).

18.3.1.2.2 Antibiotics

Manganese(IV) oxides are able to oxidatively degrade many of the leading groups of antibacterial agents used in the treatment of humans and animals, in livestock feed additives, or in active ingredients in personal care products (Zhang et al., 2008).

Manganese(IV) oxide is able to catalyze the oxidative degradation of the popular and potent fluoroquinolone (FQ) antibiotic agents under environmental conditions (Zhang and Huang, 2005a). The oxidative degradation mechanism (Figure 18.4) involves a surface reaction mechanism that likely begins with the formation of a surface complex between FQ and the surface-bound Mn(IV), followed by oxidation at the aromatic N1 atom of FQ's piperazine moiety to generate an anilinyl radical intermediate. The radical intermediates subsequently undergo N-dealkylation, C-hydroxylation, and possibly coupling to yield a range of products. The δ-MnO_2 exhibited reactivity to the FQ antibiotics in the order of ciprofloxacin ≈ enrofloxacin ≈ norfloxacin ≈ ofloxacin > lomefloxacin > pipemidic acid ≫ flumequine (Zhang and Huang, 2005a). The veterinary N-oxide antibiotics, carbadox, olaquindox, quinoline N-oxide, and quindoxin also exhibit high oxidation reactivity toward MnO_2 (Zhang and Huang, 2005b). The transformation mechanism involves the formation of an N-oxide radical intermediate and the generation of Mn(II) (Zhang and Huang, 2005b).

Manganese(IV) oxides are also able to promote the appreciable degradation of tetracycline antibiotics (Rubert and Pedersen, 2006). Reactivity of tetracycline antibiotics toward MnO_2 increases in the following order: rolitetracyline ≈ oxytetracycline ≤ tetracycline ≈ meclocycline < chlortetracycline. Abiotic degradation of the veterinary ionophore and macroclide antibiotics, monensin and tylosin, does take place in soil, however at a much slower rate than by microbial degradation in soil (Sassman and Lee, 2007; Sassman et al., 2007).

18.3.1.2.3 Explosives and Dyes

The explosive CL-20 (hexanitrohexaazaisowurtzitane) was found to be quickly (in minutes) abiotically, and oxidatively degraded in the presence of 2:1 layer silicates (montmorillonite, hectorite, and nontronite), micas (biotite and illite), and birnessite (Szecsody et al., 2004). The exact mechanism remains unclear, but it is known that the cage structure of CL-20 is broken and that cleavage of C–C and C–N bonds occurs due to the formation of nitrite and formate anions (Szecsody et al., 2004). Birnessite has also been found to effectively catalyze the complete abiotic transformation of the Fe^0-induced reduction products of the explosive 2,4,6-trinitrotoluene (TNT) via oxidative-coupling reactions (Kang et al., 2006).

Magnetite and other mixed metal (Co, Cu, and Mn)-ferrite oxides are able to efficiently catalyze the oxidative degradation and decolorization of various dyes (bromophenol blue, Chicago sky blue, Cu-phthalocyanine, eosin yellowish, Evans blue, naphthol blue black, phenol red, poly B-411, and reactive orange 16) in the presence of hydrogen peroxide by generating hydroxyl-free radicals (Baldrian et al., 2006). The metal oxide catalysts are

FIGURE 18.4   Proposed reaction scheme for oxidation of fluoroquinolone antibiotics by δ-MnO$_2$. (Reprinted with permission from Zhang, H.C., and C.H. Huang. 2005a. Oxidative transformation of fluoroquinolone antibacterial agents and structurally related amines by manganese oxide. Environ. Sci. Technol. 39:4474–4483. Copyright 2005 American Chemical Society.)

able to retain their activity even after repeated catalytic cycles and over a broad pH range unlike enzymatic catalysts such as laccase (Baldrian et al., 2006). The ferric iron species in layer silicate clays are also able to catalyze the oxidative degradation of the organic dye, malachite green, in the presence of hydrogen peroxide (Cheng et al., 2008). The exchangeable interlayer Fe(III) ions exhibit a much higher reactivity than structural Fe(III) in the octahedral sheet. The Fe-containing clays are chemically and mechanically stable, and no loss of activity was observed after 14 recycles (Cheng et al., 2008).

18.3.2 Polycondensation of Phenolic Compounds and Amino Acids

The transformation of colorless amino acids and polyphenols by quinone–amino acid polycondensation reactions into darkly colored polymers is a significant browning reaction in nature (Bittner, 2006). Birnessite efficiently catalyzes the polycondensation of phenolic compounds and amino acids such as in hydroquinone–glycine systems in the common pH range (4–8) of soils

with the formation of NH$_3$–N and nitrogenous polymers (Table 18.3). The proposed processes are as follows: (1) the Mn oxide acts as Lewis acid by accepting electrons from hydroquinone, which is thus oxidized and subsequently polymerized; (2) the products of the reaction of glycine with hydroquinone polymerize to form nitrogenous polymers, thereby incorporating glycine into the polymers during the oxidative polymerization of hydroquinone; and (3) a partial deamination of glycine occurs during the second process.

Birnessite also catalyzes the polycondensation of glycine and pyrogallol and deamination and decarboxylation of glycine especially in the presence of pyrogallol (Wang and Huang, 1987). The formation of HA and FA is not evident in the presence of glycine alone. In a N$_2$ atmosphere, CO$_2$ release in the birnessite–glycine–pyrogallol system is drastically reduced. Birnessite also greatly enhances the formation of N-containing humic polymers. Most of the released NH$_3$ can be attributed to the deamination of glycine by pyrogallol-derived free radicals as catalyzed by birnessite but an appreciable amount can also be directly derived from the deamination of glycine by birnessite.

TABLE 18.3   Formation of NH$_3$–N and Polymer–N (in mg kg^{-1}) in the Hydroquinone–Glycine System Both in the Absence and Presence of Mn Oxide at Different pH Values at the End of 24 h

	Initial pH of 4.0				Initial pH of 6.0				Initial pH of 7.3			
	Final pH	NH$_3$–N	Polymer–N	Sum-N	Final pH	NH$_3$–N	Polymer–N	Sum-N	Final pH	NH$_3$–N	Polymer–N	Sum-N
Hydroquinone–glycine	4.1	4.4	3.0	7.4	6.1	6.0	2.0	8.0	6.9	11.7	3.4	15.1
Mn oxide hydroquinone–glycine	4.1	30.5	12.6	43.1	7.6	30.2	40.1	70.3	7.9	34.1	41.2	75.3

Source: Shindo, H., and P.M. Huang. 1984b. Significance of Mn(IV) oxide in abiotic formation of organic nitrogen complexes in natural environments. Nature 308:57–58. With permission of Macmillan, London.

Carbon-14 labeling of glycine was used to measure the extent of incorporation of carboxyl C and alkyl C into the polycondensates of pyrogallol and glycine (Wang and Huang, 1997). Birnessite promotes the incorporation of carboxyl and especially alkyl C of glycine into the polycondensates formed in the presence of pyrogallol (Wang and Huang, 1997, 2005). The results indicate that the formation of the C–C bond is more prevalent than that of the C–O bond in the polycondensation of glycine and pyrogallol catalyzed by birnessite.

Besides birnessite, clay size layer silicates also have the ability to catalyze the polycondensation of phenolic compounds and amino acids. Wang et al. (1985) show the catalytic effect of Ca-illite on the formation of N-containing HAs in the systems containing phenolic compounds and amino acids at neutral pH. The yields and N contents of the synthesized HAs depend on the nature of amino acids.

Nontronite in which structural Fe(III) acts as an electron acceptor analogous to Mn(IV) in MnO_2 (Solomon, 1968; Theng, 1974) also catalyzes the polycondensation of glycine and pyrogallol (Wang and Huang, 1991). The IR and electron spin resonance (ESR) spectra of the HA and FA formed, which show the presence of a stabilized organic free radical (semiquinone), are similar to those of natural soil HA and FA (Schnitzer, 1977; Senesi and Schnitzer, 1977; Schnitzer and Levesque, 1979; Hatcher et al., 1980; Schnitzer and Ghosh, 1982). The formation of semiquinone-free radicals appears to proceed through electron transfer from pyrogallol to Fe(III) or other variable-valence transition metal ions on the edges or in the nontronite structure (Solomon, 1968). Electrons diffuse or tunnel to octahedral sites from layer edges or basal surfaces (Tennakoon et al., 1974). The strong oxidation power of chemisorbed O_2 on silicates (Solomon and Hawthorne, 1983) such as nontronite seems to cause the ring cleavage of pyrogallol and the decarboxylation and deamination of glycine (Wang and Huang, 1991). Semiquinone-free radicals, aliphatic fragments, and glycine apparently undergo polycondensation to form humic polymers.

Natural soil clays, from temperate and tropical environments, have been shown to catalyze the polycondensation of pyrogallol and glycine, the ring cleavage of pyrogallol, its polymerization, and the deamination of glycine (Wang and Huang, 2003). The reactivity of the mollisol and oxisol clays was attributed to the catalysis of the layer silicates and oxides of Fe, Al, and Mn present in the natural soil clays (Wang and Huang, 2003).

18.3.3 The Maillard Reaction and Integrated Polyphenol–Maillard Reactions

The Maillard reaction (Maillard, 1913), involving condensation reactions between reducing sugars and amino acids, is considered to be an important pathway in natural humification (Ikan et al., 1996). Sugars and amino acids are among the most abundant constituents of terrestrial and aquatic environments (Anderson et al., 1989). The Maillard reaction involves numerous reactions between the degradation products of the precursor sugars and amino acids, and reaction intermediates, known as Amadori and Heyns compounds (Yaylayan, 1997).

The initial reaction in the Maillard reaction involves condensation between the α-hydroxy carbonyl group of a reducing sugar and the amino group from an amino acid with the formation of a Schiff base, which rearranges to form either Amadori or Heyns compounds. Aldohexoses generate Amadori compounds while ketohexoses generate Heyns products (Yaylayan, 1997). These compounds can undergo retroaldolization reactions forming α-dicarbonyl and α-hydroxyketone compounds. All these compounds are highly reactive and readily polymerize in the presence of amino compounds to form brown-colored melanoidins. Amino acids can react with the α-dicarbonyl compounds, undergoing the Strecker degradation and then forming α-amino ketones. The α-amino ketones may then condense resulting in the formation of pyrazines (Ho, 1996; Yaylayan, 1997). In addition, the Maillard reaction can result in the formation of polyphenols, such as catechol, resorcinol, hydroquinone, and pyrogallol, which can enhance the degree of browning during the Maillard reaction by affecting the redox potential of the system (Haffenden and Yaylayan, 2005) and undergoing polymerization and polycondensation (Wang and Huang, 2005).

Commonly found soil metal oxides, birnessite (Jokic et al., 2001b), goethite (Gonzalez and Laird, 2004), and smectite clays (Gonzalez and Laird, 2004), have been shown to catalyze the Maillard reaction under typical pH and temperature ranges found in the natural environment. Jokic et al. (2001b) were the first to report that birnessite catalyzes the Maillard reaction between glucose and glycine. Their data showed that the presence of a redox-reactive mineral, namely, birnessite, significantly accelerates the reaction by one to two orders of magnitude under environmentally relevant temperatures (25°C and 45°C) and a neutral pH (7.00). This reaction is kinetically sluggish under ambient temperatures (Jokic et al., 2001c) but the presence of birnessite significantly catalyzes the reaction by decreasing the activation energy required. Other groups (Bosetto et al., 1995, 1997; Arafaioli et al., 1997, 1999) have also investigated the clay mineral catalysis of the Maillard reaction between various amino acids and glucose at relatively high temperatures (at 70°C).

Light enhances birnessite catalysis of the Maillard reaction, but the reaction still proceeds in the absence of light (Jokic et al., 2001a). This means that the reaction could readily occur in the subsoil in the presence of a mineral catalyst such as birnessite. The Maillard reaction between glucose and glycine as catalyzed by birnessite results in the formation of humic substances that contain significant amounts of (1) amides, which are the dominant N types in humic substances, soils, and sediments, and (2) heterocyclic N compounds, which are often referred to as *unknown* N (Jokic et al., 2004a) (Figure 18.5). The mineral-catalyzed Maillard reaction provides a possible explanation for one of the pathways for the formation of heterocyclic and amide N found in humic substances in the environment (Jokic et al., 2004a). The action of birnessite on glucose should promote the autoxidation of glucose, which results in the
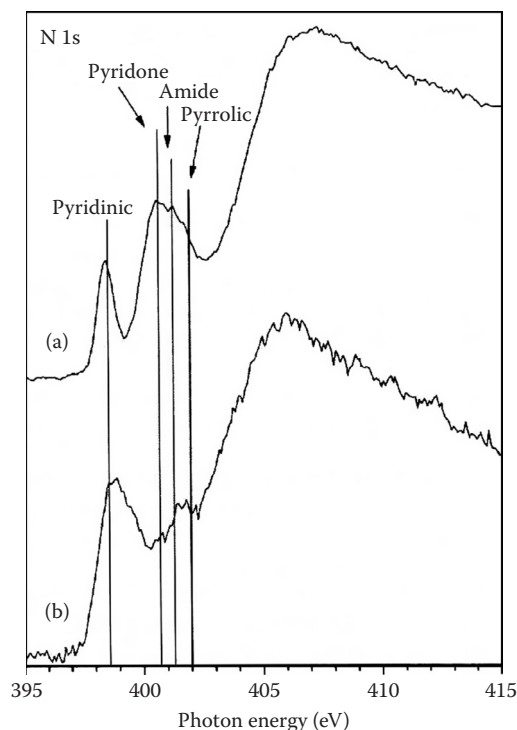
FIGURE 18.5 N 1s XANES spectra of (a) fulvic acid isolated from a glucose–glycine–δ-MnO$_2$ system and (b) the lyophilized solid phase. The peaks are assigned to pyridinic (398.6 eV), pyridone (400.7 eV), amide (401.3 eV), and pyrrolic (402.0 eV) moieties. (Reprinted from Jokic, A., H.-R. Schulten, J.N. Cutler, M. Schnitzer, and P.M. Huang. 2004a. A significant abiotic pathway for the formation of *unknown* nitrogen in nature. Geophys. Res. Lett. 31:L05502. With permission of the American Geophysical Union.)

1. Oxidation of carbohydrate or Strecker aldehyde by δ-MnO$_2$ to form carboxylic acid, or formation of carboxylic acid during Strecker degradation

2. Deamination of amino acid by δ-MnO$_2$

3. Amide formation

FIGURE 18.6 Possible amide formation pathway including the key role of MnO$_2$. (Reprinted from Jokic, A., H.-R. Schulten, J.N. Cutler, M. Schnitzer, and P.M. Huang. 2005. Catalysis of the Maillard reaction by δ-MnO$_2$: A significant abiotic sorptive condensation pathway for the formation of refractory N-containing biogeomacromolecules in nature, p. 127–152. *In* P.M. Huang, A. Violante, J.-M. Bollag, and P. Vityakon (eds.) Soil abiotic and biotic interactions and impact on the ecosystem and human welfare. Science Publishers Inc., Enfield, NH. Copyright Science Publishers Inc. With permission.)

generation of reactive dicarbonyl compounds, hydrogen peroxide, and hydroxylating agents (Wolff, 1996). These reactive dicarbonyl compounds would then react with ammonia formed by the known deamination of glycine catalyzed by birnessite (Wang and Huang, 1987), as proposed by Vairavamurthy and Wang (2002) or with glycine (undergoing the Strecker degradation), resulting in either case in the formation of heterocyclic N compounds (Wong and Shibamoto, 1996). Ammonium ions are known to react with compounds containing reactive carboxyl groups to form pyridinic structures (Steelink, 1994). Carboxylic acids are produced during the Strecker degradation of amino acids (Wong and Shibamoto, 1996) or by the action of manganese dioxides on simple carbohydrates (Bose et al., 1959; Gonzalez and Laird, 2006). When heated, carboxylic acids react with ammonia to form amides (Smith and March, 2001). The scheme for possible amide formation is shown in Figure 18.6.

Gonzalez and Laird (2004) showed that four different smectites saturated with various metal cations (Ca, Na, Cu(II), and Al) could catalyze the Maillard reaction between arginine and glucose at an environmentally relevant temperature (37°C). They found that only saturation with Cu(II) significantly altered the amount of humic substances produced. They also observed that some of the adsorbed humic substances were intercalated into the smectites.

Jokic et al. (2004b) were the first to study an integrated Maillard reaction and polyphenol (glucose, glycine, and catechol) pathway of humification, using birnessite as catalyst. Their data showed that birnessite significantly accelerates humification processes in an integrated polyphenol–Maillard reaction system under ambient conditions (25°C and 45°C). In nature, it is likely that the Maillard reaction and polyphenol pathways interact closely, since sugars, amino acids, and polyphenols coexist in soil solutions and natural waters. Jokic et al. (2004b) also showed that the integrated polyphenol–Maillard reaction system is more effective in generating humic polymers than the Maillard reaction alone. The type of polyphenol and the molar ratio of polyphenol to Maillard reagents in the polyphenol–Maillard reaction system significantly affect not only the humification processes but also the biomolecule-induced formation of inorganic C, namely, rhodochrosite (MnCO$_3$) (Hardie et al., 2009). Increasing the molar ratio of polyphenols to Maillard reagents results in a significant enhancement of humification in the polyphenol–Maillard reaction in the presence of birnessite. Polyphenols with *ortho*-OH groups (pyrogallol) participate in direct electron transfer reactions with birnessite more readily than polyphenols with *meta*-OH groups (resorcinol), and subsequently undergo oxidative polymerization and ring-cleavage reactions to a greater extent. Thus, the humic polymers formed in the pyrogallol–Maillard system are significantly more aliphatic in character than those formed in the resorcinol–Maillard system. Furthermore, the greater extent of polymerization reactions in the birnessite-catalyzed pyrogallol–Maillard system results in the suppression of MnCO$_3$ formation. Hardie et al. (2007) also compared the HA fraction formed in the presence of birnessite from the Maillard reaction and two

polyphenol–Maillard reaction systems with natural soil and peat HA using C K-edge NEXAFS and found that these HAs are basically similar, especially the HA from the Maillard reaction and pyrogallol–Maillard systems.

18.3.4 Surface Brønsted and Lewis Acidity and Hydrolysis of Organic Compounds

Mineral colloids may detoxify adsorbed pesticides by catalyzing their decompositions through the ability of mineral colloids to behave as Brønsted acids and donate protons or to act as Lewis acids and accept electron pairs. Brønsted acidity derives essentially from the dissociation of water molecules coordinated to surface-bound cations, resulting in the formation of active protons. This acidity is, thus, strongly influenced by the hydration status and polarizing power of surface-bound and structural cations on mineral colloids (Mortland, 1970, 1986). Low water contents and highly polarizing cations (small, high charge) promote Brønsted acidity. Lewis acidity arises from constituent ions such as Al and Fe ions exposed at the edges of mineral colloids (Sposito, 1984; McBride, 1994).

Many uncharged organic molecules that require extremely acid conditions to accept a H^+ ion in solution can be protonated on the surface of mineral colloids (Huang, 1990; McBride, 1994). The Brønsted acid strength and the degree of protonation are related to the electronegativity and polarizing power of the exchangeable and structural metal cations in the following order (McBride, 1994): $H^+ > Al^{3+}$, $Fe^{3+} > Mg^{2+} > Ca^{2+} > Na^+ > K^+$. As clay surfaces become drier, the Brønsted acidity increases and protons are concentrated in a smaller volume of water resulting in more extreme surface acidity. Even very weak bases, that is, poor proton acceptors can be protonated on the surfaces of such mineral colloids.

The degradation of certain pesticides is catalyzed as the result of their adsorption on mineral colloids, which was first observed by formulation chemists who used clays as carriers and diluents (Fowkes et al., 1960). Surface-catalyzed degradation of pesticides on clays was later demonstrated for several organophosphate and *s*-triazine pesticides and has been attributed to the surface acidity of clay minerals (Brown and White, 1969; Salztman et al., 1974; Minglegrin et al., 1977). Minglegrin and Salztman (1979) demonstrated that the nature of the clay, its saturating cation, and hydration status determine the rate and mechanism of degradation of parathion. The adsorption-catalyzed degradation of parathion is a hydrolysis reaction that proceeds either directly or through a rearrangement step.

Magnesium-saturated montmorillonite converts the amino form of 3-aminotriazole to the imino form (Russell et al., 1968a) as shown below:

$$(18.8)$$
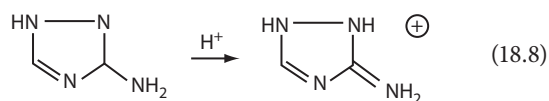
Triazine compounds can also be protonated on dry mineral colloids. Armstrong et al. (1967) showed that the addition of sterilized soil to atrazine solution increased the hydrolysis rate of atrazine 10-fold. Soil catalytic degradation of atrazine through hydrolysis is an important pathway. Mineral surface acidity also catalyzes hydrolysis of the chloro-*s*-triazine herbicides to the nonphytotoxic 2-hydroxy-*s*-triazines (Russell et al., 1968b) and may be of significance in a wide range of other organic compounds (McBride, 1994).

Besides the Brønsted acid, the Lewis acid properties of metals appear to be of significance in mineral-catalyzed hydrolysis reactions. The oxides of Fe and Al in water, and especially in the dry state, have some catalytic function in organic hydrolysis reactions, at least for those that are known to be hydroxyl ion catalyzed (Hoffmann, 1990). Those hydrolyzable organics that have a suitable structure to chelate with the surface metal cations are, in general, the most susceptible to mineral-catalyzed degradation.

Dissolved metals and metal-containing surfaces play an important role in catalyzing the hydrolysis of organic pollutants. The catalytic effectiveness of a metal ion depends on its ability to complex reactant molecules (e.g., ester linkage, leaving group, and attacking nucleophile) and shift electron density and conformation in ways favorable to reaction (Hoffmann, 1980). Metals vary in their complex formation constants, which reflect difference in metal–ligand bond strengths and solvation forces. The following rules can explain the general features: (1) complex formation constants generally increase as the charge to radius ratio of the metal ion is increased, because the electrostatic contribution to bond formation is increased; (2) polarizable metals and ligands gain additional complex stability through covalent bond formation; and (3) as complex formation constants increase, competition for the metal among available ligands (including OH^- and other inorganic constituents) becomes more pronounced. Trivalent metals such as Al and Fe(III) and tetravalent metals such as Ti(IV) are classic hard metals, which form strong complexes with classic hard ligands (Pearson, 1966). Organic compounds possessing O and N donor atoms are classified as hard ligands, whereas those possessing S donor atoms are classified as soft ligands. Oxygen donor organic ligands form strong complexes with metals such as Al, Fe(III), and Ti(IV). However, they compete with hydroxo (OH^-) and oxo (O^{2-}) species that limit metal solubility near-neutral pH values. Although Ca^{2+} and Mg^{2+} are also classified as "hard" metals, their s and p orbitals form bonds that are primarily ionic in nature (Stone and Torrents, 1995).

Complex formation constants of metals with a ligand primarily vary with the ionic radius of metals, although increases in polarizability with increasing atomic number are also important (Houghton, 1979). Even if differences in metal–ligand complex formation constants are taken into consideration, differences in electronic structure cause some metals to emerge as particularly effective catalysts.

Metals can form a series of metal hydroxo complexes (Bases and Mesmer, 1986). Although metal hydroxo complexes are less reactive nucleophiles than hydroxide ions, their concentrations can be substantially higher, especially in the neutral to acidic

pH range. Because the pK_1 values for strong Lewis acid metals are low, they generate hydroxo species across a wider pH range. The nucleophiles of various metal hydroxo complexes are governed by pK_1 and the polarizability of the metal (Stone and Torrents, 1995).

The reactivity of hydrolyzable organic pollutants arises from the presence of electrophilic (electron deficient) sites within the molecules (Stone and Torrents, 1995). The S_N2 mechanism (neucleophilic substitution) involves attack of the electrophilic sites by OH^- or H_2O, generation of a higher coordination number intermediate, subsequent elimination of the leaving group, and the formation of a hydrolysis product:

S_N2

$$(18.9)$$
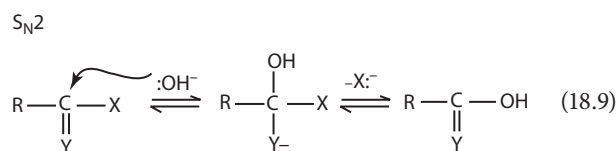
In the case of the S_N1 mechanism (nucleophilic substitution monomolecular), the reaction proceeds with the loss of the leaving group to generate a lower coordination number intermediate and then followed by generation of the hydrolysis product by nucleophilic addition as shown below:

S_N1

$$(18.10)$$
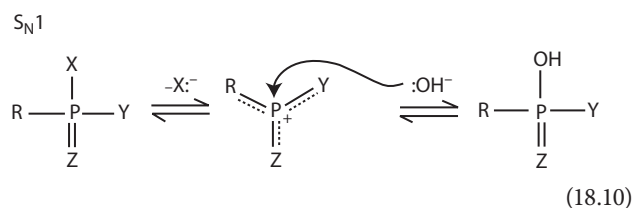
It is generally accepted that metal ions can catalyze hydrolysis in a way similar to acid catalysis. Metal ions and protons coordinate to the pollutants so that electron density is shifted away from the site of nucleophilic attack to facilitate the reaction. Because protons have an extremely high charge density and great polarizing power, metal catalysis is insignificant in acidic solutions. However, metal ions can readily coordinate two or more ligand donor sites on a molecule and can greatly outnumber protons in neutral and alkaline conditions (Plastourgou and Hoffmann, 1984). A list of metal catalysis mechanisms is given in Table 18.4.

The hydrolysis of compounds possessing good leaving groups is limited by the rate of nucleophilic attack. These compounds are susceptible to types 1, 3, 4, 6, and possibly 5 of metal *catalysis*. The hydrolysis of compounds possessing poor leaving groups is limited by the breakdown of the tetragonal intermediate; therefore, types 2, 7, and possibly 5 are important. A number of pesticides and other pollutants are in the intermediate region, where metal catalysis may shift the rate-limiting step from nucleophilic attack to breakdown of the tetrahedral intermediate or vice versa.

Many organic compounds are susceptible to metal ion catalysis. These include carboxylic acid esters, amides, anilides, phosphate-containing esters, and other hydrolyzable compounds

TABLE 18.4 Mechanisms of Metal Catalysis

Type 1	The metal coordinates the electrophile, shifting the electronic distribution in the molecule in a way that enhances its reactivity
Type 2	The metal coordinates to the leaving group, increasing its leaving ability
Type 3	The metal acts as a center for simultaneous attachment of both the electrophile and attacking nucleophile (template effect)
Type 4	The metal coordinates the nucleophile and induces deprotonation (which increases the reactivity of the nucleophile)
Type 5	Coordination of the substrate with the metal induces confirmation changes that facilitate reaction
Type 6	Coordination of the substrate with the metal makes the molecule more positive, lessening unfavorable electrostatic interaction with the nucleophile
Type 7	Coordination with the metal blocks inhibitory reverse reaction paths, such as (1) loss of OH^- from a tetrahedral intermediate instead of loss of the leaving group X^- or (2) nucleophilic attack by X^-

Source: From Stone, A.T., and A. Torrents. 1995. The role of dissolved metals and metal-containing surfaces in catalyzing the hydrolysis of organic pollutants, p. 275–298. *In* P.M. Huang, J. Berthelin, J.-M. Bollag, W.B. McGill, and A.L. Page (eds.) Environmental impact of soil component interactions, Vol. 1. Natural and anthropogenic organics. CRC Press/Lewis Publishers, Boca Raton, FL. With permission.

(Stone and Torrents, 1995). In the mid-1950s, the catalytic ability of metals in hydrolysis of thionophosphate pesticides such as parathion and EPN (thionobenzene phosphoric acid *O*-ethyl-*O*-*p*-nitrophenyl ester) was recognized (Ketelaar et al., 1956). Rate enhancements arising from the addition of 1.0×10^{-4} mol L^{-1} Cu^{2+} are as high as 20-fold for parathion at pH 8.5 and 48-fold for EPN at pH 8.2. Coordination of the S donor group by Cu^{2+} activates the P center to nucleophilic attack (type 1 catalysis). Chlorpyrifos, a phosphorothionate pesticide, contains a pyridyl N capable of forming a six-membered chelate ring with the Cu^{2+} ion (Blanchet and St.-George, 1982).

Metal ion catalysis arises from the increased electrophilicity of the P atoms in the chelate ring (type 1 catalysis) (Blanchet and St.-George, 1982) and from possible leaving group effects (type 2 catalysis). Copper(II) is by far the most reactive catalyst among a number of metals [including Mg(II) and Al(III)], which have been observed to catalyze chlorpyrifos hydrolysis. However, Cu^{2+} concentration would have to exceed 10^{-5} mol L^{-1}, which seldom occurs in soil solutions and other natural waters, in order for the metal-catalyzed pathway to surpass the rate of the uncatalyzed hydrolysis pathway (Mill and Mabey, 1988).

Many metals, which are potentially catalytic, form low-solubility inorganic solids within the pH domain of soils and associated environments. Therefore, surface-bound metals must be accessible to reactants in order for a metal-catalyzed effect to be observed. Since the mid-1950s, research has established that metal oxide/hydroxide precipitates act as hydrolysis catalysts for phosphate esters (Butcher and Westheimer, 1955; Wilkins, 1991). Much has been subsequently learned about the chemical properties and reactivities of naturally occurring solids such as oxides, carbonates, sulfides, and aluminosilicates (Stone and Torrents, 1995; Smolen and Stone, 1998).

Complex formation equilibria among dissolved and surface-bound metal species are similar in many aspects. Surface complex formation constants for the metal ions and organic ligands increase in magnitude as the formation constants of analogous complexes in solution are increased (Kummert and Stumm, 1980; Stumm et al., 1980; Davis and Hayes, 1986). Attempts have been made to compare reactivities of dissolved and surface-bound metal species (Wehrli, 1990; Wehrli et al., 1990). To compare the catalytic reactivity of dissolved and surface-bound metals, any unusual or unique characteristic of the mineral/water interface must be investigated. The high Lewis and Brønsted acidity of unoccupied coordinative sites is responsible for the well-known hydrolytic reactivity of dehydrated and partially dehydrated mineral surfaces (Theng, 1982; Voudrias and Reinhard, 1986). Although the catalysis by surface-bound metals is influenced by hydration, dramatic effects of solid surfaces on hydrolysis rates in aqueous suspensions have been demonstrated (Sanchez-Camazano and Sanchez-Martin, 1983).

The catalysis by surface-bound metals is observed when all participating reactants are adsorbed to a significant extent and when rate constants for the reactions at the mineral–water surface are comparable to or exceed rate constants for the reaction in homogeneous solution. Although adsorption phenomena have accounted for catalysis, relatively little is known about the conformation and stoichiometry of adsorbed species. This hampers our understanding of surface catalysis on a fundamental level. However, based on the existing literature, two generalities can be made: (1) auxiliary donor groups that facilitate metal catalysis by metal ions in solution should also facilitate surface catalysis and (2) phenomena of only secondary importance in reactions of dissolved complexes, for example, electrostatic and hydrophobic interactions may play a much greater role in surface catalysis (Stone and Torrents, 1995). The nature and potential significance of surface catalysis are summarized later.

The hydrolysis of phenyl picolinate (PHP) is catalyzed by both surface-bound and homogeneous solution metal ions (type 1 catalysis; Fife and Przystas, 1985; Torrents and Stone, 1991). Appropriate metal ions chelate the heterocyclic N and carbonyl O of PHP and increase the partial positive charge at the carbonyl C, thus facilitating nucleophilic attack and increasing its susceptibility to hydrolysis. Suspensions containing FeOOH or TiO_2 dramatically accelerate PHP hydrolysis rate, while in particle-free solution and suspensions containing Al_2O_3 or SiO_2, hydrolysis is negligible (Figure 18.7). Similar results are obtained with methyl picolinate (MEP; Stone and Torrents, 1995). Catalysis arises from reaction at the oxide surface and not from release of soluble metals, since removal of particles by filtration causes an immediate halt to any catalytic activity.

The susceptibility of hydrolyzable compounds to metal catalysis depends on the nature of the auxiliary donor group. The pyridyl N of PHP and MEP encourages chelate formation with Ti(IV) and Fe(III) in preference to less polarizable Al. Therefore, TiO_2 and FeOOH are effective hydrolysis catalysts and Al_2O_3 is not. In contrast, O donor groups, such as the phenolic group in phenyl salicylate and the O heteroatom of methyl furanoate, are harder and thus more favorable to Al chelate formation (Stone and Torrents, 1995). Hydrolysis of PHP in the presence of TiO_2 is virtually pH independent and in the presence of FeOOH exhibits only a slight pH dependence. In particle-free solution, PHP exhibits classic base-catalyzed hydrolysis behavior; the reaction rate of the hydrolysis increases by an order of magnitude for every unit increase in pH. Activation of the ester linkage through chelate formation with a surface-bound metal ion is apparently strong enough to react with the weak nucleophile H_2O, which dominates over the pH-dependent reaction with OH^-.

Positively charged surfaces such as Al_2O_3 (below pH 8.6) and TiO_2 (below pH 6.4) accelerate the hydrolysis of monophenyl terephthalate (MPT) by an order of magnitude or more (Stone and Torrents, 1995). Hydrolysis of MPT catalyzed by Al_2O_3
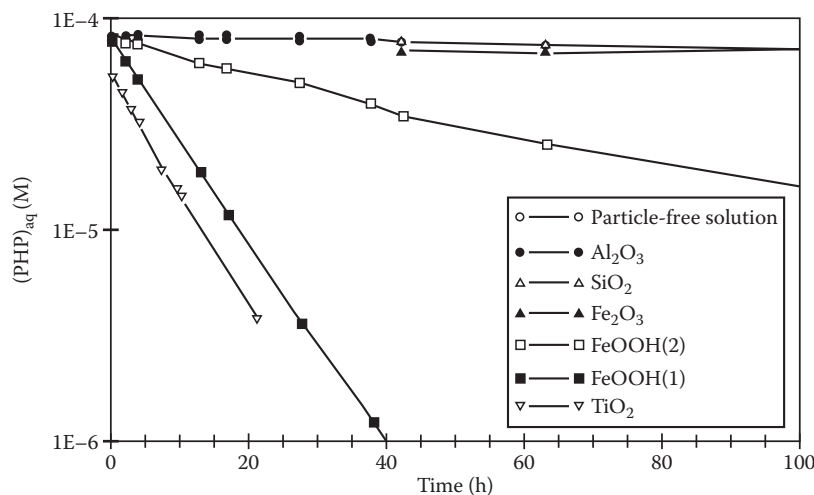
FIGURE 18.7 Effect of various metal oxides on the loss of phenyl picolinate (PHP) from solution via hydrolysis. All suspensions contained $10\,g\,L^{-1}$ oxide, $1.0\,m\,mol\,L^{-1}$ acetate buffer (pH 5.0), and 50 mM NaCl. (Reprinted with permission from Torrents, A., and A.T. Stone. 1991. Hydrolysis of phenyl picolinate at the mineral/water interface. Environ. Sci. Technol. 25:143–149. Copyright 1991 American Chemical Society.)

and TiO_2 exhibits elements of both type 3 and type 6 catalyses. Positive oxide surfaces serve as the template to accumulate both anionic reactants, OH^- and MPT ions, increasing their encounter frequency. Type 6 catalysis (electrostatic interactions favoring encounter between like-charge species) is more important for surface chemical reactions than for solution reactions due to the additive electrostatic effect arising from neighboring charged groups. Type 3 catalysis (simultaneous attachment of substrate and nucleophile) is also typical of surface chemical reactions.

Many pesticides are susceptible to surface-catalyzed hydrolysis (Torrents, 1992). For example, catalysis of chlorpyrifos hydrolysis by TiO_2, FeOOH, and Al_2O_3 apparently involves chelate formation between the surface-bound metal ion, the thionate S, and pyridinyl N.

A surface-catalyzed reaction following type 4 catalysis has been postulated for the degradation of the fungicide oxycarboxin stored in borosilicate glass (Stanton, 1987). Because of low hydroxide ion concentration in neutral and acidic pH, hydrolytic ring cleavage is not observed in the absence of a catalyst. The glass exerts its catalytic effect by providing sufficiently strong nucleophiles (surface-bound OH^- and O_2^{2-}) for attacking the oxathiin ring. Quartz and aluminosilicate surfaces present in soils and associated environments could also catalyze oxycarboxin hydrolysis. It has been reported that metal oxides catalyze the hydrolysis of *p*-nitrophenyl acetate by this mechanism (Hoffmann, 1990). However, in surface-catalyzed

hydrolysis reactions, the nature and relative significance of type 4 catalysis are poorly understood and, thus, merit closer attention.

Among the clay minerals, montmorillonite, beidellite, illite, and vermiculite, only montmorillonite is able to enhance the hydrolysis of the carbamate pesticides, carbosulfan, and aldicarb, probably due to its stronger surface acidity and chelating ability compared to the other clay minerals (Wei et al., 2001). Suspensions of sea sand, TiO_2, α-Fe_2O_3, α-FeO(OH), laponite (smectite clay), and SiO_2 are able to catalyze the hydrolysis of the cyclodiene pesticide, endosulfan to the less toxic endosulfan diol (Walse et al., 2002). The rates of endosulfan hydrolysis over the different surfaces correspond to their tritium-exchange site-density and suggest a mechanism involving surface coordination prior to nucleophilic attack (Walse et al., 2002).

Birnessite is able to catalyze the hydrolysis of the herbicide atrazine (2-chloro-4-(ethylamino)-6-(isopropylamino)-1,3,5-triazine) by acting as a Lewis acid activator and a source of metal-bound hydroxide as a nucleophile (Figure 18.8) (Shin and Cheney, 2004, 2005). Furthermore, birnessite is also able to catalyze the N-dealkylation of atrazine via a nonoxidative mechanism involving proton transfer to Mn(IV)-stabilized oxo and imido bonds (Wang et al., 1999; Shin and Cheney, 2005).

Recently, it has been shown that dissolved HAs are able to catalyze the abiotic hydrolysis of phenylurea herbicides by acting as bifunctional buffers, similar to HCO_3^-/CO_3^{2-} or $H_2PO_4^-/HPO_4^{2-}$. The hydrolysis reaction rate increases with the carboxyl group
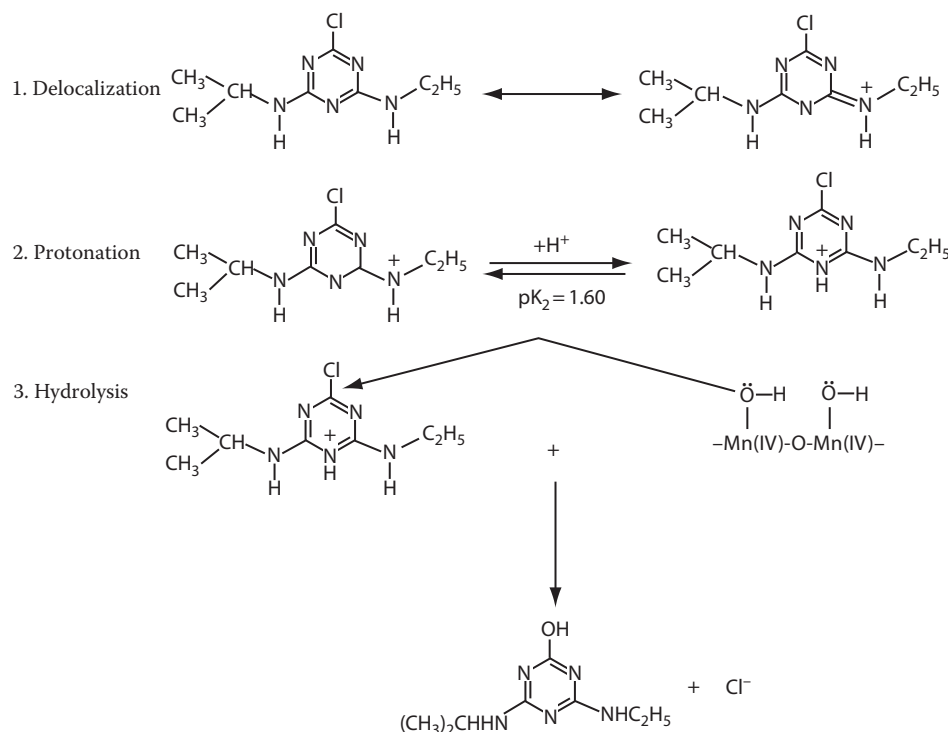
FIGURE 18.8 Hypothesized hydroxyatrazine formation pathway from the reaction of atrazine with birnessite (δ-MnO_2) at pH 4.3 and 20°C. (From Shin, J.Y., and M.A. Cheney. 2005. Abiotic dealkylation and hydrolysis of atrazine by birnessite. Environ. Toxicol. Chem. 24:1353–1360. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.)

content of the HAs, indicating that the carboxyl groups are responsible for the catalysis (Salvestrini et al., 2008).

18.3.5 Reductive Transformation of Organic Compounds

In anoxic groundwater and surficial water, biogeochemical processes provide a variety of potential reductants for abiotic contaminant transformation. Persistent xenobiotics, such as polyhalogenated alkanes and alkenes, aromatic nitro and azo compounds, may be transformed to reaction products that can be further microbially degraded or to products that can be more toxic than the parent compounds (Hofstetter et al., 2006).

18.3.5.1 Ferrous Iron as Catalyst

Soluble Fe(II) is an important reductant in anaerobic aqueous environments; however, its availability is limited in neutral and alkaline pH ranges by the precipitation of FeS, $Fe_3(PO_4)_2$, $FeCO_3$, and other Fe(II)-rich minerals (Charlet et al., 1998). Ferrous iron present within or sorbed on iron oxide mineral particles is highly reactive even at high pH conditions (Charlet et al., 1998). Reduction of nitrite ions to N_2O by Fe(II) sorbed on lepidocrocite at a pH range of 6.0–8.5 was demonstrated by Sorensen and Thorling (1991). Above pH 5.5, Fe^{2+} starts to be adsorbed onto particles, and this adsorption is completed on most mineral phases around pH 7.5, that is, before the onset of precipitation. Above pH 6.5, a significant part of sorbed Fe(II) is present as a highly reactive hydroxylated surface complex, =Fe(III)–O–Fe(II)–OH (Charlet et al., 1998). This hydrolyzed surface species is a very effective reductant toward many persistent xenobiotics, such as 4-chloronitrobenzene (Charlet et al., 1998).

Mossbauer spectra of Fe(II) adsorbed to rutile (TiO_2) and aluminum oxide (Al_2O_3) show only Fe(II) species, whereas spectra of Fe(II) reacted with the iron oxides, goethite, hematite, and ferrihydrite demonstrate electron transfer between the adsorbed Fe(II) and the underlying iron(III) oxide (Williams and Scherer, 2004). The electron transfer reactions that occur during the reduction of nitrobenzene induce growth of an Fe(III) layer on the oxide surface that is similar to the bulk oxide (Williams and Scherer, 2004). Similarly, it has been shown that the reductive degradation of 4-chloronitrobenzene and trichloronitromethane in the presence of Fe(II) and goethite results in the growth of goethite in the c-direction; however, the newly formed goethite is progressively less reactive than the original goethite (Chun et al., 2006).

Ferrous iron in the presence of goethite can catalyze the abiotic reductive degradation of the widely used dinitroaniline herbicides (trifluralin, pendimethalin, nitralin, and isopropalin) at neutral solution pH (Klupinski and Chin, 2003; Wang and Arnold, 2003). Increasing the solution pH results in increased rate constants due to the increase in Fe(II) sorbed and creation of highly reactive species on the surface of the goethite (Figure 18.9) (Wang and Arnold, 2003). Nitrobenzenes and N-hydroxylanilines are reaction intermediates of the reduction of nitroaromatics. Nitrobenzenes, the first intermediates,
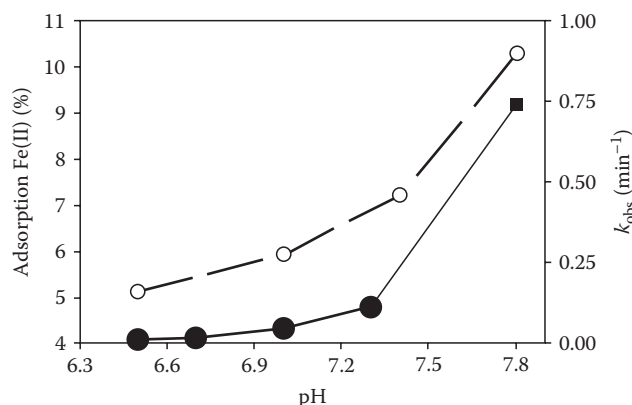
FIGURE 18.9 Role of pH on the sorbed Fe(II) (○) and observed pseudo-first-order rate constants (●) for the reduction of trifluralin with $1\,mmol\,L^{-1}$ Fe(II) in a suspension of $7.2\,m^2\,L^{-1}$ ($0.65\,mg\,mL^{-1}$) goethite. The k_{obs} value (■) at pH 7.8 was extrapolated from a regression of log k_{obs} versus pH. (Reprinted from Wang, S., and W.A. Arnold. 2003. Abiotic reduction of dinitroaniline herbicides. Water Res. 37:4191–4201. Copyright 2003, With permission from Elsevier.)

are readily reduced by soluble Fe(II) and Fe(II)-treated goethite suspensions, especially if there are electron-withdrawing substituents in the *para* position (Colon et al., 2006b). However, N-hydroxylanilines, the second intermediates, are only reduced in the presence of Fe(II)-treated goethite suspensions, and in contrast to nitrobenzenes, electron-withdrawing substituents in the *para* position decrease the rate of reduction (Colon et al., 2006b). The rate of reduction of p-cyano-N-hydroxylaniline showed a linear relationship against the concentration of mineral surface-associated Fe(II) in hematite, goethite, and lepidocrocite suspensions (Colon et al., 2006b). The reactivity of mineral surface-associated Fe(II) toward the reduction of p-cyano-nitrobenzene decreases in the order hematite > goethite > lepidocrocite > ferrihydrite, and interestingly, the surface density of surface-bound Fe(II) does not play a major role in determining the observed reactivity trend (Colon et al., 2006a).

Ferrous iron in the presence of goethite and magnetite abiotically catalyzes the degradation of trichloronitromethane via reduction, while trichloroacetonitrile, 1,1,1-trichloropropanone, and trichloroacetaldyde hydrate are transformed via both hydrolysis and reduction reactions (Chun et al., 2005). The Fe(II) bound to the iron minerals has a greater reactivity than either aqueous Fe(II) or structural Fe(II) present in magnetite (Chun et al., 2005). Ferrous iron bound to magnetite is an effective catalyst for the degradation of the high-energy explosive RDX (hexahydro-1,3,5-trinitro-1,3,5-triazine), ultimately breaking it down into NH_4^+, N_2O, and formaldehyde (Gregory et al., 2004).

Ferrous iron sorbed on the surface of goethite is able to catalyze the abiotic reductive dehalogenation of carbon tetrachloride (CCl_4) to form the even more toxic compound, chloroform ($CHCl_3$) (Amonette et al., 2000). The reductive dehalogenation of carbon tetrachloride did not occur in the presence of soluble Fe(II) without goethite (Amonette et al., 2000). Reduction of CCl_4 by surface-associated Fe(II) on goethite occurs in a single

electron transfer reaction and results in the cleavage of one C–Cl bond and the generation of unstable trichloromethyl radicals, as shown below (Elsner et al., 2004a):

$$CCl_4 \xrightarrow[{-1\ Cl^-}]{+1\ e^-} {}^{\bullet}CCl_3 \begin{array}{l} \xrightarrow{O_2} \longrightarrow \longrightarrow (O{=}CCl_2),\ CO_2 \\ \xrightarrow{R\text{-}S^-} \text{Sulfur adducts} \\ \xrightarrow{e^-} {:}CCl_3^- \\ \xrightarrow{R\text{-}H} CHCl_3 \end{array} \qquad (18.11)$$
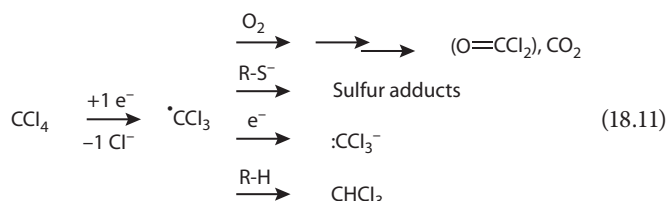
The presence of oxygen and sulfur species can thus lead to completely dehalogenated reaction products in radical reactions. However, the presence of H radical donors, such as trace amounts of natural organic matter, promotes the formation of toxic chloroform (Elsner et al., 2004a). Ferrous iron bound to the surface of iron oxides (goethite, hematite, lepidocrocite, and magnetite) is capable of reducing a variety of other polyhalogenated methanes under aquifer conditions, and the pseudo-first-order reaction rate constants, k_{obs}, of the polyhalogenated methanes increased in the order $CHBrCl_2 < CHBr_2Cl < CHBr_3 < CCl_4 < CFBr_3 < CBrCl_3 < CBr_2Cl_2$ (Pecher et al., 2002).

Reactive Fe(II) does form not only on the surface of oxides but also on the surfaces of other iron minerals such as iron sulfide (FeS), pyrite (FeS_2), and siderite ($FeCO_3$) (Elsner et al., 2004b). Elsner et al. (2004b) compared the 4-chloronitrobenzene, 4-chlorophenyl hydroxylamine, and hexachloroethane-reducing reactivity of Fe(II) aqueous suspensions in the presence of siderite, hematite, lepidocrocite, goethite, magnetite, sulfate green rust, pyrite, and mackinawite. The surface area-normalized reaction rates generally increase in the order Fe(II) + siderite < Fe(II) + iron oxides < Fe(II) + iron sulfides (Elsner et al., 2004b).

Ferrous iron-bearing minerals are also important environmental abiotic reductants. Magnetite, pyrite, and green rust are able to catalyze the abiotic reductive dechlorination of chlorinated ethylenes (Kriegman-King and Reinhard, 1994; Butler and Hayes, 1998; Lee and Batchelor, 2002a, 2002b). Surface area-normalized pseudo-first-order initial rate constants for the reduction of chlorinated ethylenes (tetrachloroethylene, trichloroethylene, *cis*-dichloroethylene, and vinyl chloride) by green rust were found to be 3.4–8.2 times greater than those by pyrite (Lee and Batchelor, 2002b), and the rate constants by pyrite were found to be 23.5–40.3 times greater than those by magnetite (Lee and Batchelor, 2002a).

In clay minerals, Fe(II) may be present as (1) structural Fe(II), (2) Fe(II) complexed by surface hydroxyl groups, and (3) Fe(II) bound by ion exchange. Hofstetter et al. (2003) investigated the accessibility and reactivity of these three types of Fe(II) species in suspensions of ferrous iron-bearing nontronite or iron-free hectorite. They found that both structural Fe(II) and Fe(II) complexed by surface hydroxyl groups of nontronite reduced the nitroaromatic compounds to anilines. Fe(II) bound by ion exchange did not contribute to the observed reduction of the nitroaromatic compounds. The abiotic reductive reactivity of

biotite, vermiculite, and montmorillonite on chlorinated ethylenes (tetrachloroethylene, trichloroethylene, *cis*-dichloroethylene, and vinyl chloride) with or without the addition of Fe(II) was investigated by Lee and Batchelor (2004). Biotite had the greatest rate constant among the phyllosilicates, both with and without Fe(II) addition (Figure 18.10), which is attributed to the higher content of Fe(II) sites on biotite. The Fe(II) content of biotite is 8 and 97.5 times higher than that of vermiculite and montmorillonite, respectively (Lee and Batchelor, 2003). The addition of Fe(II) to the mineral suspensions increased the rate constants of the dechlorination reactions, which is attributed to the regeneration of active sites on the phyllosilicates resulting from redox reaction with Fe(II) or due to the reactivity of Fe(II) that binds to the phyllosilicate surfaces (Lee and Batchelor, 2004).

Ferruginous smectite, chemically reduced with dithionite is able to reduce and dechlorinate trichloroacetonitrile and chloropicrin (Cervini-Silva et al., 2001). Similarly, chemically reduced montmorillonite and ferruginous smectite are able to catalyze the abiotic reduction of nitrobenzene to aniline, and as much as 40% of the structural Fe(II) in the reduced smectite was oxidized to Fe(III) during the reaction (Yan and Bailey, 2001). Hofstetter et al. (2006) investigated the reactivity of Fe(II) species in suspensions of chemically reduced montmorillonite, with a low structural iron content, and reduced iron-rich nontronite
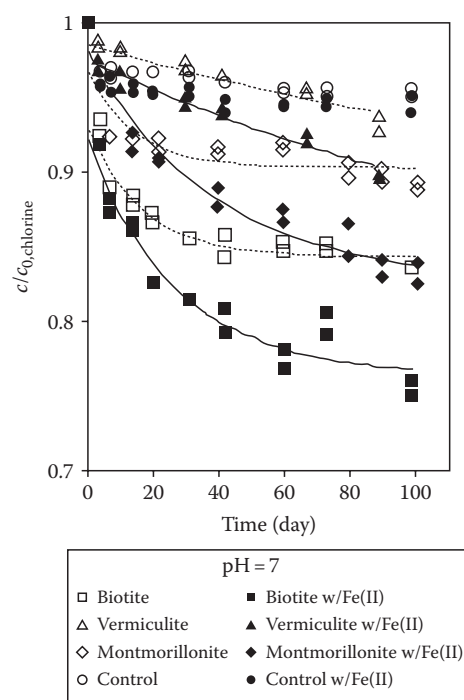
FIGURE 18.10 Reductive dechlorination of chlorinated ethylenes (0.19 mM) in iron-bearing phyllosilicate suspensions (0.085 g g^{-1}) with and without Fe(II) addition (4.28 mM). (Reprinted from Lee, W.J., and B. Batchelor. 2004. Abiotic reductive dechlorination of chlorinated ethylenes by iron-bearing phyllosilicates. Chemosphere 56:999–1009. Copyright 2004, with permission from Elsevier.) C_0 and C stand for the concentration (mM) of the organic chlorine compound at time 0 and the sampling time.

using acetylnitrobenzene compounds. Their results indicate that Fe(II) bound in the octahedral sheet of reduced smectites is the predominant reductant, and that electron transfer presumably occurs via basal siloxane planes. Nonreactive electron donor–acceptor complexation of the nitroaromatic compounds occurred at the basal smectite surfaces of the reduced montmorillonite but was not observed in suspensions of reduced iron-rich nontronite.

Reduced aquifer sediments from the perched Pantex aquifer (Amarillo, TX) were found to efficiently, abiotically degrade the explosives RDX, HMX (octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine), and TNT (Boparai et al., 2008). When Fe(II) was partially removed from the reduced sediments by washing (citrate-bicarbonate buffer), RDX degradation was suppressed, but the addition of Fe(II) restored the degradation efficiency of the sediment (Boparai et al., 2008).

18.3.5.2 Natural Organic Matter as Catalyst and Redox Mediator/Inhibitor

Humus plays a significant role in the anaerobic degradation of xenobiotics by shuttling electrons. Humic functional groups, in particular quinones, can act as redox mediators and electron donors for accelerating abiotic reductive transformation processes (Field et al., 2000; Doong and Chiang, 2005; Fimmen et al., 2007; Ratasuk and Nanny, 2007). Electrochemically reduced soil and aquatic HAs are able to reduce hexachloroethane at appreciable rates, which indicates that HA could play an important role in the reductive transformation of xenobiotics in anoxic environments (Kappler and Haderlein, 2003).

Natural organic matter can act as an electron transfer mediator in the presence of a bulk electron donor such a Fe(II) or HS⁻ and thus significantly enhances the rate of reduction of halogenated hydrocarbons, such as carbon tetrachloride or bromoform (Curtis and Reinhard, 1994) or substituted nitrobenzenes (Dunnivant et al., 1992). Strathmann and Stone (2002) demonstrated that natural organic matter from the Great Dismal Swamp facilitates oxime carbamate pesticide reduction by Fe(II) in the same manner as individual organic ligands, carboxylate, and aminocarboxylate.

Complexation of Fe(II) by catechol and organothiol ligands, which are similar to the functional groups found in natural humic substances, leads to the formation of highly reactive aqueous species that are capable of reducing substituted nitroaromatic compounds (Naka et al., 2006, 2008). The nitroaromatic compounds are not reduced in Fe(II)-only or ligand-only solutions (Naka et al., 2006, 2008). The high reactivity of these Fe(II)–organic ligand complexes is attributed to a lowering of standard one-electron reduction potential of the Fe(III)/Fe(II) redox couple on complexation by the ligands. The organothiol ligands are also able to re-reduce the Fe(III) that forms when Fe(II) complexes are oxidized by reactions with the nitroaromatic compounds (Naka et al., 2008). Ferrous iron–catechol and ferrous iron–organothiol complexes are also able to completely reductively degrade the explosive RDX (Kim and Strathmann, 2007) and to catalyze the reductive degradation of polyhalogenated alkanes

(Bussan and Strathmann, 2007). Iron(II)–tiron complexes are able to quantitatively reduce trichloroethane to acetaldehyde, a product previously only reported for reactions with Cr(II), but not with Fe-based reductants (Bussan and Strathmann, 2007).

The explosive RDX is rapidly reductively degraded in the presence of black carbon (BC) and sulfide to low-molecular-weight ring-cleavage products (Kemper et al., 2008). RDX was not degraded in the presence of BC-only or sulfide-only suspensions. Due to the heterogeneous nature of BC, it remains unclear whether the quinone material in the BC or the ability of BC to sorb the hydrophobic RDX facilitated the reductive degradation (Kemper et al., 2008).

On the other hand, natural organic matter can decrease the reductive potential of Fe(II) by competing for reactive sites on Fe(II)-mineral surfaces. Hakala et al. (2007) found that dissolved organic matter only enhances the reductive capacity of Fe(II) toward nitroaromatic compounds in the absence of iron oxide colloids. Colon et al. (2008) found that Suwannee River HA decreases the reduction capacity of Fe(II)-treated goethite toward nitrobenzene, by complexing the surface-associated Fe(II) species or hindering access of nitrobenzenes to the Fe(II) sites.

18.3.6 Genotoxicity and Bioavailability of Xenobiotics as Influenced by Mineral Catalysis

Much is known about the toxicity of xenobiotic parent compounds, whereas, relatively little work has been carried out on investigating the toxicity of the degradation products of xenobiotics (Sinclair and Boxall, 2003). In some cases, the transformation products of xenobiotics, such as commonly used pesticides and herbicides, can be even more toxic than the parent compound (Osano et al., 2002; Sinclair and Boxall, 2003). To date, very little research has been carried out on investigating the bioavailability and toxicity of xenobiotics after reaction with abiotic catalysts, such as metal oxides or clay minerals.

The oxidative coupling of the polycyclic aromatic hydrocarbon, phenanthrene, with catechol as catalyzed by birnessite significantly reduces the bioavailability of phenanthrene to microorganisms (Russo et al., 2005). The oxidative coupling of phenol induced by birnessite or soil significantly decreases the toxicity of the compound as determined with a Microtox System (Jung et al., 2008b).

The pesticides, oxamyl (carbamate family) and alachlor (chloroacetanilide family), reacted with reduced nontronite clay (ferruginous smectite) show a significant decrease in their overall mammalian cell cytotoxic (toxicity to cells) potential, whereas the pesticide 2,4-D (phenoxy family) shows no difference in toxicity after treatment with the reduced clay (Sorensen et al., 2004). The pesticide, dicamba (benzoic acid family), reacted with reduced nontronite clay actually generates products, which are more cytotoxic than the parent compound (Sorensen et al., 2004). Oxidized nontronite had no effect on the cytotoxicity of any of the aforementioned pesticides, which indicates reduced

Fe in smectites plays an important role in altering the cytotoxicity of pesticides and should thus be taken into account in pesticide management programs (Sorensen et al., 2004).

The genotoxic potency (potential damaging effect on DNA) of most pesticides in mammalian cells is unknown, let alone their abiotic transformation products. Sorensen et al. (2005) were the first group to investigate the effect of reduced ferruginous smectite on the genotoxicity of the pesticides 2,4-D, dicamba and oxamyl. Neither 2,4-D alone nor its transformation products after reaction with the reduced clay were found to be genotoxic. The genotoxicity of oxamyl actually decreased after reaction with the reduced clay. Dicamba alone was not genotoxic but became genotoxic after reaction with the reduced clay.

18.4 Abiotic Catalysis in the Transformation of Metals, Metalloids, and Other Inorganics

18.4.1 Transformation of Metals and Metalloids

18.4.1.1 Arsenic

Arsenic bioavailability and toxicology depend on its chemical state (Huang and Fujii, 1996) with As(III) (arsenite) being much more toxic than As(V) (arsenate). Manganese(IV) oxides, such as birnessite (Table 18.5), are very effective oxidants of As(III) (Oscarson et al., 1981a, 1981c, 1983b; Feng et al., 2006b). In the control experiment, no detectable As(III) is oxidized in the absence of Mn(IV) oxide (Table 18.5). Arsenic(III) oxidation by birnessite proceeds through two single electron transfer steps with the formation of Mn(III) as intermediate (MnOOH), as follows (Nesbitt et al., 1998; Tournassat et al., 2002):

$$2MnO_2 + H_3AsO_3 = 2MnOOH + H_3AsO_4, \qquad (18.12)$$

$$2MnOOH + H_3AsO_3 = 2MnO + H_3AsO_4 + H_2O. \qquad (18.13)$$

Poorly crystalline Fe(III) oxyhydroxides (Devitre et al., 1991; Nicholas et al., 2003; Jang and Dempsey, 2008) and the surfaces

FIGURE 18.11 Oxidation of As(III) to As(V) and the sorption of As by lake sediments as a function of time. (a) Concentration of As(III) in solution, (b) concentration of As(V) in solution, and (c) As(III)/As(III) + As(V). Ten μg mL^{-1} of As(III) were added initially. During the reaction period, the pH of the As-sediment suspension ranged from 8.0 to 8.2 for the Buffalo Pound sediment and from 7.3 to 7.6 for the other four sediments. (Reprinted from Oscarson, D.W., P.M. Huang, and W.K. Liaw. 1980. The oxidation of arsenite by aquatic sediments. J. Environ. Qual. 9:700–703. With permission of the American Society of Agronomy.)

TABLE 18.5 Oxidation of As(III) and Sorption of As by Mn(IV) Oxide (in μg mL^{-1} in Solution)

As(III) or As(V) Added (μg mL^{-1})	As(III)	As(V)	Mn	Final pH
100 As(III)	ND[a]	83.5 ± 1.4[b]	0.41 ± 0.12	7.1
300 As(III)	63.2 ± 7.0	186 ± 5	8.08 ± 0.36	7.1
500 As(III)	213 ± 4	205 ± 1	6.06 ± 0.60	7.3
1000 As(III)	665 ± 5	216 ± 4	4.16 ± 0.86	7.5
300 As(V)	ND	298 ± 1	0.06 ± 0.02	7.5

Source: Oscarson, D.W., P.M. Huang, C. Defosse, and A. Herbillon. 1981a. Oxidative power of Mn(IV) and Fe(III) oxides with respect to As(III) in terrestrial and aquatic environments. Nature 291:50–51. With permission of Macmillan, London.

[a] ND, not detectable.

[b] Mean ± SD; *n* = 3.

of the clay minerals illite and kaolinite (Manning and Goldberg, 1997) are also able to catalyze the oxidation of As(III).

Although As(V) is a thermodynamically stable species in oxygenated water at common pH values (Penrose, 1974), the kinetics of oxidation of As(III) with O$_2$ is very slow at near-neutral pH values (Kolthoff, 1921). Lake sediments from Saskatchewan, Canada, can oxidize As(III) (700 μg As) to As(V) within 48 h (Figure 18.11). The oxidation of As(III) is not detectable within 72 h in the absence of sediment. The oxidation of As(III) to As(V) is not affected by flushing N$_2$ or air through the sediment suspensions, nor does the addition of HgCl$_2$ to the system eliminate the conversion of As(III) to As(V). This indicates that the oxidation of As(III) to As(V) is an abiotic process.

When the lake sediments are treated with hydroxylamine hydrochloride or sodium acetate, which are effective extractants for Mn, the oxidation of As(III) to As(V) by the treated sediments is greatly decreased relative to the untreated sediments

(Oscarson et al., 1981c). Hydroxylamine hydrochloride treatment also removes Fe oxide; however, the evidence obtained from colorimetry and x-ray photoelectron spectroscopy shows that a redox reaction between Fe oxide and As(III) does not occur within 72 h, indicating that the redox reaction between As(III) and Fe(III) is kinetically slow (Oscarson et al., 1981a). This supports the evidence that Mn oxide is the primary component responsible for catalyzing the conversion of As(III) to As(V). The rate constant increases with increasing temperature from 278 to 298 K; the heat of activation for the process varies from 13.8 to 35.6 kJ mol⁻¹, indicating that the depletion of As(III) is predominantly a diffusion-controlled process (Oscarson et al., 1981b). Other recent studies have also suggested that Mn oxides are largely responsible for the oxidation of As(III) in aquifer sediments from Maine, USA (Amirbahman et al., 2006) and Dhaka, Bangladesh (Stollenwerk et al., 2007).

The ability of Mn oxides to deplete As(III) from solution (oxidation plus sorption) varies with their crystallinity and specific surface (Oscarson et al., 1983b). The depletion of As(III) by Mn oxides follows the first-order kinetics. Pyrolusite is highly ordered and has a low specific surface; conversely, birnessite and cryptomelane are poorly crystalline and have relatively high specific surfaces (Table 18.6). Because birnessite and cryptomelane have relatively high specific surface areas due to their porous nature (Huang, 1991b), their rate constants for the depletion of As(III) are much higher than those for pyrolusite (Table 18.6). On the other hand, the rate constants for the depletion of As(III) by birnessite are significantly greater than those for cryptomelane, despite its higher specific surface. Birnessite has a greater negative charge density than cryptomelane at pH 7 (McKenzie, 1981). Because As(V) is also negatively charged at pH 7, the repulsive interaction energy would be greater between birnessite than cryptomelane and As(V), which is why birnessite does not sorb a detectable amount of As(V) relative to cryptomelane (Oscarson et al., 1983b). Differences in the point of zero charge (p.z.c.) of birnessite and cryptomelane and their ability to sorb As(V) explain the greater As(III) depletion by birnessite than by cryptomelane even though cryptomelane has the greater surface area. Little As(V) is sorbed from solution by birnessite upon oxidation of As(III), and less total As is sorbed by birnessite than by cryptomelane; the electron-accepting sites on the surface of birnessite are, thus, blocked to a lesser extent than those on the surface

of cryptomelane. Consequently, the rate of depletion of As(III) from solution is greater for birnessite than for cryptomelane.

The structure and composition of Mn oxides, as well as the solution pH, ionic strength, and presence of organic ligands all affect the rate of As(III) oxidation by Mn oxides. Tunnel-structured cryptomelane has the strongest As(III) oxidation ability (842.2 mmol kg⁻¹), compared with layered–structured birnessite (480.4 mmol kg⁻¹), and Mn(II/III)-bearing hausmannite (117.9 mmol kg⁻¹) (Feng et al., 2006b). Increasing the reaction system pH from 3.0 to 6.5 results in a significant increase in the rate of oxidation of As(III) by the three Mn oxides, whereas increasing the ionic strength results in a decrease in the oxidation of As(III) by birnessite and cryptomelane, while hausmannite is unaffected (Feng et al., 2006b). The presence of tartaric acid at a concentration below 4 mol L⁻¹ promotes oxidation of As(III) by birnessite, cryptomelane, and hausmannite, whereas it decreases As(III) oxidation by cryptomelane and hausmannite when the concentration is increased above 4 mol L⁻¹ (Feng et al., 2006b). The oxidation of As(III) by the Mn oxyhydroxide, manganite [MnO(OH)], occurs on the time scale of hours. The presence of competing anions such as phosphate decreases the oxidation rate of As(III) to As(V) by manganite (Chiu and Hering, 2000). The presence of a divalent cation, Zn^{2+}, significantly suppresses the oxidation of As(III) by birnessite, and this suppression is more pronounced when the Zn^{2+} is presorbed on the birnessite (Power et al., 2005).

The surfaces of many soil Mn oxides are coated with various chemical species (McKenzie, 1989). The rate constants for the depletion (oxidation plus sorption) of As(III) are generally substantially smaller for the Mn oxides with higher levels of coatings of Fe and Al oxides and $CaCO_3$ than those for the untreated MnO_2 or MnO_2 with lower levels of coatings (Oscarson et al., 1983a). HA coatings on natural Mn nodules also significantly reduce the As(III) depletion capacity of the nodules (Chen et al., 2006). The electron-accepting sites on MnO_2 are partially masked by the oxides, $CaCO_3$, and organic matter. There is evidence that shows that Fe(III) and Al oxides and $CaCO_3$ do not oxidize As(III) to As(V) (Oscarson et al., 1981a, 1983a), and thus, the oxidation of As(III) to As(V) on the coated MnO_2 is solely attributed to MnO_2 in this case.

However, some studies have shown that Fe oxides and oxyhydroxides can catalyze the oxidation of As(III) to As(IV). Devitre et al. (1991) showed that As(III) is rapidly oxidized by diagenetic

TABLE 18.6 Specific Surface and Point of Zero Charge (p.z.c.) of the Mn Dioxides and Rate Constants and Energies of Activation for the Depletion of As(III) by the Mn Dioxides

Mineral	Specific Surface (m² g⁻¹)	p.z.c.	Rate Constant × 10⁻³ (h⁻¹)			Energies of Activation (kJ mol⁻¹)
			278 K	298 K	318 K	
Birnessite	277 ± 5[a]	2.3 ± 0.1	126 ± 13	267 ± 6	533 ± 38	26.0 ± 0.2
Cryptomelane	346 ± 4	2.8 ± 0.1	54 ± 10	189 ± 8	318 ± 22	32.3 ± 6.7
Pyrolusite	8 ± 1	6.4 ± 0.3	0.12 ± 0.02	0.44 ± 0.03	0.58 ± 0.05	29.0 ± 9.8

Source: Oscarson, D.W., P.M. Huang, W.K. Liaw, and U.T. Hammer. 1983b. Kinetics of oxidation of arsenite by various manganese dioxides. Soil Sci. Soc. Am. J. 47:644–648. With permission of the Soil Science Society of America.

[a] Mean ± SE.

amorphous Fe(III) oxyhydroxides. The rapid oxidation of As(III) is observed on the Fe oxyhydroxide, ferrihydrite, in the presence of hydrogen peroxide (Nicholas et al., 2003). Jang and Dempsey (2008) found that the oxidation of As(III) by Fe oxyhydroxides is significant only in the presence of As(V). Magnetite (Fe_3O_4) is able to catalyze the oxidation of As(III), which is attributed to the presence of structural Fe(III) and possible Mn impurities in the magnetite (Su and Puls, 2008).

The oxidation of As(III) by the Mn oxides, birnessite, todorokite, and hausmannite, increases significantly in the presence of the Fe(III) oxide, goethite (Feng et al., 2006a). Goethite promotes As(III) oxidation on the Mn oxide minerals by adsorption of the As(V) produced during the reaction, thus decreasing the As(V) concentration in solution. Therefore, the combined effects of the Mn and Fe oxide minerals can lead to the rapid oxidation and immobilization of As in soils (Deschamps et al., 2003; Feng et al., 2006a).

The conversion of As(III) to As(V) by uncoated and coated Mn oxides has important implications for the transport, fate, and toxicity of As in terrestrial and associated environments. In some environments that have been contaminated with As(III), the addition of reactive Mn oxides may alleviate the toxicity of As(III) through their catalytic reactions.

18.4.1.2 Selenium

Selenium is an essential trace element in the human and animal nutrition; however, there is a very narrow optimal concentration range between deficiency and toxicity (selenosis). The mobility and availability of Se are largely controlled by its redox state, which affects its sorption to mineral and organic matter surfaces (Masscheleyn et al., 1990; Charlet et al., 2007). Selenium(VI) is the stable valence state in oxic environments and exists as the selenate (SeO_4^{2-}) oxyanion. Selenium(IV) is the stable valence state under mildly reducing or anoxic conditions and exists as the selenite (SeO_3^{2-}) oxyanion. Selenate is weakly sorbed by minerals and is more soluble and bioavailable than selenite, which binds strongly to the surfaces of metal oxides by inner-sphere complexes. In reducing sediments, Se(VI) and Se(IV) can be reduced to Se(0) and other reduced species (oxidation states of −1 and −2), such as Se-pyrite or ferroselite. The formation of elemental and pyritic Se is regarded as an important mechanism for controlling the availability of Se in the environment (Belzile et al., 2000). The reduction of Se(VI) and Se(IV) may be facilitated by sorption of these species on the surfaces of Fe–Mn oxyhydroxides or Fe(II)-bearing minerals or by the involvement of HS^-, pyrite, and organic matter (Myneni et al., 1997; Belzile et al., 2000; Bruggeman et al., 2005; Charlet et al., 2007; Scheinost and Charlet, 2008). Previously it was thought that these reactions are principally mediated by microorganisms (Myneni et al., 1997).

The abiotic reduction of Se(VI) and Se(IV) to Se(0) can be catalyzed by Fe(II) either sorbed on the surfaces of clays or structurally present in Fe-bearing minerals. Green rust, an Fe(II/III) oxide commonly present in reduced sediments and soils, is able to reduce Se(VI) to Se(IV) and Se(0) by adsorption
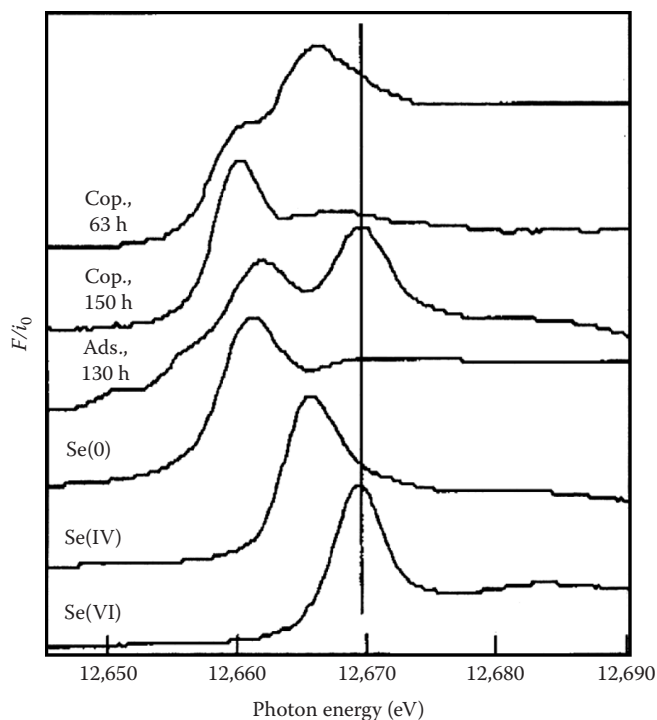
FIGURE 18.12 In situ XANES spectra of Se(VI) reaction with green rust as a function of sorption mechanisms (coprecipitation [Cop.] and adsorption [Ads]). The XANES spectra of Se models in different oxidation states are shown for comparison. The vertical line in the center shows the position of Se(VI). (Reprinted from Myneni, S.C.B., T.K. Tokunaga, and G.E. Brown. 1997. Abiotic selenium redox transformations in the presence of Fe(II,III) oxides. Science 278:1106–1109. With permission of the American Association for the Advancement of Science.)

and coprecipitation mechanisms (Figure 18.12) (Myneni et al., 1997). The reaction proceeds as follows; magnetite is produced during the oxidation of the green rust:

$$HSeO_4^- + 4Fe_4^{II}Fe_2^{III}(OH)_{12}SO_4\,3H_2O$$

$$\leftrightarrow HSe^- + 8Fe_3O_4 + 4SO_4^{2-} + 8H^+ + 32H_2O. \qquad (18.14)$$
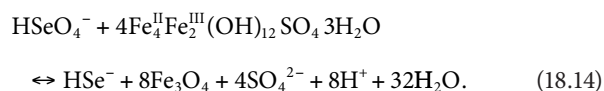
Similarly, zero-valent iron reacted with Se(VI) catalyzes the formation of Se(0) and produces a green rust layer on the exposed surfaces of the iron(0) (Scheidegger et al., 2003).

The reduction of selenite [Se(IV)] to solid Se(0) can be catalyzed by Fe(II)-bearing mineral, pyrite (FeS_2), via the process of adsorption on the surface of pyrite (Bruggeman et al., 2005; Breynaert et al., 2008). The presence of illite-rich Boom clay sediments (Belgium) and natural humic substances decreases the rate of selenite reduction by pyrite due to the competitive adsorption (Bruggeman et al., 2005). Troilite (FeS) is also able to catalyze the reduction of selenite, however, through a different mechanism that involves reduction by dissolved sulfide and the formation of Se(0) as an intermediate (Breynaert et al., 2008). The Se(0) is then further reduced to $FeSe_x$ species, which precipitate on the surface of the troilite. Nanoparticulate Fe(II)-bearing

minerals mackinawite [(Fe, Ni)$_{1+x}$S$_x$] and magnetite (Fe$_3$O$_4$), and micrometer-sized siderite (FeCO$_3$) are able to rapidly reduce selenite (Scheinost and Charlet, 2008). Mackinawite reduces selenite to Se(0) at pH 6.3 and FeSe at pH 4, whereas magnetite reduces selenite to FeSe and Fe$_7$Se$_8$ species at pH 5. Siderite is weakest catalyst of three minerals and only partially reduces selenite to Se(0) (Scheinost and Charlet, 2008).

Montmorillonite equilibrated with Fe^{2+} is able to catalyze the slow reduction of selenite [Se(IV)] to nanoparticulate Se(0) (Charlet et al., 2007). Mossbauer and XAS spectroscopy suggest that the Se and Fe redox reactions are not directly coupled, but rather that a small amount of Fe(II) that is presorbed on the montmorillonite is oxidized to Fe(III), and the electrons that are produced during this reaction are stored by the formation of surface H$_2$ species, which then later are able to reduce selenite (Charlet et al., 2007).

18.4.1.3 Iron

Manganese oxides, which have different structural and surface properties, vary in their ability to catalyze the conversion of Fe(II) to Fe(III) and the subsequent formation of Fe oxides and oxyhydroxides (Krishnamurti and Huang, 1987, 1988). The standard electrode potential (E°) of the redox pairs Fe^{2+}–MnO$_2$ and Fe^{2+}–Mn$_3$O$_4$ can be described by the following equations (Bricker, 1965):

$$2Fe^{2+} + MnO_2 + 4H^+ = Mn^{2+} + 2Fe^{3+} + 2H_2O \quad E° = +0.438\,V,$$

(18.15)

$$2Fe^{2+} + Mn_3O_4 + 8H^+ = 3Mn^{2+} + 2Fe^{3+} + 4H_2O \quad E° = +1.04\,V.$$

(18.16)
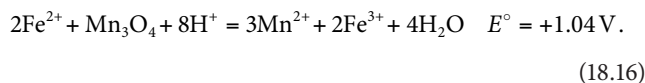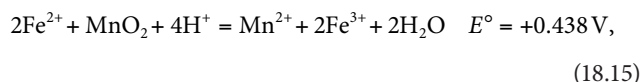
These *E*° values indicate that oxidation of Fe^{2+} by Mn oxides is thermodynamically feasible. The Eh–pH diagram (Figure 18.13) indicates the feasibility of the conversion of Mn^{4+} and Mn^{3+} to Mn^{2+} in the Mn oxides in the Eh–pH ranges of the formation of Fe oxides and oxyhydroxides. Further, the ESR spectra of the filtrates after the reaction of Fe^{2+} with Mn oxides show the presence of a significant amount of Mn^{2+}, demonstrating the reduction of Mn(IV) and Mn(III) oxides to Mn^{2+} in the presence of Fe^{2+} in solution. Simultaneously, the oxidation of Fe^{2+} to Fe^{3+} by Mn oxides leads to subsequent hydrolysis of Fe^{3+} to form a series of the precipitation products including lepidocrocite, akaganeite, feroxyhyte, and magnetite. Because various Mn oxides differ in their ability to catalyze the formation of Fe oxides and oxyhydroxides, their role in the transformation of Fe deserves attention.

At neutral pH conditions, structural Fe(III) in hematite is able to catalyze the oxidation of sorbed Fe(II) to amorphous ferric oxide on the surface of hematite (Jeon et al., 2003). Solid-phase contact among hematite, amorphous ferric oxide, and structural Fe(II) eventually produces magnetite (Fe$_3$O$_4$) (Jeon et al., 2003). Hydrous ferric oxide is able to catalyze the abiotic oxidation of
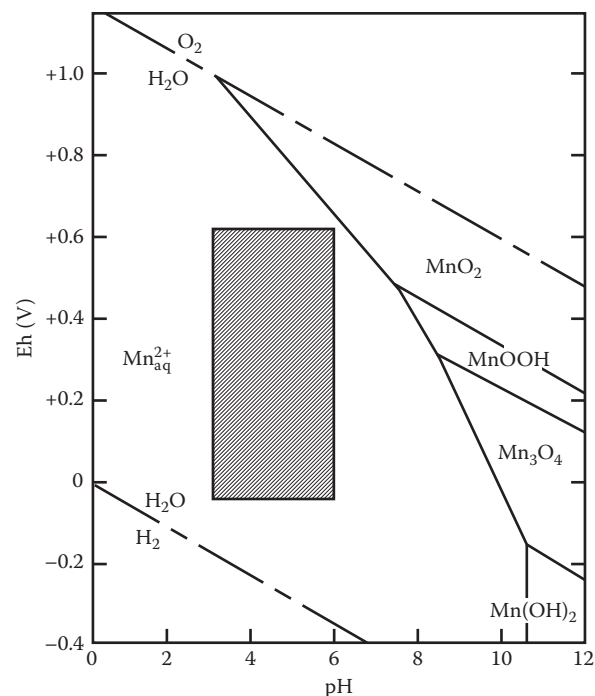
FIGURE 18.13 Stability relations of different species of Mn at 25°C, 0.101 MPa, and a$_{Mn^{2+}}$ = 10^{-6} (Bricker 1965). The Eh and pH ranges of the systems during the formation of iron oxide precipitates in the presence of Mn oxides in the present study are shown as shaded region. (Reprinted from Krishnamurti, G.S.R., and P.M. Huang. 1988. Influence of manganese oxide minerals on the formation of iron oxides. Clay. Clay Miner. 36:467–475. With permission of the Clay Minerals Society.)

Fe^{2+} under anoxic conditions at neutral pH (Park and Dempsey, 2005). During this reaction, the hydrous ferric oxide is partially converted to goethite. The reaction mechanism involves an anode/cathode mechanism with O$_2$ reduced at electron deficient sites with Fe(II) strongly sorbed, and Fe(II) is oxidized at electron-rich sites without sorbed Fe(II). Iron(III) has been found to play a dominant role in pyrite (FeS$_2$) oxidation in comparison with dissolved oxygen under acidic conditions (Mazumdar et al., 2008).

18.4.1.4 Manganese

Although oxidation of Mn in soils by atmospheric O$_2$ is a thermodynamic possibility throughout the pe and pH range common in well-aerated soils, Mn is not readily oxidized in solutions unless the pH is raised above pH 8.5 (Ross and Bartlett, 1981). A catalytic mechanism of some sort is apparently necessary. Oxidation of Mn in soil environments is generally assumed to be microbially mediated (Tebo et al., 2005; Thompson et al., 2005).

Oxidation of Mn(II) by soils has been shown to be proportional to the level of existing reactive Mn oxides (Ross and Bartlett, 1981). Arrhenius plots of rates of the oxidation between 1°C and 35°C show that the oxidation of Mn(II) by soils has nonbiological characteristics. Generally, oxidation of Mn(II) is initially rapid with no lag period and may be related to the mechanism proposed for the accumulation of Co and Zn on Mn oxide surfaces (Loganathan et al., 1977). The adsorption on

the oxide surfaces is specific and not a simple function of surface charge. Above pH 5–6, this adsorption reverses the surface charge of the Mn oxides from negative to positive, as measured by electrophoretic mobility. A positively charged surface should result in a higher OH^- concentration near the surface, which could enable oxidation of adsorbed Mn(II) by atmospheric O_2. The mechanism of specific adsorption is not fully understood, but the preferential adsorption of transition metals by Mn oxides is well documented (Jenne, 1977). The oxidation of Mn(II) by Mn oxides is theorized to be autocatalytic, involving specific adsorption of Mn(II) on existing Mn oxide surfaces (Ross and Bartlett, 1981).

The Fe(III) oxyhydroxide, akaganeite, promotes the formation of birnessite (δ-MnO_2) by catalyzing the air oxidation of surface adsorbed Mn(II) (Cornell and Giovanoli, 1991). Hematite, especially at a particle size below 4 nm, is able to promote the oxidation Mn(II) in the presence of O_2 by donating electron density to adsorbed Mn^{2+}, which promotes reaction with O_2 (Madden and Hochella, 2005).

18.4.1.5 Radionuclides

Radionuclides are unstable isotopes of elements that are radioactive. Most elements with an atomic number greater than 83 have an unstable nucleus and are thus radioactive, and therefore, all of their isotopes are radionuclides. Many radionuclides are man-made, for example, plutonium and technetium, and are released into the environment as a result of nuclear waste, airborne emissions from nuclear reactors, nuclear weapons testing and processing, and infrequently due to incorrect disposal of medical or industrial radiation sources (Salbu et al., 2004; USEPA, 2009). Most of the commonly encountered polluting radionuclides have multiple oxidation states, and thus, redox chemistry plays a dominant role in their environmental mobility and availability.

Plutonium and neptunium are the chief radionuclides of concern for long-term storage of nuclear waste due to their long half-lives ($t_{1/2} = 2.14 \times 10^6$ years for ^{237}Np and 2.41×10^4 years for ^{239}Pu), radiotoxicity, and chemotoxicity (Runde et al., 2002). Plutonium exists in the III–VI oxidation states as Pu^{3+}, Pu^{4+}, PuO_2^+, and PuO_2^{2+} in strong acid conditions. Plutonium(III/IV) cations are sorbed by soil constituents, and therefore, immobile in most aqueous and soil environments, whereas, Pu(VI) cations are not sorbed by temperate soils and thus quite mobile and bioavailable. $Pu(V)O_2^+$ can be sorbed onto soil surfaces from dilute solutions or seawater (Keeneykennicutt and Morse, 1985). Neptunium tends to occur as highly mobile $Np(V)O_2^+$ oxidation state under a wide range of environmental conditions (Wilk et al., 2005).

Manganese(III/IV) oxides and oxyhydroxides can oxidize Pu(III/IV) (Cleveland, 1970; Amacher and Baker, 1982; Morgenstern and Choppin, 2002) and Pu(V) (Keeneykennicutt and Morse, 1985; Duff et al., 1999) to the more mobile and toxic Pu(VI). Conversely, Mn(II)-bearing oxides, hausmannite and manganite, can reduce Pu(VI) to Pu(IV), which then forms strong inner-sphere complexes with the surface of the mineral

(Shaughnessy et al., 2003). Surprisingly, it has also been shown that pyrolusite (α-MnO_2), a Mn(IV)-bearing oxide, can promote the reduction of Pu(VI) (Kersting et al., 1999; Powell et al., 2006). Powell et al. (2006) attributed this to the observation that pyrolusite can initially oxidize Pu(IV) to Pu(VI); however, if the oxidized species remains associated with the solid phase, the initial oxidation step is followed by reduction to a stable hydrolyzed Pu(IV) species, which is not affected by the oxidizing surface, and over time becomes the predominant solid-phase Pu species. These findings are important when assessing the risk of the geological burying of nuclear waste in areas containing Mn(IV)-bearing minerals, such as Yucca Mountain or the Hanford sites (Powell et al., 2006).

Iron(II)-bearing minerals can reduce highly mobile Np(V) and Pu(V/VI) species, as well as, many other mobile higher valency species of radionuclides such as uranium and technetium. Natural uranium is mildly radioactive and occurs in trace amounts in most rocks, soils, and water. The more strongly radioactive isotope, U-235, is selectively enriched and used for nuclear weapons and for fuel for nuclear reactors. The U-235-depleted uranium is used for making weapons and armor, and has been used for counter weights in aircraft tails. All the above-mentioned uses result in the release of U into the environment. Similar to Pu and Np, U has a variety of oxidation states with uranyl $U(VI)O_2^{2+}$ species being the most mobile. Technetium is an artificial element produced during nuclear fission and is commonly found in groundwater at sites where nuclear waste has been processed or stored. It has a similar chemistry to Mn and has several oxidation states. Under oxic conditions, it occurs as a pertechnetate anion, $Tc(VII)O_4^-$, which is poorly sorbed across the environmental pH range. The lower valency Tc(IV) readily precipitates as insoluble hydrous oxides or is sorbed on mineral surfaces.

Magnetite (Fe_3O_4) can reduce Pu(V) to strongly sorbed Pu(IV) over a wide pH range (pH 3–8; Powell et al., 2004). Similarly, magnetite can reduce Tc(VII) to Tc(IV) (Cui and Eriksen, 1996), Np(V) to Np(IV) (Nakata et al., 2004), and U(VI) to U(IV) (Scott et al., 2005; El Aamrani et al., 2007). Biotite, which contains significant amounts of Fe(II), is also able to reduce U(VI) on exposed edge sites over a broad pH range (Ilton et al., 2004).

Green rusts, which are mixed Fe(II/III) hydrous oxides, are potent reducing agents in suboxic environments and can readily reduce U(VI) to U(IV) to precipitate nanoparticular UO_2 (O'Loughlin et al., 2003a) and reduce Tc(VII) pertechnetate anions to strongly surface-complexed Tc(IV) (Pepper et al., 2003). Likewise, amorphous FeS can reduce U(VI) to U(V) and U(IV) oxides (Hua and Deng, 2008). Mackinawite (FeS) is able to reduce Tc(VII) to form $Tc(IV)S_2$, U(VI) to U(V/IV) oxide phases, and Np(V) to Np(IV) surface complexes (Livens et al., 2004).

Fe(II) sorbed to the surface of Fe(III)-containing oxides, forming the reactive surface species $\equiv Fe^{III}O$–$Fe^{II}OH^0$, is also a potent reducing agent of radionuclides. In contrast to the abiotic reduction of organics (e.g., nitrobenzenes) by Fe(II) sorbed to hematite or magnetite, which is an outer-sphere mechanism, the

reduction of inorganics such as U(VI) involves an inner-sphere electron transfer mechanism (Charlet et al., 1998; Liger et al., 1999). The first step is the adsorption of U(VI) ion by the reactive surface species and the formation of an inner-sphere complex, followed by two reductive steps, which lead to the formation of a mixed $U(IV)O_2/Fe(OH)_3$ solid phase (Charlet et al., 1998; Liger et al., 1999). The reduction of U(VI) by Fe(II) sorbed to synthetic Fe oxides (magnetite, goethite, and hematite) and natural Fe(III) oxide-containing sediments has been demonstrated, and the natural sediments are less efficient than the pure oxides due to their lower Fe(II) sorption capacity (Jeon et al., 2005). Iron(II) sorbed to hematite and goethite is much more efficient Tc(VII)-reducing agent than Fe(II) sorbed to phyllosilicates (vermiculite, illite, and muscovite) or structural Fe(II) in the phyllosilicates (Peretyazhko et al., 2008a). Iron(II) sorbed to Al oxyhydroxide, diaspore, and Al oxide, corundum, is also able to reduce Tc(VII) to insoluble Tc(IV) (Peretyazhko et al., 2008b).

Natural humic substances and quinoid-enriched humic substances are able to catalyze the abiotic reduction of Pu(V) to insoluble Pu(IV) over a wide pH range and under oxic and anoxic conditions (Andre and Choppin, 2000; Shcherbina et al., 2007a). Humic substances are only able to reduce Np(V) to Np(IV) under acidic, anoxic conditions (Shcherbina et al., 2007a, 2007b).

18.4.1.6 Other Trace Metals

In addition to oxidation of metalloids such as As(III) and metals such as Fe(II), Mn(II), Mn oxides, and oxyhydroxides can catalyze the oxidation of trace metals by disproportionation of Mn^{2+} and MnO_2. The disproportionation facilitates electron transfer processes that can either greatly decrease or increase the equilibrium solubility of certain metals (Hem, 1978).

Chromium and Pu are similar in chemical behavior in aqueous environments (Rai and Serne, 1977; Bartlett and James, 1979). Chromium occurs in the II, III, and VI oxidation states in water. Chromium(II) is unstable. Chromium(III) has broad stability and exists as the cation Cr^{3+} and its hydrolysis products or as the anion CrO_2^-. Chromium(VI) exists under strongly oxidizing conditions, occurs as dichromate $Cr_2O_7^{2-}$, or chromate $HCrO_4^-$ and CrO_4^{2-} anions. Chromium(III) cations are readily sorbed to soil constituents, and therefore, immobile in most aqueous and soil environments; however, Cr(VI) is not sorbed by temperate soils to any extent and is thus quite mobile in soils and associated aqueous environments. Therefore, hexavalent Cr is readily bioavailable and extremely toxic (Amacher and Baker, 1982) and, thus, of concern in food-chain contamination and ecosystem health. Manganese(III/IV) oxides can oxidize Cr(III) and are the only known oxidizers of Cr(III) in the soil environment (Stepniewska et al., 2004; Negra et al., 2005).

The kinetics of Cr(III) oxidation by Mn oxide is very rapid with most of the conversion of Cr(III) to Cr(VI) occurring during the first hour (Amacher and Baker, 1982). The oxidation of Cr(III) increases with increasing temperature. Manganite (γ-MnOOH) is also able to catalyze the oxidation of Cr(III), and the reaction rate increases under acidic conditions (Weaver et al., 2002). At higher

pH, Cr(III) hydroxy hydrate precipitates on the manganite surface, and the oxidation reactions are minimized (Weaver et al., 2002). Feng et al. (2007) compared the Cr(III) oxidative power of four different Mn oxides and found that order of reactivity was as follows (mmol Cr(III) kg^{-1} oxide): birnessite (1330.0) > cryptomelane (422.6) > todorokite (59.7) > hausmannite (36.6). The Cr(III) oxidative ability of the Mn oxides depends on the crystallinity, structure, composition, and surface properties of the minerals (Feng et al., 2007).

The kinetics of Cr(III) oxidation on Mn(III/IV) oxides shows a trend similar to that observed for soils (Amacher and Baker, 1982). Negra et al. (2005) demonstrated that greater Mn(IV) oxide abundance in soils enhances the Cr(III) oxidation potential. It has also been shown that the oxidation state of Cr in soils in the long-term depends on balances between organic carbon availability (enhances Cr reduction) and the Mn redox status [Mn(III)/(IV) oxides enhance oxidation] (Stepniewska et al., 2004; Tokunaga et al., 2007). Nonliving biomaterials, such as seaweed, dead fungi, sawdust and pine bark, and needles, are able to abiotically (nonenzymatically) catalyze the adsorption-coupled reduction of Cr(VI) to Cr(III) (Park et al., 2004, 2007, 2008b). The Cr(VI)-reducing capacity of pine bark is 8.7 times higher than that of a common chemical Cr(VI) reductant, $FeSO_4 \cdot 7H_2O$. When Cr(VI) comes in contact with biomaterials, there are two possible reduction mechanisms (Figure 18.14) (Park et al., 2005, 2007). Mechanism I, the direct reduction mechanism, involves the direct reduction of Cr(VI) to Cr(III) by contact with electron-donor groups of the biomaterial, such as amino and carboxyl groups (Figure 18.14). Mechanism II, the indirect reduction mechanism, involves three steps: (1) the binding of anionic Cr(VI) to the positively charged groups present on the biomaterial surface, (2) the reduction of Cr(VI) to Cr(III) by adjacent electron-donor groups, and (3) the release of the reduced Cr(III) into the aqueous phase due to electronic repulsion between the positively charged groups and the Cr(III), or the complexation of the reduced Cr(III) with adjacent groups (Figure 18.14). A lower pH makes the biosurface more positively charged and accelerates redox reactions, which enhances Cr(VI) removal (Park et al., 2007). The addition of soil slurry to nonliving biomaterials, such as pine bark, significantly enhances the Cr(VI)-reducing capability of the pine bark, by sorbing Cr(III) from solution (Park et al., 2008a). Therefore, although Mn oxides can enhance the mobility and toxicity of Cr in soils and related environments, the presence of organic matter decreases its oxidation potential.

Manganese oxides and oxyhydroxides can also catalyze the oxidation of other trace metals such as Co, Pb, Ni, Cu (Hem, 1978), and Sb (Belzile et al., 2001). The catalytic transformation of these metal ions may be directly influenced by redox processes coupled to disproportionation of Mn mixed valence oxide, to catalyzed oxidation by aqueous O_2, or to other redox reactions involving changes from one Mn oxide species to another. When the oxidized form of the element has a lower solubility than the reduced form, this effect can be of major significance.

X-ray photoelectron spectroscopy measurements of Co adsorbed on MnO_2 reveal strong evidence that Co(II) is oxidized
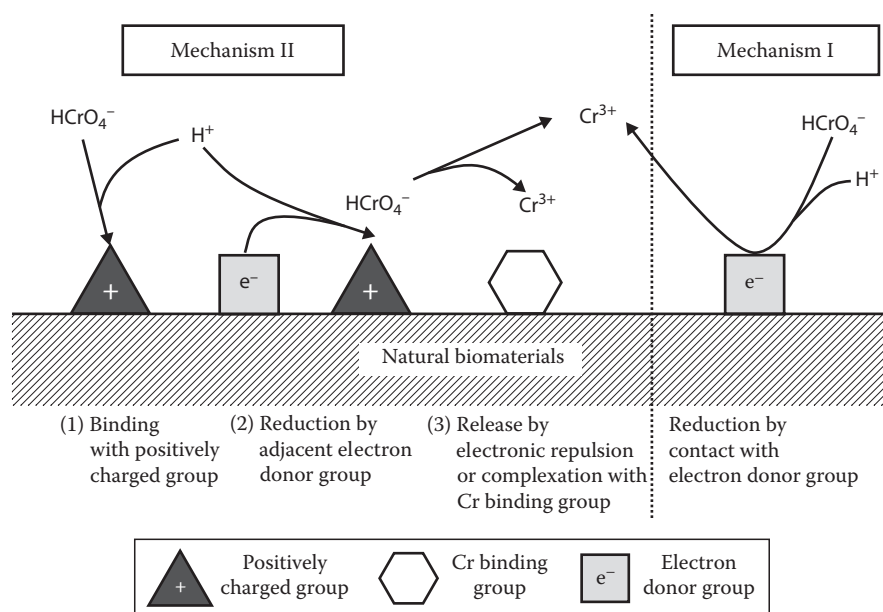
FIGURE 18.14 Proposed direct and indirect mechanisms for Cr(VI) biosorption and reduction by natural biomaterials. (Reprinted from Park, D., Y.S. Yun, and J.M. Park. 2005. Studies on hexavalent chromium biosorption by chemically-treated biomass of *Ecklonia* sp. Chemosphere 60:1356–1364. Copyright 2005, with permission from Elsevier.)

to Co(III) in the presence of the strong electric field at the MnO$_2$–solution interface (Murray and Dillard, 1979). Nickel(II), however, cannot be oxidized at the interface except at very high concentrations. Strong experimental evidence for the oxidation of other trace metals catalyzed by Mn oxides still remains to be attained. Further, more information on the kinetics and mechanisms of redox reactions of these trace metals on the surfaces of Mn oxides is needed.

Other than Mn oxides, heterogeneous oxidation/reduction reactions involving electron transfer between transition metals and Fe-containing minerals have been investigated in numerous studies. Vanadium(II) and V(IV) can be oxidized by Fe(III) oxyhydroxides (Wehrli and Stumm, 1989). Amorphous Fe(III) oxyhydroxides present in natural waters and sediments can also oxidize the more toxic Sb(III) into Sb(V) (Belzile et al., 2001). The oxidation pseudo-first-order rate of Sb(III) by natural Fe oxyhydroxides is slower compared to synthetic Fe oxyhydroxides. This is attributed to the slight crystallinity of the natural oxides (compared to amorphous synthetic compounds) and to their more complex chemical composition, which can include adsorbed ions and organic matter (Belzile et al., 2001). Goethite has also been shown to catalyze the oxidation of Sb(III) to Sb(V) (Leuz et al., 2006).

Iron(II)-bearing minerals are able to reduce a variety of trace metals. Reduction of Cr(VI) by biotite has been shown by Eary and Rai (1989) and Ilton and Veblen (1994). Magnetite and ilmenite are the most common Fe(II)-containing oxide minerals in the earth's crust and potentially important in controlling heterogeneous redox reactions involving aqueous transition metals in soil environments. Direct evidence of Cr(VI) reduction on magnetite surfaces has been documented by Peterson et al. (1996) using x-ray adsorption fine structure (EXAFS)

spectroscopy. Structural Fe(II) in magnetite and ilmenite heterogeneously reduces aqueous Cu(II), V(IV), and Cr(VI) ions at the oxide surfaces over a pH range of 1–7 at 25°C (White and Peterson, 1996). Calcium carbonate coatings on the surface of magnetite reduce or eliminate its ability to reduce Cr(VI) (Doyle et al., 2004). Siderite (FeCO$_3$) is also able to reduce aqueous Cr(VI), and its reactivity increases with addition of acid and with increasing temperature (Erdem et al., 2004). Pyrite is also an effective reductant of Cr(VI), and the reaction is about 100 times faster than reduction by biotite, which is related to the relative amount and dissolution rate of Fe(II) from the minerals (Chon et al., 2006). Similarly, Lin and Huang (2008) investigated the kinetics and effect of solution characteristics on Cr(VI) reduction by pyrite under dark and anaerobic conditions and found that the reduction is highly dependent on the solution pH, which influences the dissolution of Fe(II) and sulfide ions from pyrite, which are responsible for the reduction of Cr(VI). Magnetite (within days) and mackinawite (within minutes) are able to reduce Sb(V) to the more toxic Sb(III) form (Kirsch et al., 2008). The Sb(V) reduction by mackinawite proceeds solely by oxidation of surface Fe(II), while the oxidation state of sulfide is conserved, with the formation of amorphous or nanoparticulate SbS$_3$-like solids (Kirsch et al., 2008).

The Fe(II)-rich chlorite and corrensite clays are able to reduce Cr(VI) under acidic conditions, and during the reaction, structural Fe(II) is oxidized to Fe(III) (Brigatti et al., 2000). Mixed Fe(II)/Fe(III) hydroxides, known as green rusts, commonly found in suboxic environments are able to readily reduce a variety of trace metals, such as Ag(I), Au(III), Cu(II), and Hg(II) to Ag(0), Au(0), Cu(0), and Hg(0) (O'Loughlin et al., 2003b). Sulfate green rusts are able to effectively reduce Cr(VI) to Cr(III), which

initially involves chromate anions replacing sulfate in the interlayer of the green rust, and then the chromate is reduced by Fe(II) (Skovbjerg et al., 2006). Even though the formation of sparing soluble Cr(III) blocks further chromate entry, Cr(VI) reduction continues at the green rust solid/solution interface due to electron transfer from the center of the green rust crystals. During this process, Cr-substituted goethite is formed, which is sparingly soluble and thus essentially locks away the Cr, which has important implications for environmental remediation (Skovbjerg et al., 2006). Sulfate green rust is also able to reduce the highly stable Sb(V) to the more toxic Sb(III) (Mitsunobu et al., 2008).

Iron(II) sorbed to the surface of Fe oxides and clay minerals are also able to catalyze the reduction of Cr(VI) to Cr(III). Buerge and Hug (1999) compared the Cr(VI) reduction reactivity of Fe(II) in the presence of aluminum oxide, silicon dioxide, goethite, lepidocrocite, montmorillonite, and kaolinite and found that only Al_2O_3 did not accelerate the reduction reaction in the presence of Fe(II). The order of reactivity was as follows (Buerge and Hug, 1999): goethite ≈ lepidocrocite ≫ montmorillonite > kaolinite ≈ SiO_2 ≫ Al_2O_3. Similarly, Fe(II) added to magnetite significantly improves the Cr(VI) reductive ability of magnetite (Jung et al., 2007).

Organic complexes of Fe(II) also strongly affect the reduction of Cr(VI) to Cr(III) (Buerge and Hug, 1998). Fe(III)-stabilizing ligands such as bi- and multidentate carboxylates and phenolates generally accelerate the reduction reaction, whereas Fe(II)-stabilizing ligands such as phenanthroline essentially stop the reaction. Natural dissolved organic matter from a forest soil shows quantitatively the same accelerating behavior as the investigated Fe(II)-carboxylate complexes (Buerge and Hug, 1998).

18.4.2 Transformation of Other Inorganics

Nitrate formation in soils from NO_2^- is related to soil level of reactive Mn oxides, which catalyze the oxidation (Bartlett, 1981). Nitrite oxidation, NO_3^- formation, and MnO_2 reduction are stoichiometrically related reactions both in the presence and the absence of atmospheric O_2. The relationship shown in Figure 18.15 suggests that at high ratios of MnO_2/NO_2^-, most of the Mn(IV) is reduced only to nonexchangeable solid Mn(III) oxide rather than to exchangeable Mn(II) ions. When the MnO_2/NO_2^- ratio is high, little Mn(II) becomes exchangeable relative to the NO_3^- formed, as indicated by high values of $NO_3^-/Mn(II)$. At low MnO_2/NO_2^- ratios, more exchangeable Mn(II) relative to NO_3^- is formed. Therefore, the more MnO_2 relative to NO_2^-, the greater is the efficiency of the NO_2^- to NO_3^- transformation and the less exchangeable Mn(II) formed relative to NO_3^-. High MnO_2/NO_2^- ratios likely prevail in many soils. Equation 18.17, which does not involve a pH change, may describe the NO_2^- to NO_3^- transformation in high Mn oxide soils better than Equation 18.18:

$$NO_2^- + 2MnO_2 = Mn_2O_3 + NO_3^-, \qquad (18.17)$$

$$NO_2^- + MnO_2 + 2H^+ = Mn^{2+} + NO_3^- + H_2O. \qquad (18.18)$$
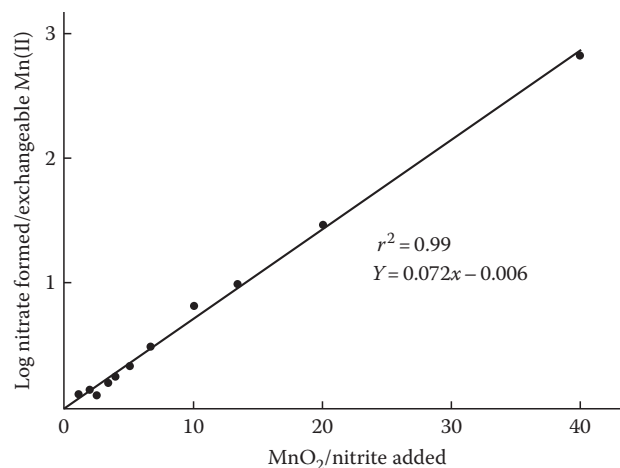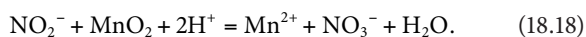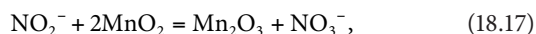
FIGURE 18.15   The logs of the ratios of nitrate formed from nitrite over Mn reduced to the exchangeable form plotted against the MnO_2/nitrite ratios in the original suspensions. (Reprinted from Bartlett, R.J. 1981. Nonmicrobial nitrite-to-nitrate transformation in soils. Soil Sci. Soc. Am. J. 45:1054–1058. With permission of the Soil Science Society of America.)

Such a process of substantial reduction of Mn in the solid phase relative to the release of exchangeable Mn(II) might be described better as the insertion of electrons into the overlapping electronic orbitals of the solid, such that the electron excess is delocalized over the solid, rather than the reduction of one Mn^{4+} atom to Mn^{2+} at a particular surface site. The overlapping atomic orbitals form bands of many energy levels rather than a single energy level.

Luther and Popp (2002) studied the kinetics of nitrite oxidation by polymeric Mn(IV) oxides, which are highly reactive Mn oxide phases. The energy of activation and the entropy of activation show the oxidation reaction to be associative and diffusion controlled, occurring via an inner-sphere mechanism, likely with O atom transfer from MnO_2 to HNO_2. The reaction requires protons and slows down at a pH above 5.5, where NO_2^- and MnO_2 (negatively charged) become the dominant species (Luther and Popp, 2002).

Manganese(IV) oxide is also able to catalyze the oxidation of NH_3 and organic N to N_2, which has traditionally been assumed to be a microbially mediated process in sediments (Luther et al., 1997). The MnO_2-catalyzed formation of NO_3^{2-} from NH_3 can also occur, but the formation of N_2 from NH_3 is more favorable (Luther et al., 1997). Luther et al. also observed the reduction NO_3^{2-} to N_2 by dissolved Mn^{2+}. These oxidation and reduction reactions can occur in the presence or absence of O_2. Hulth et al. (1999) found that introduction of Mn oxides from oxidized surface sediments into anoxic sediments results in anoxic nitrification and the net production of NO_3^- and NO_2^-. They found that the anoxic nitrification rates were directly related to the Mn oxide content of the sediment.

Iron(II)-bearing minerals can catalyze the abiotic reduction of nitrate and nitrite. Some field studies have suggested that structural Fe(II) in clay minerals catalyzes the reduction of NO_3^- to NO_2^-, as indicated by significant negative correlations in the

NO_3^- and Fe(II) content in minerals in clayey till and catchment sediments (Ernstsen and Morup, 1992; Ernstsen, 1996). Chloride and sulfate intercalated green rusts (mixed Fe(II/III) hydrous oxides), which commonly occurs in anaerobic soils and sediments, are able to rapidly reduce NO_3^- to NH_4^+ NH_4^+, with the formation of magnetite (Fe_3O_4) (Hansen et al., 1996; Hansen and Koch, 1998; Hansen et al., 2001). Wustite (FeO) is able to catalyze the rapid reduction of NO_3^{2-} to NO_2^- and ultimately NH_4^+ (Rakshit et al., 2005). During this reaction, an Fe(II/III) oxide, magnetite (Fe_3O_4), is formed. Siderite ($FeCO_3$) is able to catalyze the abiotic reduction of NO_2^- to N_2O, during which siderite is transformed to lepidocrocite (γ-FeOOH) (Rakshit et al., 2008).

Nitrate immobilization into dissolved and humic organic N fractions can be abiotically catalyzed in soils and sediments (Dail et al., 2001; Davidson et al., 2003; Huygens et al., 2008; Torres-Canabate et al., 2008). It has been suggested that anaerobic microsites in soils that contain reduced Fe and Mn minerals are responsible for catalyzing the reactions (Davidson et al., 2003; Torres-Canabate et al., 2008). The mechanism is thought to involve the abiotic reduction of NO_3^{2-} by Fe(II) (present in anaerobic microsites) to NO_2^-, which then rapidly abiotically reacts with existing organic matter via nitrosation (Thorn and Mikita, 2000) and thus is incorporated into the organic N fraction (Davidson et al., 2003). Further research is required to elucidate the exact mechanism of abiotic nitrate immobilization into organic N.

Manganese oxides also catalyze the oxidation of gaseous CO (Klier and Kuchynka, 1966). This reaction occurs by consumption of the structural O_2^{2-} of MnO_2. The catalyst must then be regenerated by chemisorption of O_2. Whether or not a similar cycle of Mn oxide reduction and O_2 consumption occurs in soil environments, where adsorbed molecules such as CO_2 and H_2O can poison the catalyst surface, remains unanswered.

Soil clay minerals and metal oxides, such as montmorillonite and goethite, can catalyze the abiotic hydrolytic degradation of high- and low-molecular-weight silicones, such as polydimethylsiloxane, by cleaving siloxane (Si–O–Si) bonds (Graiver et al., 2003). The resultant low-molecular-weight silanol units are volatile and thus evaporate into the atmosphere where they are degraded to silica, water, and CO_2 by hydroxyl-free radicals in the atmosphere (Graiver et al., 2003).

18.5 Role of Nanoparticles in Abiotic Catalysis

Mineral nanoparticles are important constituents of natural terrestrial and aquatic environments (Gilbert and Banfield, 2005; Waychunas et al., 2005; Wigginton et al., 2007; Hochella et al., 2008; Wilson et al., 2008). Iron oxides, oxyhydroxides, and sulfides commonly occur as nanoparticles in nature; and various studies have investigated the abiotic transformation of organics (Klupinski et al., 2004; McCormick and Adriaens, 2004; Chun et al., 2006; Cwiertny et al., 2008) and inorganics, such as Sb (Kirsch et al., 2008) and Se (Loyo et al., 2008; Scheinost and Charlet, 2008) by nanoparticulate Fe-bearing minerals.

Mineral nanoparticles exhibit considerably different chemical and physical properties than their larger counterparts, due to the fact that a large percentage of their atoms are on the surface (e.g., for a particle with a diameter of 4 nm approximately 50% of its atoms are on the surface) (Grassian, 2008; Hochella et al., 2008). This can result in surface strain or reconstruction (relaxation), structural disorder, and variations in surface topography and crystallographic structure that is exposed. Although nanoparticles are highly reactive, which is attributable to their very high specific surface area, it has become clear that their increased reactivity cannot be solely attributed to this characteristic. Surface area-normalized studies have reported enhanced nanoparticle reactivity such as cation adsorption and redox reactions, which could only be attributed to a greater density of reactive sites per surface area unit or a greater inherent reactivity of nanoparticle surface sites (Waychunas et al., 2005; Grassian, 2008; Hochella et al., 2008).

Several studies have shown that the reactivity of Fe oxides and oxyhydroxides, which commonly occur as nanoparticles (3–10 nm range), is strongly dependent on particle size. The surface area-normalized reductive dissolution of 4 nm particles of ferrihydrite by hydroquinone is 20 times faster than that of 6 nm particles (Anschutz and Penn, 2005). Similarly, Erbs et al. (2008) investigated the effect of particle size on reductive dissolution of nanoparticulate ferrihydrite by hydroquinone, and they attributed the increased reactivity to increased reagent-surface encounters as indicated by a sixfold increase in the pre-exponential factor when comparing 3.9 nm particles to 5.9 nm particles. They hypothesized that the rate of diffusion of the reagent to the nanoparticle surface is size dependent, as the structure of the water is affected by the size of the nanoparticle.

Nanoparticulate hematite oxidation of soluble Mn^{2+} in the presence of O_2 is dependent on the size of the nanoparticles as related to surface structure and topography (Madden and Hochella, 2005). The surface area-normalized oxidation rate of hematite particles with an average diameter of 7.3 nm is at least one order of magnitude higher than those with an average diameter of 37 nm. The oxidation reaction proceeds most readily at topographic features that distort the octahedral Mn^{2+} coordination environment. These distorting surface features increase as the particle size of hematite decreases. This has the effect of lowering the reorganization energy, which effectively controls the magnitude of the transition state barrier (Madden and Hochella, 2005).

Particle diameter of nanoparticulate magnetite strongly affects its reactivity toward the reductive degradation of carbon tetrachloride (Vikesland et al., 2007). Magnetite with a 9 nm diameter is far more reactive than 80 nm diameter magnetite on both a mass and surface area-normalized basis. The aggregation state of the 9 nm magnetite significantly affects its reactivity, and an increase in ionic strength leads to a decrease in reactivity due to increased aggregation of the nanoparticles. Jung et al. (2008a) compared the reactivity of biogenic and commercial nanoparticulate magnetite toward catalyzing the ozonation of parachlorobenzoic acid. Biogenic magnetite was found to be a less

efficient catalyst due to its tendency to form larger aggregates than the commercial magnetite (Jung et al., 2008a). While investigating the abiotic reductive transformation of nitrobenzene by Fe(II)-sorbed on goethite nanoparticles, Cwiertny et al. (2008) found that aggregation of the goethite nanoparticles resulted in the decreased BET-specific surface area-normalized rate constants, implying that they are less reactive than larger particles. They concluded that new methods are needed to quantify the amount of surface area accessible for reaction in wet nanoparticle suspensions rather than assuming the specific surface area of the dry nanoparticles is relevant.

In view of the high specific surface area and the density and nature of reactive sites of nanoparticles, their role in the catalytic reactions in soil and related environments merits increasing attention for years to come.

18.6 Conclusions

The significance of abiotic catalysis in the transformation of natural organic, anthropogenic organic compounds, metals, metalloids, and other inorganics is far greater than had previously been perceived. Mineral colloids are abundant in nature. Abiotic catalysts include metal (oxy)hydroxides and oxides, SRO aluminosilicates, layer silicates, and primary minerals. The ability of these catalysts to promote the transformation of vital elements and environmental pollutants varies significantly with their atomic bonding and structural configuration and surface chemistry and the structure and functionality of the substrates.

The transformation of biomolecules to humic substances is substantially promoted by abiotic catalysis. Phenolics are the most widely distributed class of plant secondary metabolites and play an important role in regulating nutrient cycling in the terrestrial environment. Abiotic catalysis promotes the formation of humic substances from the phenolic compounds especially in the presence of the amino acids through oxidative polymerization, polycondensation, ring cleavage, decarboxylation, dealkylation, and/or deamination. Sugars and amino acids are among the most abundant constituents of the terrestrial ecosystem. The Maillard reaction, involving condensation reactions between reducing sugars and amino acids, is substantially promoted by abiotic catalysis. In nature, the Maillard reaction and polyphenol pathways should interact closely, since sugars, amino acids, and polyphenols coexist in soil solutions and natural waters. The integrated polyphenol–Maillard reaction system under abiotic catalysis is more effective in generating humic polymers than the Maillard reaction or the polyphenol pathway alone. A series of mineral colloids promotes such catalytic reactions. Among abiotic catalysts, Mn oxides are most reactive in the transformation of these biomolecules.

Mineral colloids can catalyze the hydrolysis of many xenobiotics through the ability of their metal-containing surfaces to behave as Brønsted acids and donate protons or to act as Lewis acids and accept electron pairs. The catalytic effectiveness of a metal ion depends on its ability to complex reactant molecules and shift electron density and conformation in ways favorable to the reactions. A wide range of xenobiotics such as pesticides, antibiotics, dyes, and explosives are subject to abiotic soil mineral-catalyzed oxidative transformation. In anoxic groundwater and surficial waters, biogeochemical processes provide a variety of potential reductants to catalyze abiotic transformation of persistent xenobiotics. Abiotic degradation pathways of biocides (pesticides and antibiotics) are particularly important in the environment as the toxicity of these substances can limit their biotic degradation. Much is known about the toxicity of xenobiotic parent compounds. To date, relatively little is known about the bioavailability and toxicity of xenobiotics after reaction with abiotic catalysts, such as metal (oxy)hydroxides and oxides and clay minerals.

The transformations of metals and metalloids are substantially influenced by Mn oxides and Fe-bearing minerals. Manganese(IV) oxides have the ability to catalyze many reactions such as the oxidation of As(III) to As(V), Fe(II) to Fe(III), Pu(III, IV, V) to Pu(VI), and Cr(III) to Cr(IV). The ability of Mn(IV) oxides to catalyze these transformation reactions varies with their structure and surface properties and can also be influenced by the kinds and levels of surface coatings. Iron(III)-bearing minerals also have the ability to catalyze the oxidation of some transition metals such as the transformation of V(II) to V(IV) and Sb(III) to Sb(V). On the other hand, the ability of Fe(II)-bearing minerals to catalyze the reduction reactions of a variety of trace metals and metalloids (e.g., Cr, Sb, V, Hg, U, Tc, Pu, Np, and Se) has been demonstrated. Besides Fe(II)-bearing minerals, nonliving biomaterials are able to abiotically catalyze the adsorption-coupled reduction of Cr(VI) to Cr(III). The solubility, mobility, toxicity, and food-chain contamination of metals and metalloids can be significantly influenced by abiotic catalysis.

Manganese(IV) oxides have been shown to catalyze the conversion of NO_2^- to NO_3^-. The reaction is theorized as mainly the insertion of electrons from NO_2^- into the overlapping electron orbitals of the Mn oxides. Manganese(IV) oxides also catalyze the oxidation of gaseous CO through consumption of structural O_2^-. Furthermore, nitrate immobilization into dissolved and humic organic N fractions can be abiotically catalyzed by reduced Fe and Mn (present in anaerobic microsites) in soils and sediments. Further research is required to elucidate the exact mechanism of nitrate immobilization into organic N.

Nanoparticles are important environmental particles, which influence the pedosphere, hydrosphere, biosphere, and atmosphere. Mineral nanoparticles exhibit considerably different chemical and physical properties, compared with their larger counterparts. This is ascribed to the exposure of a large percentage of atoms on the surface of nanoparticles and the resultant high specific surface area and the increase in density and reactivity of reactive sites of nanoparticles. The catalytic role of mineral nanoparticles in the transformation of organics and inorganics has been demonstrated. More in-depth research on the abiotic catalysis of environmental nanoparticles and on the biogeochemical and ecological impacts is warranted.

Therefore, abiotic catalysis deserves close attention in sustaining environmental quality and ecosystem health. Furthermore, abiotic and biotic catalysts coexist in soil and associated environments. Abiotic catalysts can influence microbial formation of enzymes (biotic catalysts) and enzymatic activity. The interactions and relative importance of abiotic and biotic catalysts and the impact on terrestrial ecosystem integrity should, thus, be an important and exciting issue for years to come.

Acknowledgment

References

Ahn, M.-Y., C.E. Martínez, D.D. Archibald, A.R. Zimmerman, J.-M. Bollag, and J. Dec. 2006. Transformation of catechol in the presence of a laccase and birnessite. Soil Biol. Biochem. 38:1015–1020.

Amacher, M.L., and D.E. Baker. 1982. Redox reactions involving chromium, plutonium, and manganese in soils. *DOE/DP/04515-1*. Pennsylvania State University, University Park, PA.

Amirbahman, A., D.B. Kent, G.P. Curtis, and J.A. Davis. 2006. Kinetics of sorption and abiotic oxidation of arsenic(III) by aquifer materials. Geochim. Cosmochim. Acta 70:533–547.

Amonette, J.E., D.J. Workman, D.W. Kennedy, J.S. Fruchter, and Y.A. Gorby. 2000. Dechlorination of carbon tetrachloride by Fe(II) associated with goethite. Environ. Sci. Technol. 34:4606–4613.

Anderson, H.A., W. Bick, A. Hepburn, and M. Stewart. 1989. Nitrogen in humic substances, p. 223–253. *In* M.H.B. Hayes, P. MacCarthy, R.L. Malcolm, and R.S. Swift (eds.) Humic substances II. In search of structure. Wiley-Interscience, Chichester, U.K.

Andre, C., and G.R. Choppin. 2000. Reduction of Pu(V) by humic acid. Radiochim. Acta 88:613–616.

Anschutz, A.J., and R.L. Penn. 2005. Reduction of crystalline iron(III) oxyhydroxides using hydroquinone: Influence of phase and particle size. Geochem. Trans. 6:60–66.

Arafaioli, P., O.L. Pantani, M. Bosetto, and G.G. Ristori. 1999. Influence of clay minerals and exchangeable cations on the formation of humic-like substances (melanoidins) from D-glucose and L-tyrosine. Clay Miner. 34:487–497.

Arafaioli, P., G.G. Ristori, M. Bosetto, and P. Fusi. 1997. Humic-like compounds formed from L-tryptophan and D-glucose in the presence of Cu(II). Chemosphere 35:575–584.

Armstrong, D.E., G. Chesters, and R.F. Harris. 1967. Atrazine hydrolysis in soils. Soil Sci. Soc. Am. Proc. 31:61–66.

Baldrian, P., V. Merhautová, J. Gabriel et al. 2006. Decolorization of synthetic dyes by hydrogen peroxide with heterogeneous catalysis by mixed iron oxides. Appl. Catal. B 66:258–264.

Barrett, K.A., and M.B. McBride. 2005. Oxidative degradation of glyphosate and aminomethylphosphonate by manganese oxide. Environ. Sci. Technol. 39:9223–9228.

Bartlett, R.J. 1981. Nonmicrobial nitrite-to-nitrate transformation in soils. Soil Sci. Soc. Am. J. 45:1054–1058.

Bartlett, R.J. 1986. Soil redox behavior, p. 179–207. *In* D.L. Sparks (ed.) Soil physical chemistry. CRC Press, Boca Raton, FL.

Bartlett, R.J., and B. James. 1979. Behavior of chromium in soils. III. Oxidation. J. Environ. Qual. 8:31–35.

Bases, C.F., and R.E. Mesmer. 1986. The hydrolysis of cations. Krieger Publishing Company, Malabar, FL.

Basolo, F., and R.G. Pearson. 1968. Mechanisms of inorganic reactions, 2nd Ed. John Wiley & Sons, New York.

Bell, R.P. 1941. Acid–base catalysis. Oxford University Press, Oxford, U.K.

Belzile, N., Y.W. Chen, and Z.J. Wang. 2001. Oxidation of antimony (III) by amorphous iron and manganese oxyhydroxides. Chem. Geol. 174:379–387.

Belzile, N., Y.W. Chen, and R.R. Xu. 2000. Early diagenetic behaviour of selenium in freshwater sediments. Appl. Geochem. 15:1439–1454.

Bittner, S. 2006. When quinones meet amino acids: Chemical, physical and biological consequences. Amino Acids 30:205–224.

Blanchet, P.-F., and A. St.-George. 1982. Kinetics of chemical degradation of organophosphorus pesticides: Hydrolysis of chlorpyrifos and chlorpyrifos-methyl in the presence of copper (II). Pestic. Sci. 13:85–91.

Bollag, J.-M., J. Dec, and P.M. Huang. 1998. Formation mechanisms of complex organic structures in soil habitats. Adv. Agron. 63:237–266.

Bollag, J.-M., S.-Y. Liu, and R.D. Minard. 1982. Enzymatic oligomerization of vanillic acid. Soil Biol. Biochem. 14:157–163.

Bollag, J.-M., C. Myers, S. Pal, and P.M. Huang. 1995. The role of abiotic and biotic catalysts in the transformation of phenolic compounds, p. 299–310. *In* P.M. Huang, J. Berthelin, J.-M. Bollag, W.B. McGill, and A.L. Page (eds.) Environmental impact of soil component interactions, Vol. 1. Natural and anthropogenic organics. Lewis Publishers, Boca Raton, FL.

Boparai, H.K., S.D. Comfort, P.J. Shea, and J.E. Szecsody. 2008. Remediating explosive-contaminated groundwater by in situ redox manipulation (ISRM) of aquifer sediments. Chemosphere 71:933–941.

Bose, J.L., A.B. Foster, M. Stacey, and J.M. Webber. 1959. Action of manganese dioxide on simple carbohydrates. Nature 184:1301–1302.

Bosetto, M., P. Arafaioli, O.L. Pantani, and G.G. Ristori. 1997. Study of the humic-like compounds formed from L-tyrosine on homoionic clays. Clay Miner. 32:341–349.

Bosetto, M., P. Arafaioli, G.G. Ristori, and P. Fusi. 1995. Formation of melanin-type compounds from L-tryptophan on Ca-saturated and Al-saturated clays. Fresenius Environ. Bull. 4:369–374.

Breynaert, E., C. Bruggeman, and A. Maes. 2008. XANES-EXAFS analysis of Se solid-phase reaction products formed upon contacting Se(IV) with FeS_2 and FeS. Environ. Sci. Technol. 42:3595–3601.

Bricker, O. 1965. Some stability relations in the system Mn–O_2–H_2O at 25°C and one atmosphere total pressure. Am. Mineral. 50:1296–1354.

Brigatti, M.F., C. Lugli, G. Cibin et al. 2000. Reduction and sorption of chromium by Fe(II)-bearing phyllosilicates: Chemical treatments and X-ray absorption spectroscopy (XAS) studies. Clay. Clay Miner. 48:272–281.

Brønsted, J.N. 1928. Acid and base catalysis. Chem. Rev. 5:231–338.

Brown, C.B., and J.L. White. 1969. Reactions of 12 *s*-triazines with soil clays. Soil Sci. Soc. Am. J. Proc. 33:863–867.

Bruggeman, C., A. Maes, J. Vancluysen, and P. Vandenmussele. 2005. Selenite reduction in Boom clay: Effect of FeS_2, clay minerals and dissolved organic matter. Environ. Pollut. 137:209–221.

Brunetti, G., N. Senesi, and C. Plaza. 2007. Effects of amendment with treated and untreated olive oil mill wastewaters on soil properties, soil humic substances and wheat yield. Geoderma 138:144–152.

Buerge, I.J., and S.J. Hug. 1998. Influence of organic ligands on chromium(VI) reduction by iron(II). Environ. Sci. Technol. 32:2092–2099.

Buerge, I.J., and S.J. Hug. 1999. Influence of mineral surfaces on chromium(VI) reduction by Iron(II). Environ. Sci. Technol. 33:4285–4291.

Bussan, A.L., and T.J. Strathmann. 2007. Influence of organic ligands on the reduction of polyhalogenated alkanes by iron(II). Environ. Sci. Technol. 41:6740–6747.

Butcher, W.W., and F.H. Westheimer. 1955. The lanthanum hydroxide gel promoted hydrolysis of phosphate esters. J. Am. Chem. Soc. 77:2420–2424.

Butler, E.C., and K.F. Hayes. 1998. Effects of solution composition and pH on the reductive dechlorination of hexachloroethane by iron sulfide. Environ. Sci. Technol. 32:1276–1284.

Cervini-Silva, J., R.A. Larson, J. Wu, and J.W. Stucki. 2001. Transformation of chlorinated aliphatic compounds by ferruginous smectite. Environ. Sci. Technol. 35:805–809.

Chang, H.M., and G.G. Allen. 1971. Oxidation, p. 433–485. *In* K.V. Sarkanene and C.H. Ludwig (eds.) Lignins. Wiley-Interscience, New York.

Charlet, L., A.C. Scheinost, C. Tournassat et al. 2007. Electron transfer at the mineral/water interface: Selenium reduction by ferrous iron sorbed on clay. Geochim. Cosmochim. Acta 71:5731–5749.

Charlet, L., E. Silvester, and E. Liger. 1998. N-compound reduction and actinide immobilisation in surficial fluids by Fe(II): The surface Fe(III)OFe(II)OH degrees species, as major reductant. Chem. Geol. 151:85–93.

Chen, Z., K.W. Kim, Y.G. Zhu, R. McLaren, F. Liu, and J.Z. He. 2006. Adsorption (As-III, As-V) and oxidation (As-III) of arsenic by pedogenic Fe–Mn nodules. Geoderma 136:566–572.

Cheney, M.A., G. Sposito, A.E. McGrath, and R.S. Criddle. 1996. Abiotic degradation of 2,4-D (dichlorophenoxyacetic acid) on synthetic birnessite: A calorespirometric method. Colloids Surf. A 107:131–140.

Cheng, H.H. 1991. Pesticides in the soil environment: Processes, impacts, and modeling. SSSA, Madison, WI.

Cheng, M.M., W.J. Song, W.H. Ma et al. 2008. Catalytic activity of iron species in layered clays for photodegradation of organic dyes under visible irradiation. Appl. Catal. B 77:355–363.

Chiu, V.Q., and J.G. Hering. 2000. Arsenic adsorption and oxidation at manganite surfaces. 1. Method for simultaneous determination of adsorbed and dissolved arsenic species. Environ. Sci. Technol. 34:2029–2034.

Chon, C.M., J.G. Kim, and H.S. Moon. 2006. Kinetics of chromate reduction by pyrite and biotite under acidic conditions. Appl. Geochem. 21:1469–1481.

Chun, C.L., R.M. Hozalski, and T.A. Arnold. 2005. Degradation of drinking water disinfection byproducts by synthetic goethite and magnetite. Environ. Sci. Technol. 39:8525–8532.

Chun, C.L., R.L. Penn, and W.A. Arnold. 2006. Kinetic and microscopic studies of reductive transformations of organic contaminants on goethite. Environ. Sci. Technol. 40:3299–3304.

Cleveland, J.M. 1970. The chemistry of plutonium. Gordon and Breach, New York.

Colarieti, M.L., G. Toscano, M.R. Ardi, and G. Greco. 2006. Abiotic oxidation of catechol by soil metal oxides. J. Hazard. Mater. 134:161–168.

Colon, D., E.J. Weber, and J.L. Anderson. 2006a. QSAR study of the reduction of nitroaromatics by Fe(II) species. Environ. Sci. Technol. 40:4976–4982.

Colon, D., E.J. Weber, and J.L. Anderson. 2008. Effect of natural organic matter on the reduction of nitroaromatics by Fe(II) species. Environ. Sci. Technol. 42:6538–6543.

Colon, D., E.J. Weber, J.L. Anderson, P. Winget, and L.A. Suarez. 2006b. Reduction of nitrosobenzenes and *N*-hydroxylanilines by Fe(II) species: Elucidation of the reaction mechanism. Environ. Sci. Technol. 40:4449–4454.

Cornell, R.M., and R. Giovanoli. 1991. Transformation of akaganeite into goethite and hematite in the presence of Mn. Clay. Clay Miner. 39:144–150.

Cui, D.Q., and T.E. Eriksen. 1996. Reduction of pertechnetate by ferrous iron in solution: Influence of sorbed and precipitated Fe(II). Environ. Sci. Technol. 30:2259–2262.

Curtis, G.P., and M. Reinhard. 1994. Reductive dehalogenation of hexachlorethane, carbon-tetrachloride, and bromoform by anthrahydroquinone disulfonate and humic-acid. Environ. Sci. Technol. 28:2393–2401.

Cwiertny, D.M., R.M. Handler, M.V. Schaefer, V.H. Grassian, and M.M. Scherer. 2008. Interpreting nanoscale size-effects in aggregated Fe-oxide suspensions: Reaction of Fe(II) with goethite. Geochim. Cosmochim. Acta 72:1365–1380.

Dail, D.B., E.A. Davidson, and J. Chorover. 2001. Rapid abiotic transformation of nitrate in an acid forest soil. Biogeochemistry 54:131–146.

Daintith, J. 1990. A concise dictionary of chemistry. Oxford University Press, Oxford, U.K.

Davidson, E.A., J. Chorover, and D.B. Dail. 2003. A mechanism of abiotic immobilization of nitrate in forest ecosystems: The ferrous wheel hypothesis. Global Change Biol. 9:228–236.

Davis, J.A., and K.F. Hayes. 1986. Geochemical processes at mineral surfaces. American Chemical Society, Washington, DC.

Dec, J., K. Haider, and J.-M. Bollag. 2001. Decarboxylation and demethoxylation of naturally occurring phenols during coupling reactions and polymerization. Soil Sci. 166:660–671.

Dec, J., K. Haider, and J.-M. Bollag. 2003. Release of substituents from phenolic compounds during oxidative coupling reactions. Chemosphere 52:549–556.

Deschamps, E., V.S.T. Ciminelli, P.G. Weidler, and A.Y. Ramos. 2003. Arsenic sorption onto soils enriched in Mn and Fe minerals. Clay. Clay Miner. 51:197–204.

Devitre, R., N. Belzile, and A. Tessier. 1991. Speciation and adsorption of arsenic on diagenetic iron oxyhydroxides. Limnol. Oceanogr. 36:1480–1485.

Dixon, J.B., and S.B. Weed. 1989. Minerals in soil environments. SSSA, Madison, WI.

Doong, R.A., and H.C. Chiang. 2005. Transformation of carbon tetrachloride by thiol reductants in the presence of quinone compounds. Environ. Sci. Technol. 39:7460–7468.

Doyle, C.S., T. Kendelewicz, and G.E. Brown. 2004. Inhibition of the reduction of Cr(VI) at the magnetite–water interface by calcium carbonate coatings. Appl. Surf. Sci. 230:260–271.

Duff, M.C., D.B. Hunter, I.R. Triay et al. 1999. Mineral associations and average oxidation states of sorbed Pu on tuff. Environ. Sci. Technol. 33:2163–2169.

Dunnivant, F.M., R.P. Schwarzenbach, and D.L. Macalady. 1992. Reduction of substituted nitrobenzenes in aqueous solutions containing natural organic matter. Environ. Sci. Technol. 26:2133–2141.

Eary, L.E., and D. Rai. 1989. Kinetics of chromate reduction by ferrous ions derived from hematite and biotite at 25°C. Am. J. Sci. 289:180–213.

El Aamrani, S., J. Gimenez, M. Rovira et al. 2007. A spectroscopic study of uranium(VI) interaction with magnetite. Appl. Surf. Sci. 253:8794–8797.

Elsner, M., S.B. Haderlein, T. Kellerhals et al. 2004a. Mechanisms and products of surface-mediated reductive dehalogenation of carbon tetrachloride by Fe(II) on goethite. Environ. Sci. Technol. 38:2058–2066.

Elsner, M., R.P. Schwarzenbach, and S.B. Haderlein. 2004b. Reactivity of Fe(II)-bearing minerals toward reductive transformation of organic contaminants. Environ. Sci. Technol. 38:799–807.

Erbs, J.J., B. Gilbert, and R.L. Penn. 2008. Influence of size on reductive dissolution of six-line ferrihydrite. J. Phys. Chem. C 112:12127–12133.

Erdem, M., F. Gur, and F. Tumen. 2004. Cr(VI) reduction in aqueous solutions by siderite. J. Hazard. Mater. 113:219–224.

Ernstsen, V. 1996. Reduction of nitrate by Fe^{2+} in clay minerals. Clay. Clay Miner. 44:599–608.

Ernstsen, V., and S. Morup. 1992. Nitrate reduction in clayey till by Fe(II) in clay-minerals. Hyperfine Interact. 70:1001–1004.

Faust, G.T. 1940. Staining of clay minerals as a rapid means of identification in natural and beneficiation products. Report of Investigation 3522. Bureau of Mines, Washington, DC.

Feng, X.H., W.F. Tan, F. Liu, H.D. Ruan, and J.Z. He. 2006a. Oxidation of As-III by several manganese oxide minerals in absence and presence of goethite. Acta Geol. Sin. 80:249–256.

Feng, X.H., L.M. Zhai, W.F. Tan, F. Liu, and J.Z. He. 2007. Adsorption and redox reactions of heavy metals on synthesized Mn oxide minerals. Environ. Pollut. 147:366–373.

Feng, X.H., Y.Q. Zu, W.F. Tan, and F. Liu. 2006b. Arsenite oxidation by three types of manganese oxides. J. Environ. Sci. 18:292–298.

Field, J.A., F.J. Cervantes, F.P. van der Zee, and G. Lettinga. 2000. Role of quinones in the biodegradation of priority pollutants: A review. Water Res. 42:215–222.

Fife, T.H., and T.J. Przystas. 1985. Divalent metal ion catalysis in the hydrolysis of ester of picolinic acid. Metal ion promoted hydroxide ion and water catalyzed reactions. J. Am. Chem. Soc. 107:1041–1047.

Filip, Z., W. Flaig, and E. Ritz. 1977. Oxidation of some phenolic substances as influenced by clay minerals. Soil Org. Matter Stud. Proc. Symp. 2:91–96.

Fimmen, R.L., R.M. Cory, Y.P. Chin, T.D. Trouts, and D.M. McKnight. 2007. Probing the oxidation–reduction properties of terrestrially and microbially derived dissolved organic matter. Geochim. Cosmochim. Acta 71:3003–3015.

Flaig, W., H. Beutelspacher, and E. Rietz. 1975. Chemical composition and physical properties of humic substances, p. 1–211. *In* J.E. Gieseking (ed.) Soil components, Vol. 1. Organic components. Springer-Verlag, New York.

Fowkes, F.M., H.A. Benesi, R.B. Ryland et al. 1960. Clay catalyzed decomposition of insecticides. J. Agric. Food Chem. 8:203–210.

Furukawa, T., and G.W. Brindley. 1973. Adsorption and oxidation of benzidine and aniline by montmorillonite and hectorite. Clay. Clay Miner. 21:279–288.

Gianfreda, L., G. Iamarino, R. Scelza, and M.A. Rao. 2006. Oxidative catalysts for the transformation of phenolic pollutants: A brief review. Biocatal. Biotransform. 24:177–187.

Gilbert, B., and J.F. Banfield. 2005. Molecular-scale processes involving nanoparticulate minerals in biogeochemical systems. Rev. Mineral. Geochem. 59:109–155.

Gonzalez, J.M., and D.A. Laird. 2004. Role of smectites and Al-substituted goethites in the catalytic condensation of arginine and glucose. Clay. Clay Miner. 52:443–450.

Gonzalez, J.M., and D.A. Laird. 2006. Smectite-catalyzed dehydration of glucose. Clay. Clay Miner. 54:38–44.

Graiver, D., K.W. Farminer, and R. Narayan. 2003. A review of the fate and effects of silicones in the environment. J. Polym. Environ. 11:129–136.

Grassian, V.H. 2008. When size really matters: Size-dependent properties and surface chemistry of metal and metal oxide nanoparticles in gas and liquid phase environments. J. Phys. Chem. C 112:18303–18313.

Gregory, K.B., P. Larese-Casanova, G.F. Parkin, and M.M. Scherer. 2004. Abiotic transformation of hexahydro-1,3,5-trinitro-1,3,5-triazine by Fe^{II} bound to magnetite. Environ. Sci. Technol. 38:1408–1414.

Haffenden, L.J.W., and V.A. Yaylayan. 2005. Mechanism of formation of redox-reactive hydroxylated benzenes and pyrazine in ^{13}C-labelled glycine/D-glucose model systems. J. Agric. Food Chem. 53:9742–9746.

Hakala, J.A., Y.P. Chin, and E.J. Weber. 2007. Influence of dissolved organic matter and Fe(II) on the abiotic reduction of pentachloronitrobenzene. Environ. Sci. Technol. 41:7337–7342.

Hansen, H.C.B., S. Guldberg, M. Erbs, and C.B. Koch. 2001. Kinetics of nitrate reduction by green rusts—Effects of interlayer anion and Fe(II): Fe(III) ratio. Appl. Clay Sci. 18:81–91.

Hansen, H.C.B., and C.B. Koch. 1998. Reduction of nitrate to ammonium by sulfate green rust: Activation energy and reaction mechanism. Clay Miner. 33:87–101.

Hansen, H.C.B., C.B. Koch, H. NanckeKrogh, O.K. Borggaard, and J. Sorensen. 1996. Abiotic nitrate reduction to ammonium: Key role of green rust. Environ. Sci. Technol. 30:2053–2056.

Hardie, A.G., J.J. Dynes, L.M. Kozak, and P.M. Huang. 2007. Influence of polyphenols on the integrated polyphenol–Maillard reaction humification pathway as catalyzed by birnessite. Ann. Environ. Sci. 1:91–110.

Hardie, A.G., J.J. Dynes, L.M. Kozak, and P.M. Huang. 2009. Biomolecule-induced carbonate genesis in abiotic formation of humic substances in nature. Can. J. Soil Sci. 89:445–453.

Hatcher, P.G., I.A. Breger, and M.A. Mattingly. 1980. Structural characteristics of fulvic acids from continental shelf sediments. Nature 285:560–562.

Hatcher, P.G., M. Schnitzer, L.W. Dennis, and G.E. Maciel. 1981. Aromaticity of humic substances in soils. Soil Sci. Soc. Am. J. 45:1089–1094.

Hättenschwiler, S., and P.M. Vitousek. 2000. The role of polyphenols in terrestrial ecosystem nutrient cycling. Trends Ecol. Evol. 15:238–243.

Hauser, E.A., and M.B. Legget. 1940. Color reactions between clays and amines. J. Am. Chem. Soc. 62:1811–1814.

Hayes, M.H.B., and R.S. Swift. 1978. The chemistry of soil organic colloids, p. 179–319. In D.J. Greenland and M.H.B. Hayes (eds.) The chemistry of soil constituents. John Wiley & Sons, New York.

Hem, J.D. 1978. Redox processes at surfaces of manganese oxide and their effects on aqueous metal ions. Chem. Geol. 21:199–218.

Hightower, J.W., W.N. Delgass, E. Iglesia, and A.T. Bell. 1996. Studies in surface science and catalysis. Elsevier, Amsterdam, the Netherlands.

Ho, C.-T. 1996. Thermal generation of Maillard aromas, p. 27–35. In R. Ikan (ed.) The Maillard reaction. Consequences for the chemical and life sciences. John Wiley & Sons, Chichester, U.K.

Hochella, M.F., S.K. Lower, P.A. Maurice, R.L. Penn, M. Sahai, D.L. Sparks, and B.S. Twining. 2008. Nanominerals, mineral nanoparticles, and earth systems. Science 319:1631–1635.

Hoffmann, M.R. 1980. Trace metal catalysis in aquatic environments. Environ. Sci. Technol. 14:1061–1066.

Hoffmann, M.R. 1990. Catalysis in aquatic environments, p. 71–111. In W. Stumm (ed.) Aquatic chemical kinetics. John Wiley & Sons, New York.

Hofstetter, T.B., A. Neumann, and R.P. Schwarzenbach. 2006. Reduction of nitroaromatic compounds by Fe(II) species associated with iron-rich smectites. Environ. Sci. Technol. 40:235–242.

Hofstetter, T.B., R.P. Schwarzenbach, and S.B. Haderlein. 2003. Reactivity of Fe(II) species associated with clay minerals. Environ. Sci. Technol. 37:519–528.

Houghton, R.P. 1979. Metal complexes in organic chemistry. Cambridge University Press, Cambridge, U.K.

Hua, B., and B.L. Deng. 2008. Reductive immobilization of uranium(VI) by amorphous iron sulfide. Environ. Sci. Technol. 42:8703–8708.

Huang, P.M. 1990. Role of soil minerals in transformation of natural organics and xenobiotics in soil, p. 29–115. In J.-M. Bollag and G. Stozky (eds.) Soil biochemistry, Vol. 6. Marcel Dekker, New York.

Huang, P.M. 1991a. Ionic factors affecting the formation of short-range ordered aluminosilicates. Soil Sci. Soc. Am. J. 55:1172–1180.

Huang, P.M. 1991b. Kinetics of redox reactions on surfaces of Mn oxides and its impact on environmental quality, p. 191–230. In D.L. Sparks and D.L. Suares (eds.) Rates of chemical processes in soils. SSSA Special Publication 27. SSSA, Madison, WI.

Huang, P.M. 2004. Soil mineral–organic matter–microorganism interactions: Fundamentals and impacts. Adv. Agron. 82:391–472.

Huang, P.M., and R. Fujii. 1996. Selenium and arsenic, p. 793–831. In D.L. Sparks (ed.) Methods of soil analysis: Part 3. Chemical methods. SSSA, Madison, WI.

Huang, P.M., and A.G. Hardie. 2009. Formation mechanisms of humic substances in the environment, p. 41–109. In N. Senesi, B. Xing, and P.M. Huang (eds.) Biophysico-chemical processes involving natural non-living organic matter in environmental systems, Vol. 2. Wiley–IUPAC series in bio-physico-chemical processes in environmental systems. John Wiley & Sons, Hoboken, NJ.

Huang, P.M., M.K. Wang, and C.Y. Chiu. 2005. Soil mineral–organic matter–microbe interactions: Impacts on biogeochemical processes and biodiversity in soils. Pedobiologia 49:609–635.

Hulth, S., R.C. Aller, and F. Gilbert. 1999. Coupled anoxic nitrification/manganese reduction in marine sediments. Geochim. Cosmochim. Acta 63:49–66.

Huygens, D., P. Boeckx, P. Templer et al. 2008. Mechanisms for retention of bioavailable nitrogen in volcanic rainforest soils. Nature Geosci. 1:543–548.

Ikan, R., Y. Rubinsztain, A. Nissenbaum, and I.R. Kaplan. 1996. Geochemical aspects of the Maillard reaction, p. 1–25. *In* R. Ikan (ed.) The Maillard reaction: Consequences for the chemical and life sciences. John Wiley & Sons, Chichester, U.K.

Iler, R.K. 1979. The chemistry of silica. John Wiley & Sons, New York.

Ilton, E.S., A. Haiduc, C.O. Moses et al. 2004. Heterogeneous reduction of uranyl by micas: Crystal chemical and solution controls. Geochim. Cosmochim. Acta 68:2417–2435.

Ilton, E.S., and D.R. Veblen. 1994. Chromium sorption by phlogopite and biotite in acidic solutions at 25°C: Insights from X-ray photoelectron spectroscopy and electron microscopy. Geochim. Cosmochim. Acta 58:2777–2788.

Jang, J.H., and B.A. Dempsey. 2008. Coadsorption of arsenic(III) and arsenic(V) onto hydrous ferric oxide: Effects on abiotic oxidation of arsenic(III), extraction efficiency, and model accuracy. Environ. Sci. Technol. 42:2893–2898.

Jauregui, M.A., and H.M. Reisenauer. 1982. Dissolution of oxides of manganese and iron by root exudate components. Soil Sci. Soc. Am. J. 46:314–317.

Jenne, E.A. 1977. Trace element sorption by sediments and soils-sites and processes, p. 425–553. *In* W.R. Chappell and K.K. Peterson (eds.) Molybdenum in the environment, Vol. 2. Marcel Dekker, New York.

Jeon, B.H., B.A. Dempsey, and W.D. Burgos. 2003. Kinetics and mechanisms for reactions of Fe(II) with iron(III) oxides. Environ. Sci. Technol. 37:3309–3315.

Jeon, B.H., B.A. Dempsey, W.D. Burgos, M.O. Barnett, and E.E. Roden. 2005. Chemical reduction of U(VI) by Fe(II) at the solid–water interface using natural and synthetic Fe(III) oxides. Environ. Sci. Technol. 39:5642–5649.

Jernberg, K.M., and W.L. Philip. 1999. Fate of famoxadone in the environment. Pestic. Sci. 55:587–589.

Jokic, A., A.I. Frenkel, and P.M. Huang. 2001a. Effect of light on birnessite catalysis of the Maillard reaction and its implication in humification. Can. J. Soil Sci. 81:277–283.

Jokic, A., A.I. Frenkel, M.A. Vairavamurthy, and P.M. Huang. 2001b. Birnessite catalysis of the Maillard reaction: Its significance in natural humification. Geophys. Res. Lett. 28:3899–3902.

Jokic, A., H.-R. Schulten, J.N. Cutler, M. Schnitzer, and P.M. Huang. 2004a. A significant abiotic pathway for the formation of *unknown* nitrogen in nature. Geophys. Res. Lett. 31:L05502.

Jokic, A., H.-R. Schulten, J.N. Cutler, M. Schnitzer, and P.M. Huang. 2005. Catalysis of the Maillard reaction by δ-MnO$_2$: A significant abiotic sorptive condensation pathway for the formation of refractory N-containing biogeomacromolecules in nature, p. 127–152. *In* P.M. Huang, A. Violante, J.-M. Bollag, and P. Vityakon (eds.) Soil abiotic and biotic interactions and impact on the ecosystem and human welfare. Science Publishers Inc., Enfield, NH.

Jokic, A., M.C. Wang, C. Liu, A.I. Frenkel, and P.M. Huang. 2004b. Integration of the polyphenol and Maillard reactions into a unified abiotic pathway for humification in nature: The role of δ-MnO$_2$. Org. Geochem. 35:747–762.

Jokic, A., Z. Zimpel, P.M. Huang, and P.G. Mezey. 2001c. Molecular shape analysis of a Maillard reaction intermediate. SAR QSAR Environ. Res. 12:297–307.

Jung, H., J.W. Kim, H. Choi, J.H. Lee, and H.G. Hur. 2008a. Synthesis of nanosized biogenic magnetite and comparison of its catalytic activity in ozonation. Appl. Catal. B 83:208–213.

Jung, J.W., S. Lee, H. Ryu, K.H. Kang, and K. Nam. 2008b. Detoxification of phenol through bound residue formation by birnessite in soil: Transformation kinetics and toxicity. J. Environ. Sci. Health. Part A Toxic/Hazard. Subst. Environ. Eng. 43:255–261.

Jung, Y., J. Choi, and W. Lee. 2007. Spectroscopic investigation of magnetite surface for the reduction of hexavalent chromium. Chemosphere 68:1968–1975.

Kang, K.H., J. Dec, H. Park, and J.M. Bollag. 2004. Effect of phenolic mediators and humic acid on cyprodinil transformation in presence of birnessite. Water Res. 38:2737–2745.

Kang, K.H., D.M. Lim, and H. Shin. 2006. Oxidative-coupling reaction of TNT reduction products by manganese oxide. Water Res. 40:903–910.

Kappler, A., and S.B. Haderlein. 2003. Natural organic matter as reductant for chlorinated aliphatic pollutants. Environ. Sci. Technol. 37:2714–2719.

Keeneykennicutt, W.L., and J.W. Morse. 1985. The redox chemistry of Pu(V)O$_2^+$ interaction with common mineral surfaces in dilute solutions and seawater. Geochim. Cosmochim. Acta 49:2577–2588.

Kemper, J.M., E. Ammar, and W.A. Mitch. 2008. Abiotic degradation of hexahydro-1,3,5-trinitro-1,3,5-triazine in the presence of hydrogen sulfide and black carbon. Environ. Sci. Technol. 42:2118–2123.

Kersting, A.B., D.W. Efurd, D.L. Finnegan et al. 1999. Migration of plutonium in ground water at the Nevada test site. Nature 397:56–59.

Ketelaar, J.A.A., H.R. Gersmann, and M.M. Beck. 1956. Metal-catalysed hydrolysis of thiophosphoric esters. Nature 177:392–396.

Kim, D., and T.J. Strathmann. 2007. Role of organically complexed iron(II) species in the reductive transformation of RDX in anoxic environments. Environ. Sci. Technol. 41:1257–1264.

Kirsch, R., A.C. Scheinost, A. Rossberg, D. Banerjee, and L. Charlet. 2008. Reduction of antimony by nano-particulate magnetite and mackinawite. Mineral. Mag. 72:185–189.

Klier, K., and K. Kuchynka. 1966. Carbon monoxide oxidation and adsorbate-gas exchange reactions on MnO$_2$-based catalysts. J. Catal. 6:62–71.

Klupinski, T.P., and Y.P. Chin. 2003. Abiotic degradation of trifluralin by Fe(II): Kinetics and transformation pathways. Environ. Sci. Technol. 37:1311–1318.

Klupinski, T.P., Y.P. Chin, and S.J. Traina. 2004. Abiotic degradation of pentachloronitrobenzene by Fe(II): Reactions on goethite and iron oxide nanoparticles. Environ. Sci. Technol. 38:4353–4360.

Kodaka, R., T. Sugano, T. Katagi, and Y. Takimoto. 2003. Clay-catalyzed nitration of a carbamate fungicide diethofencarb. J. Agric. Food Chem. 51:7730–7737.

Kolthoff, I.M. 1921. Iodometric studies. VII. Reactions between arsenic trioxide and iodine. Anal. Chem. 60:393–406.

Kresge, A.J. 1975. Water makes proton transfer fast. Acc. Chem. Res. 8:354–360.

Kriegman-King, M.R., and M. Reinhard. 1994. Transformation of carbon tetrachloride by pyrite in aqueous solution. Environ. Sci. Technol. 28:692–700.

Krishnamurti, G.S.R., and P.M. Huang. 1987. The catalytic role of birnessite in the transformation of iron. Can. J. Soil Sci. 67:533–543.

Krishnamurti, G.S.R., and P.M. Huang. 1988. Influence of manganese oxide minerals on the formation of iron oxides. Clay. Clay Miner. 36:467–475.

Kroll, H. 1952. The participation of heavy metal ions in the hydrolysis of amino acid esters. J. Am. Chem. Soc. 74:2036–2039.

Kumada, K., and H. Kato. 1970. Browning of pyrogallol as affected by clay minerals. Soil Sci. Plant Nutr. 16:195–200.

Kummert, R., and W. Stumm. 1980. The surface complexation of organic acids on hydrous-Al_2O_3. J. Colloid Interface Sci. 75:373.

Kung, K.-H., and M.B. McBride. 1988. Electron transfer processes between hydroquinone and hausmannite (Mn_3O_4). Clay. Clay Miner. 36:297–302.

Kyuma, K., and K. Kawaguchi. 1964. Oxidative changes of polyphenols as influenced by allophone. Soil Sci. Soc. Am. Proc. 28:371–374.

Laha, S., and R.G. Luthy. 1990. Oxidation of aniline and other primary aromatic amines by manganese dioxide. Environ. Sci. Technol. 24:363–373.

Lee, J.S.K., and P.M. Huang. 1995. Photochemical effect on the abiotic transformation of polyphenols as catalyzed by Mn(IV) oxide, p. 177–189. *In* P.M. Huang, J. Berthelin, J.-M. Bollag, W.B. McGill, and A.L. Page (eds.) Environmental impact of soil component interactions, Vol. 1. Natural and anthropogenic organics. Lewis Publishers, Boca Raton, FL.

Lee, W., and B. Batchelor. 2002a. Abiotic reductive dechlorination of chlorinated ethylenes by iron-bearing soil minerals. 1. Pyrite and magnetite. Environ. Sci. Technol. 36:5147–5154.

Lee, W., and B. Batchelor. 2002b. Abiotic, reductive dechlorination of chlorinated ethylenes by iron-bearing soil minerals. 2. Green rust. Environ. Sci. Technol. 36:5348–5354.

Lee, W., and B. Batchelor. 2003. Reductive capacity of natural reductants. Environ. Sci. Technol. 37:535–541.

Lee, W.J., and B. Batchelor. 2004. Abiotic reductive dechlorination of chlorinated ethylenes by iron-bearing phyllosilicates. Chemosphere 56:999–1009.

Lehmann, R.G., and H.H. Cheng. 1988. Reactivity of phenolic acids in soils and formation of oxidation products. Soil Sci. Soc. Am. J. 52:352–356.

Lehmann, R.G., H.H. Cheng, and J.B. Harsh. 1987. Oxidation of phenolic acids by soil iron and manganese oxides. Soil Sci. Soc. Am. J. 51:352–356.

Leuz, A.K., H. Monch, and C.A. Johnson. 2006. Sorption of Sb(III) and Sb(V) to goethite: Influence on Sb(III) oxidation and mobilization. Environ. Sci. Technol. 40:7277–7282.

Li, F.B., C.S. Liu, C.H. Liang, X.Z. Li, and L.J. Zhang. 2008. The oxidative degradation of 2-mercaptobenzothiazole at the interface of beta-MnO_2 and water. J. Hazard. Mater. 154:1098–1105.

Li, H., L.S. Lee, D.G. Schulze, and C.A. Guest. 2003. Role of soil manganese in the oxidation of aromatic amines. Environ. Sci. Technol. 37:2686–2693.

Lide, D.R. 2008. CRC Handbook of chemistry and physics, 89th Ed. CRC Press, Boca Raton, FL.

Liger, E., L. Charlet, and P. Van Cappellen. 1999. Surface catalysis of uranium(VI) reduction by iron(II). Geochim. Cosmochim. Acta 63:2939–2955.

Lin, Y.T., and C.P. Huang. 2008. Reduction of chromium(VI) by pyrite in dilute aqueous solutions. Sep. Purif. Technol. 63:191–199.

Liu, C., and P.M. Huang. 2000. Catalytic effects of hydroxy-aluminum and silicic acid on catechol humification, p. 37–51. *In* E. Ghabbour and G. Davies (eds.) Humic substances: Components of plants, soil and water. Royal Society of Chemistry, Cambridge, U.K.

Liu, C., and P.M. Huang. 2001. The influence of catechol humification on surface properties of metal oxides, p. 253–270. *In* E. Ghabbour and G. Davies (eds.) Humic substances: Structure, models and functions. Royal Society of Chemistry, Cambridge, U.K.

Liu, C., and P.M. Huang. 2002. Role of hydroxy-aluminosilicate ions (proto-imogolite sol) in the formation of humic substances. Org. Geochem. 33:295–305.

Liu, S.-Y., R.D. Minard, and J.-M. Bollag. 1981. Oligomerization of syringic acid, a lignin derived derivative, by a phenoloxidase. Soil Sci. Soc. Am. J. 45:110–1105.

Livens, F.R., M.J. Jones, A.J. Hynes et al. 2004. X-ray absorption spectroscopy studies of reactions of technetium, uranium and neptunium with mackinawite. J. Environ. Radioact. 74:211–219.

Loganathan, P., R.G. Burau, and D.W. Fuerstenau. 1977. Influence of pH on the sorption of Co^{2+}, Zn^{2+}, and Ca^{2+} by a hydrous manganese oxide. Soil Sci. Soc. Am. J. 41:57–62.

Loyo, R.L.D., S.I. Nikitenko, A.C. Scheinost, and M. Simonoff. 2008. Immobilization of selenite on Fe_3O_4 and Fe/Fe_3C ultrasmall particles. Environ. Sci. Technol. 42:2451–2456.

Luther, G.W., and J.I. Popp. 2002. Kinetics of the abiotic reduction of polymeric manganese dioxide by nitrite: An anaerobic nitrification reaction. Aquat. Geochem. 8:15–36.

Luther, G.W., B. Sundby, B.L. Lewis, P.J. Brendel, and N. Silverberg. 1997. Interactions of manganese with the nitrogen cycle: Alternative pathways to dinitrogen. Geochim. Cosmochim. Acta 61:4043–4052.

Madden, A.S., and M.F. Hochella. 2005. A test of geochemical reactivity as a function of mineral size: Manganese oxidation promoted by hematite nanoparticles. Geochim. Cosmochim. Acta 69:389–398.

Maillard, L.C. 1913. Formation de matières humiques par action de polypeptides sur sucres. C.R. Acad. Sci. 156:148–149.

Manning, B.A., and S. Goldberg. 1997. Adsorption and stability of arsenic(III) at the clay mineral–water interface. Environ. Sci. Technol. 31:2005–2011.

Masscheleyn, P.H., R.D. Delaune, and W.H. Patrick. 1990. Transformations of selenium as affected by sediment oxidation reduction potential and pH. Environ. Sci. Technol. 24:91–96.

Mazumdar, A., T. Goldberg, and H. Strauss. 2008. Abiotic oxidation of pyrite by Fe(III) in acidic media and its implications for sulfur isotope measurements of lattice-bound sulfate in sediments. Chem. Geol. 253:30–37.

McBride, M.B. 1979. Reactivity of adsorbed and structural iron in hectorite as indicated by oxidation of benzidine. Clay. Clay Miner. 27:224–230.

McBride, M.B. 1994. Environmental chemistry of soils. Oxford University Press, London, U.K.

McCormick, M.L., and P. Adriaens. 2004. Carbon tetrachloride transformation on the surface of nanoscale biogenic magnetite particles. Environ. Sci. Technol. 38:1045–1053.

McKenzie, R.M. 1981. The surface charge on manganese dioxides. Aust. J. Soil Res. 19:41–50.

McKenzie, R.M. 1989. Manganese oxides and hydroxides, p. 439–465. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Mill, T., and W. Mabey. 1988. Hydrolysis of organic chemicals, p. 71–111. *In* O. Hutzinger (ed.) The handbook of environmental chemistry, Vol. 2D. Reactions and processes. Springer-Verlag, Berlin, Germany.

Minglegrin, U., and S. Salztman. 1979. Surface reactions of parathion on clays. Clay. Clay Miner. 27:72–78.

Minglegrin, U., S. Salztman, and B. Yaron. 1977. A possible model for the surface-induced hydrolysis of organo-phosphorus pesticides on kaolinite clays. Soil Sci. Soc. Am. J. 41:519–523.

Mitsunobu, S., Y. Takahashi, and Y. Sakai. 2008. Abiotic reduction of antimony(V) by green rust ($Fe_4(II)Fe_2(III)(OH)_{12}SO_4 \cdot 3H_2O$). Chemosphere 70:942–947.

Moore, J.W., and R.G. Pearson. 1981. Kinetics and mechanisms. John Wiley & Sons, New York.

Morgenstern, A., and G.R. Choppin. 2002. Kinetics of the oxidation of Pu(IV) by manganese dioxide. Radiochim. Acta 90:69–74.

Mortland, M.M. 1970. Clay-organic complexes and interactions. Adv. Agron. 22:75–115.

Mortland, M.M. 1986. Mechanisms of adsorption of nonhumic organic species by clays, p. 59–76. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and microbes. SSSA, Madison, WI.

Mortland, M.M., and L.J. Halloran. 1976. Polymerization of aromatic molecules on smectites. Soil Sci. Soc. Am. J. 40:367–370.

Murray, J.W., and J.G. Dillard. 1979. The oxidation of cobalt(II) adsorbed on manganese dioxide. Geochim. Cosmochim. Acta 43:781–787.

Myneni, S.C.B., T.K. Tokunaga, and G.E. Brown. 1997. Abiotic selenium redox transformations in the presence of Fe(II,III) oxides. Science 278:1106–1109.

Naidja, A., P.M. Huang, and J.-M. Bollag. 1998. Comparison of the reaction products from the transformation of catechol catalyzed by birnessite or tyrosinase. Soil Sci. Soc. Am. J. 62:188–195.

Naidja, A., P.M. Huang, and J.-M. Bollag. 2000. Enzyme-clay interactions and their impact on transformations of natural and anthropogenic organic compounds in soil. J. Environ. Qual. 29:677–691.

Naka, D., D. Kim, R.F. Carbonaro, and T.J. Strathmann. 2008. Abiotic reduction of nitroaromatic contaminants by iron(II) complexes with organothiol ligands. Environ. Toxicol. Chem. 27:1257–1266.

Naka, D., D. Kim, and T.J. Strathmann. 2006. Abiotic reduction of nitroaromatic compounds by aqueous iron(II)—Catechol complexes. Environ. Sci. Technol. 40:3006–3012.

Nakata, K., S. Nagasaki, S. Tanaka et al. 2004. Reduction rate of neptunium(V) in heterogeneous solution with magnetite. Radiochim. Acta 92:145–149.

Nannipieri, P., and L. Gianfreda. 1998. Kinetics of enzyme reactions in soil environments, p. 449–479. *In* P.M. Huang, N. Senesi, and J. Buffle (eds.) Structure and surface reactions of soil particles, Vol. 4. IUPAC Ser. on Analytical and physical chemistry of environmental systems. John Wiley & Sons, New York.

Nasser, A., G. Sposito, and M.A. Cheney. 2000. Mechanochemical degradation of 2,4-D adsorbed on synthetic birnessite. Colloid Surf. A 163:117–123.

Negra, C., D.S. Ross, and A. Lanzirotti. 2005. Oxidizing behavior of soil manganese: Interactions among abundance, oxidation state, and pH. Soil Sci. Soc. Am. J. 69:87–95.

Nesbitt, H.W., G.W. Canning, and G.M. Bancroft. 1998. XPS study of reductive dissolution of 7 angstrom-birnessite by H_3AsO_3, with constraints on reaction mechanism. Geochim. Cosmochim. Acta 62:2097–2110.

Nicholas, D.R., S. Ramamoorthy, V. Palace et al. 2003. Biogeochemical transformations of arsenic in circumneutral freshwater sediments. Biodegradation. 14:123–137.

Nikalje, M.D., P. Phukan, and A. Sudalai. 2000. Recent advances in clay-catalyzed organic transformations. Org. Prep. Proced. Int. 32:1–32.

O'Loughlin, E.J., S.D. Kelly, R.E. Cook, R. Csencsits, and K.M. Kemner. 2003a. Reduction of uranium(VI) by mixed iron(II/iron(III) hydroxide (green rust): Formation of UO_2 nanoparticles. Environ. Sci. Technol. 37:721–727.

O'Loughlin, E.J., S.D. Kelly, K.M. Kemner, R. Csencsits, and R.E. Cook. 2003b. Reduction of Ag-I, Au-III, Cu-II, and Hg-II by Fe-II/Fe-III hydroxysulfate green rust. Chemosphere 53:437–446.

Osano, O., W. Admiraal, H.J.C. Klamer, D. Pastor, and E.A.J. Bleeker. 2002. Comparative toxic and genotoxic effects of chloroacetanilides, formamidines and their degradation products on *Vibrio fischeri* and *Chironomus riparius*. Environ. Pollut. 119:195–202.

Oscarson, D.W., P.M. Huang, C. Defosse, and A. Herbillon. 1981a. Oxidative power of Mn(IV) and Fe(III) oxides with respect to As(III) in terrestrial and aquatic environments. Nature 291:50–51.

Oscarson, D.W., P.M. Huang, U.T. Hammer, and W.K. Liaw. 1983a. Oxidation and sorption of arsenite by manganese dioxide as influenced by surface coatings of iron and aluminum oxides and calcium carbonate. Water Air Soil Pollut. 20:233–244.

Oscarson, D.W., P.M. Huang, and W.K. Liaw. 1980. The oxidation of arsenite by aquatic sediments. J. Environ. Qual. 9:700–703.

Oscarson, D.W., P.M. Huang, and W.K. Liaw. 1981b. The kinetics and components involved in the oxidation of arsenite by freshwater sediments. Verh. Int. Ver. Theor. Angew. Limnol. 21:181–186.

Oscarson, D.W., P.M. Huang, and W.K. Liaw. 1981c. The role of manganese in the oxidation of arsenite by freshwater lake sediments. Clay. Clay Miner. 28:219–225.

Oscarson, D.W., P.M. Huang, W.K. Liaw, and U.T. Hammer. 1983b. Kinetics of oxidation of arsenite by various manganese dioxides. Soil Sci. Soc. Am. J. 47:644–648.

Pal, S., J.-M. Bollag, and P.M. Huang. 1994. Role of abiotic and biotic catalysts in the transformation of phenolic compounds through oxidative coupling reactions. Soil Biol. Biochem. 26:813–820.

Park, B., and B.A. Dempsey. 2005. Heterogeneous oxidation of Fe(II) on ferric oxide at neutral pH and a low partial pressure of O_2. Environ. Sci. Technol. 39:6494–6500.

Park, D., C.K. Ahn, Y.M. Kim, Y.S. Yun, and J.M. Park. 2008a. Enhanced abiotic reduction of Cr(VI) in a soil slurry system by natural biomaterial addition. J. Hazard. Mater. 160:422–427.

Park, D., S.R. Lim, Y.S. Yun, and J.M. Park. 2007. Reliable evidences that the removal mechanism of hexavalent chromium by natural biomaterials is adsorption-coupled reduction. Chemosphere 70:298–305.

Park, D., Y.S. Yun, H.W. Lee, and J.M. Park. 2008b. Advanced kinetic model of the Cr(VI) removal by biomaterials at various pHs and temperatures. Bioresour. Technol. 99:1141–1147.

Park, D., Y.S. Yun, and J.M. Park. 2004. Reduction of hexavalent chromium with the brown seaweed *Ecklonia* biomass. Environ. Sci. Technol. 38:4860–4864.

Park, D., Y.S. Yun, and J.M. Park. 2005. Studies on hexavalent chromium biosorption by chemically-treated biomass of *Ecklonia* sp. Chemosphere 60:1356–1364.

Pearson, R.G. 1966. Acids and bases. Science 151:172–177.

Pearson, R.G. 1976. Symmetry rules for chemical reactions. Wiley-Interscience, New York.

Pecher, K., S.B. Haderlein, and R.P. Schwarzenbach. 2002. Reduction of polyhalogenated methanes by surface-bound Fe(II) in aqueous suspensions of iron oxides. Environ. Sci. Technol. 36:1734–1741.

Penrose, W.R. 1974. Arsenic in the marine and aquatic sediments: Analysis, occurrence and significance. CRC Crit. Rev. Environ. Contr. 4:465–482.

Pepper, S.E., D.J. Bunker, N.D. Bryan et al. 2003. Treatment of radioactive wastes: An X-ray absorption spectroscopy study of the reaction of technetium with green rust. J. Colloid Interface Sci. 268:408–412.

Peretyazhko, T., J.M. Zachara, S.M. Heald et al. 2008a. Heterogeneous reduction of Tc(VII) by Fe(II) at the solid–water interface. Geochim. Cosmochim. Acta 72:1521–1539.

Peretyazhko, T., J.M. Zachara, S.M. Heald et al. 2008b. Reduction of Tc(VII) by Fe(II) sorbed on Al (hydr)oxides. Environ. Sci. Technol. 42:5499–5506.

Peterson, M.L., G.E. Brown, and G.A. Parks. 1996. Direct XAFS evidence for heterogeneous redox reaction at the aqueous chromium/magnetite interface. Colloid Surf. A 107:77–88.

Petrie, R.A., P.R. Grossl, and R.C. Sims. 2002. Oxidation of pentachlorophenol in manganese oxide suspensions under controlled Eh and pH environments. Environ. Sci. Technol. 36:3744–3748.

Pinnavais, T.J., P.L. Hall, S.S. Cady, and M.M. Mortland. 1974. Aromatic radical cation formation on the intracrystal surfaces of transition metal layer silicates. J. Phys. Chem. 78:994–999.

Pizzigallo, M.D.R., A. Napola, M. Spagnuolo, and P. Ruggiero. 2004. Mechanochemical removal of organo-chlorinated compounds by inorganic components of soil. Chemosphere 55:1485–1492.

Pizzigallo, M.D.R., P. Ruggiero, C. Crecchio, and G. Mascolo. 1998. Oxidation of chloroanilines at metal oxide surfaces. J. Agric. Food Chem. 46:2049–2054.

Plastourgou, M., and M.R. Hoffmann. 1984. Transformation and fate of organic esters in layered-flow systems: The role of trace metal catalysis. Environ. Sci. Technol. 18:756–764.

Pohlman, A.A., and J.G. McColl. 1989. Organic oxidation and manganese and aluminum mobilization in forest soils. Soil Sci. Soc. Am. J. 53:686–690.

Porterfield, W.W. 1983. Inorganic chemistry. A unified approach. Harper International SI Edition, London, U.K.

Powell, B.A., M.C. Duff, D.I. Kaplan et al. 2006. Plutonium oxidation and subsequent reduction by Mn(IV) minerals in Yucca Mountain tuff. Environ. Sci. Technol. 40:3508–3514.

Powell, B.A., R.A. Fjeld, D.I. Kaplan, J.T. Coates, and S.M. Serkiz. 2004. Pu(V)O-2(+) adsorption and reduction by synthetic magnetite (Fe_3O_4). Environ. Sci. Technol. 38:6016–6024.

Power, L.E., Y. Arai, and D.L. Sparks. 2005. Zinc adsorption effects on arsenite oxidation kinetics at the birnessite–water interface. Environ. Sci. Technol. 39:181–187.

Pracht, J., J. Boenigk, M. Isenbeck-Schroter, F. Keppler, and H.F. Scholer. 2001. Abiotic Fe(III) induced mineralization of phenolic substances. Chemosphere 44:613–619.

Rai, D., and R.J. Serne. 1977. Plutonium activities in soil solutions and the stability and formation of selected plutonium minerals. J. Environ. Qual. 6:89–95.

Rakshit, S., C.J. Matocha, and M.S. Coyne. 2008. Nitrite reduction by siderite. Soil Sci. Soc. Am. J. 72:1070–1077.

Rakshit, S., C.J. Matocha, and G.R. Haszler. 2005. Nitrate reduction in the presence of wustite. J. Environ. Qual. 34:1286–1292.

Ratasuk, N., and M.A. Nanny. 2007. Characterization and quantification of reversible redox sites in humic substances. Environ. Sci. Technol. 41:7844–7850.

Ribault, L.L. 1971. Presence d'ne pellicule de silice amorphe a la surface de cristaux de quartz des formation sableuses. C.R. Acad. Sci. Paris Ser. D 272:1933–1936.

Ross, D.S., and R.J. Bartlett. 1981. Evidence for non-microbial oxidation of manganese in soil. Soil Sci. 132:153–160.

Rubert, K.F., and J.A. Pedersen. 2006. Kinetics of oxytetracycline reaction with a hydrous manganese oxide. Environ. Sci. Technol. 40:7216–7221.

Runde, W., S.D. Conradson, D.W. Efurd et al. 2002. Solubility and sorption of redox-sensitive radionuclides (Np, Pu) in J-13 water from the Yucca Mountain site: Comparison between experiment and theory. Appl. Geochem. 17:837–853.

Russell, J.D., M. Cruz, and J.L. White. 1968a. The adsorption of 3-aminotriazole by montmorillonite. J. Agric. Food Chem. 16:21–24.

Russell, J.D., M. Cruz, and J.L. White. 1968b. Model of chemical degradation of *s*-triazines by montmorillonite. Science 160:1340–1342.

Russo, F., M.A. Rao, and L. Gianfreda. 2005. Bioavailability of phenanthrene in the presence of birnessite-mediated catechol polymers. Appl. Microbiol. Biotechnol. 68:131–139.

Salbu, B., O.C. Lind, and L. Skipperud. 2004. Radionuclide speciation and its relevance in environmental impact assessments. J. Environ. Radioact. 74:233–242.

Salvestrini, S., S. Capasso, and P. Lovino. 2008. Catalytic effect of dissolved humic acids on the chemical degradation of phenylurea herbicides. Pest Manag. Sci. 64:768–774.

Salztman, S., B. Yaron, and U. Minglegrin. 1974. The surface-catalyzed hydrolysis of parathion on kaolinite. Soil Sci. Soc. Am. Proc. 38:231–234.

Sanchez-Camazano, M., and M.J. Sanchez-Martin. 1983. Montmorillonite-catalyzed hydrolysis of phosmet. Soil Sci. 136:89–93.

Sassman, S.A., and L.S. Lee. 2007. Sorption and degradation in soils of veterinary ionophore antibiotics: Monensin and lasalocid. Environ. Toxicol. Chem. 26:1614–1621.

Sassman, S.A., A.K. Sarmah, and L.S. Lee. 2007. Sorption of tylosin A, D, and A-aldol and degradation of tylosin A in soils. Environ. Toxicol. Chem. 26:1629–1635.

Scheffer, F., B. Meyer, and E.A. Niederbudde. 1959. Huminstoffbildung unter katalytischer Einwirkung natürlich vorkommender Eisenverbindungen im modelversuch. Z. Pflanzenernahr. Dung. Bodenkd. 87:26–44.

Scheidegger, A.M., D. Grolimund, D. Cui et al. 2003. Reduction of selenite on iron surfaces: A micro-spectroscopic study. J. Phys. IV 104:417–420.

Scheinost, A.C., and L. Charlet. 2008. Selenite reduction by mackinawite, magnetite and siderite: XAS characterization of nanosized redox products. Environ. Sci. Technol. 42:1984–1989.

Schnitzer, M. 1977. Recent findings on the characterization of humic substances extracted from soils from widely differing climatic zones, p. 77–101. *In* Proceedings of the symposium on soil organic matter studies II. IAEA Bulletin, Vienna, Austria.

Schnitzer, M. 1978. Humic substances: Chemistry and reactions, p. 1–64. *In* M. Schnitzer and S.U. Khan (eds.) Soil organic matter. Elsevier, Amsterdam, the Netherlands.

Schnitzer, M. 1982. Quo vadis soil organic matter research. Trans. 12th Int. Congr. Soil Sci. 5:67–78.

Schnitzer, M., and K. Ghosh. 1982. Characteristics of water–soluble fulvic acid–copper and fulvic acid–iron complexes. Soil Sci. 134:354–363.

Schnitzer, M., and M. Levesque. 1979. Electron spin resonance as a guide to the degree of humification of peats. Soil Sci. 127:140–145.

Scott, T.B., G.C. Allen, P.J. Heard, and M.G. Randell. 2005. Reduction of U(VI) to U(IV) on the surface of magnetite. Geochim. Cosmochim. Acta 69:5639–5646.

Senesi, N., and M. Schnitzer. 1977. The effect of pH, reaction time, chemical reduction and irradiation on ESR spectra of fulvic acids. Soil Sci. 123:224–234.

Shaughnessy, D.A., H. Nitsche, C.H. Booth et al. 2003. Molecular interfacial reactions between Pu(VI) and manganese oxide minerals manganite and hausmannite. Environ. Sci. Technol. 37:3367–3374.

Shcherbina, N.S., S.N. Kalmykov, I.V. Perminova, and A.N. Kovalenko. 2007a. Reduction of actinides in higher oxidation states by hydroquinone-enriched humic derivatives. J. Alloy. Comp. 444:518–521.

Shcherbina, N.S., I.V. Perminova, S.N. Kalmykov et al. 2007b. Redox and complexation interactions of neptunium(V) with quinonoid-enriched humic derivatives. Environ. Sci. Technol. 41:7010–7015.

Shin, J.Y., C.M. Buzgo, and M.A. Cheney. 2000. Mechanochemical degradation of atrazine adsorbed on four synthetic manganese oxides. Colloid Surf. A 172:113–123.

Shin, J.Y., and M.A. Cheney. 2004. Abiotic transformation of atrazine in aqueous suspension of four synthetic manganese oxides. Colloid Surf. A 242:85–92.

Shin, J.Y., and M.A. Cheney. 2005. Abiotic dealkylation and hydrolysis of atrazine by birnessite. Environ. Toxicol. Chem. 24:1353–1360.

Shindo, H. 1992. Relative effectiveness of short-range ordered Mn(IV), Fe(III), Al, and Si oxides in the synthesis of humic acids from phenolic compounds. Soil Sci. Plant Nutr. 38:459–465.

Shindo, H., and P.M. Huang. 1982. Role of Mn(IV) oxide in abiotic formation of humic substances in the environment. Nature 298:363–365.

Shindo, H., and P.M. Huang. 1984a. Catalytic effects of manganese(IV), iron(III), aluminium, and silicon oxides on the formation of phenolic polymers. Soil Sci. Soc. Am. J. 48:927–934.

Shindo, H., and P.M. Huang. 1984b. Significance of Mn(IV) oxide in abiotic formation of organic nitrogen complexes in natural environments. Nature 308:57–58.

Shindo, H., and P.M. Huang. 1985a. The catalytic power of inorganic components in the abiotic synthesis of hydroquinone-derived polymers. Appl. Clay Sci. 1:71–81.

Shindo, H., and P.M. Huang. 1985b. Catalytic polymerization of hydroquinone by primary minerals. Soil Sci. 139:505–511.

Shindo, H., and P.M. Huang. 1992. Comparison of the influence of Mn(IV) oxide and tyrosinase on the formation of humic substances in the environment. Sci. Total Environ. 117/118:103–110.

Sinclair, C.J., and A.B.A. Boxall. 2003. Assessing the ecotoxicity of pesticide transformation products. Environ. Sci. Technol. 37:4617–4625.

Skovbjerg, L.L., S.L.S. Stipp, S. Utsunomiya, and R.C. Ewing. 2006. The mechanisms of reduction of hexavalent chromium by green rust sodium sulphate: Formation of Cr-goethite. Geochim. Cosmochim. Acta 70:3582–3592.

Smith, M.B., and J. March. 2001. March's advanced organic chemistry reactions, mechanisms, and structure. John Wiley & Sons, New York.

Smith, B.A., W.E. Siems, A.L. Teel, and R.J. Watts. 2006. Pyrolusite (beta-MnO_2)-mediated, near dry-phase oxidation of 2,4,6-trichlorophenol. Environ. Toxicol. Chem. 25:1474–1479.

Smolen, J.M., and A.T. Stone. 1998. Organophosphorus ester hydrolysis catalyzed by dissolved metals and metal-containing surfaces, p. 157–171. *In* P.M. Huang, D.C. Adriano, T.J. Logan, and R.T. Checkai (eds.) Soil chemistry and ecosystem health. SSSA, Madison, WI.

Solomon, D.H. 1968. Clay minerals as electron acceptors and/or electron donors in organic reactions. Clay. Clay Miner. 16:31–39.

Solomon, D.H., and D.G. Hawthorne. 1983. Chemistry of pigments and fillers. John Wiley & Sons, New York.

Sorensen, K.C., J.W. Stucki, R.E. Warner, and M.J. Plewa. 2004. Alteration of mammalian-cell toxicity of pesticides by structural iron(II) in ferruginous smectite. Environ. Sci. Technol. 38:4383–4389.

Sorensen, K.C., J.W. Stucki, R.E. Warner, E.D. Wagner, and M.J. Plewa. 2005. Modulation of the genotoxicity of pesticides reacted with redox-modified smectite clay. Environ. Mol. Mutagen. 46:174–181.

Sorensen, J., and L. Thorling. 1991. Stimulation by lepidocrocite (gamma-FeOOH) of Fe(II)-dependant nitrite reduction. Geochim. Cosmochim. Acta 55:1289–1294.

Sposito, G. 1984. The surface chemistry of soils. Oxford University Press, Oxford, U.K.

Stanton, D.T. 1987. Glass-catalyzed decomposition of oxycarboxin in aqueous solution. J. Agric. Food Chem. 35:856–859.

Steelink, C. 1994. Application of N-15 NMR spectroscopy to the study of organic nitrogen and humic substances in the soil, p. 405–427. *In* N. Senesi and D. Miano (eds.) Humic substances in the global environment. Elsevier, Amsterdam, the Netherlands.

Steinberger, R., and F.H. Westheimer. 1949. The metal ion catalyzed decarboxylation of dimethyloxaloacetic acid. J. Am. Chem. Soc. 71:4158–4159.

Steinberger, R., and F.H. Westheimer. 1951. Metal ion-catalyzed decarboxylation: A model for an enzyme system. J. Am. Chem. Soc. 73:429–435.

Stepniewska, Z., K. Bucior, and R.P. Bennicelli. 2004. The effects of MnO_2 on sorption and oxidation of Cr(III) by soils. Geoderma 122:291–296.

Stevenson, F.J., and M. Schnitzer. 1982. Transmission electron microscopy of extracted fulvic and humic acids. Soil Sci. 133:179–185.

Stollenwerk, K.G., G.N. Breit, A.H. Welch et al. 2007. Arsenic attenuation by oxidized aquifer sediments in Bangladesh. Sci. Total Environ. 379:133–150.

Stone, A.T. 1987. Reductive dissolution of manganese(III)/(IV) oxides by substituted phenols. Environ. Sci. Technol. 21:979–988.

Stone, A.T., and J.J. Morgan. 1984a. Reduction and dissolution of manganese(III) and manganese(IV) oxides by organics. 1. Reaction with hydroquinone. Environ. Sci. Technol. 18:450–456.

Stone, A.T., and J.J. Morgan. 1984b. Reduction and dissolution of manganese(III) and manganese(IV) oxides by organics. 2. Survey of the reactivity of organics. Environ. Sci. Technol. 18:617–624.

Stone, A.T., and A. Torrents. 1995. The role of dissolved metals and metal-containing surfaces in catalyzing the hydrolysis of organic pollutants, p. 275–298. *In* P.M. Huang, J. Berthelin, J.-M. Bollag, W.B. McGill, and A.L. Page (eds.) Environmental impact of soil component interactions, Vol. 1. Natural and anthropogenic organics. CRC Press/Lewis Publishers, Boca Raton, FL.

Strathmann, T.J., and A.T. Stone. 2002. Reduction of oxamyl and related pesticides by Fe-II: Influence of organic ligands and natural organic matter. Environ. Sci. Technol. 36:5172–5183.

Stumm, W. 1992. Chemistry of the solid–water interface. John Wiley & Sons, New York.

Stumm, W., R. Kummert, and L. Sigg. 1980. A ligand exchange model for the adsorption of inorganic and organic ligands at hydrous interfaces. Croat. Chem. Acta 53:291.

Su, C.M., and R.W. Puls. 2008. Arsenate and arsenite sorption on magnetite: Relations to groundwater arsenic treatment using zero valent iron and natural attenuation. Water Air Soil Pollut. 193:65–78.

Szecsody, J.E., D.C. Girvin, B.J. Devary, and J.A. Campbell. 2004. Sorption and oxic degradation of the explosive CL-20 during transport in subsurface sediments. Chemosphere 56:593–610.

Tebo, B.M., H.A. Johnson, J.K. McCarthy, and A.S. Templeton. 2005. Geomicrobiology of manganese(II) oxidation. Trends Microbiol. 13:421–428.

Tennakoon, D.T.B., J.M. Thomas, and M.J. Tricker. 1974. Surface and intercalate chemistry of layer silicates. Part II. An iron-57 Mossbauer study of the role of lattice-substituted iron in the benzidine blue reaction of montmorillonite. J. Chem. Soc. Dalton Trans. 1974:2211–2215.

Theng, B.K.G. 1974. The chemistry of clay-organic reactions. John Wiley & Sons, New York.

Theng, B.K.G. 1979. Formation and properties of clay–polymer complexes. Elsevier Science Publishing, New York.

Theng, B.K.G. 1982. Clay-activated organic reactions, p. 197–238. *In* H. van Olphen and F. Veniale (eds.) Proc. Int. Clay Conf. 1981. Elsevier, Amsterdam, the Netherlands.

Thompson, I.A., D.M. Huber, C.A. Guest, and D.G. Schulze. 2005. Fungal manganese oxidation in a reduced soil. Environ. Microbiol. 7:1480–1487.

Thompson, T.D., and J.F. Moll. 1973. Oxidative power of smectites measured by hydroquinone. Clay. Clay Miner. 21:337–350.

Thorn, K.A., and M.A. Mikita. 2000. Nitrite fixation by humic substances: Nitrogen-15 nuclear magnetic resonance evidence for potential intermediates in chemodenitrification. Soil Sci. Soc. Am. J. 64:568–582.

Tokunaga, T.K., J. Wan, A. Lanzirotti et al. 2007. Long-term stability of organic carbon-stimulated chromate reduction in contaminated soils and its relation to manganese redox status. Environ. Sci. Technol. 41:4326–4331.

Torrents, A. 1992. Hydrolysis of organic esters at the mineral/water interface. Ph.D. Thesis. Johns Hopkins University. Baltimore, MD.

Torrents, A., and A.T. Stone. 1991. Hydrolysis of phenyl picolinate at the mineral/water interface. Environ. Sci. Technol. 25:143–149.

Torres-Canabate, P., E. Davidson, E. Bulygina, R. Garcia-Ruiz, and J. Carreira. 2008. Abiotic immobilization of nitrate in two soils of relic Abies pinsapo-fir forests under Mediterranean climate. Biogeochemistry. 91:1–11.

Tournassat, C., L. Charlet, D. Bosbach, and A. Manceau. 2002. Arsenic(III) oxidation by birnessite and precipitation of manganese(II) arsenate. Environ. Sci. Technol. 36:493–500.

Twigg, M.V. 1989. Catalyst handbook. Wolfe Publishing Ltd., London, U.K.

Ulrich, H.J., and A.T. Stone. 1989. Oxidation of chlorophenols adsorbed to manganese oxide surfaces. Environ. Sci. Technol. 23:421–428.

USEPA (United States Environmental Protection Agency). 2009. Commonly encountered radionuclides. USEPA, Washington, DC. Available online with updates at http://www.epa.gov/radiation/radionuclides/index.html (accessed on August 17, 2009).

Vairavamurthy, A., and S. Wang. 2002. Organic nitrogen in geomacromolecules: Insights on speciation and transformation with K-edge XANES spectroscopy. Environ. Sci. Technol. 36:3050–3056.

Vikesland, P.J., A.M. Heathcock, R.L. Rebodos, and K.E. Makus. 2007. Particle size and aggregation effects on magnetite reactivity toward carbon tetrachloride. Environ. Sci. Technol. 41:5277–5283.

Voudrias, E.A., and M. Reinhard. 1986. Abiotic organic reactions at mineral surfaces, p. 462–486. *In* J.A. Davis and K.F. Hayes (eds.) Geochemical processes at mineral surfaces. ACS Symp. Ser. 323. American Chemical Society, Washington, DC.

Walse, S.S., K.D. Shimizu, and J.L. Ferry. 2002. Surface-catalyzed transformations of aqueous endosulfan. Environ. Sci. Technol. 36:4846–4853.

Wang, S., and W.A. Arnold. 2003. Abiotic reduction of dinitroaniline herbicides. Water Res. 37:4191–4201.

Wang, T.S.C., J.-H. Chen, and W.-M. Hsiang. 1985. Catalytic synthesis of humic acids containing various amino acids and dipeptides. Soil Sci. 140:3–10.

Wang, M.C., and P.M. Huang. 1986. Humic macromolecule interlayering in nontronite through interaction with phenol monomers. Nature 323:529–531.

Wang, M.C., and P.M. Huang. 1987. Polycondensation of pyrogallol and glycine and the associated reactions as catalyzed by birnessite. Sci. Total Environ. 62:435–442.

Wang, M.C., and P.M. Huang. 1991. Nontronite catalysis in polycondensation of pyrogallol and glycine and the associated reactions. Soil Sci. Soc. Am. J. 55:1156–1161.

Wang, M.C., and P.M. Huang. 1992. Significance of Mn(IV) oxide in the abiotic ring cleavage of pyrogallol in natural environments. Sci. Total Environ. 113:147–157.

Wang, M.C., and P.M. Huang. 1994. Structural role of polyphenols in influencing the ring cleavage and related chemical reactions as catalyzed by nontronite, p. 173–180. *In* N. Senesi and T.M. Miano (eds.) Humic substances in the global environment and implications on human health. Elsevier, Amsterdam, the Netherlands.

Wang, M.C., and P.M. Huang. 1997. Catalytic power of birnessite in abiotic formation of humic polycondensates from glycine and pyrogallol, p. 59–65. *In* J. Drozl, S.S. Gonet, N. Senesi, and J. Webber (eds.) Proc. 8th Conf. International Humic Substances Society. Wroclaw, Poland.

Wang, M.C., and P.M. Huang. 2000a. Characteristics of pyrogallol-derived polymers formed by catalysis of oxides. Soil Sci. 165:737–747.

Wang, M.C., and P.M. Huang. 2000b. Ring cleavage and oxidative transformation of pyrogallol catalyzed by Mn, Fe, Al, and Si, oxides. Soil Sci. 165:934–942.

Wang, M.C., and P.M. Huang. 2003. Cleavage and polycondensation of pyrogallol and glycine catalyzed by natural soil clay. Geoderma 112:31–50.

Wang, M.C., and P.M. Huang. 2005. Cleavage of ^{14}C-labelled glycine and its polycondensation with pyrogallol as catalyzed by birnessite. Geoderma 124:415–426.

Wang, T.S.C., P.M. Huang, C.-H. Chou, and J.-H. Chen. 1986. The role of soil minerals in abiotic polymerization of phenolic compounds and formation of humic substances, p. 251–281. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and microbes. SSSA, Madison, WI.

Wang, T.S.C., M.M. Kao, and P.M. Huang. 1980. The effect of pH on the catalytic synthesis of humic substances by illite. Soil Sci. 129:333–398.

Wang, T.S.C., and S.W. Li. 1977. Clay minerals as heterogeneous catalysts in preparation of model humic substances. Z. Pflanzenernahr. Dung. Bodenkd. 140:669–676.

Wang, D.J., J.Y. Shin, M.A. Cheney, G. Sposito, and T.G. Spiro. 1999. Manganese dioxide as a catalyst for oxygen-independent atrazine dealkylation. Environ. Sci. Technol. 33:3160–3165.

Wang, T.S.C., M.C. Wang, and P.M. Huang. 1983. Catalytic synthesis of humic substances by using aluminas as catalysts. Soil Sci. 136:226–246.

Wang, T.S.C., K.L. Yeh, S.Y. Cheng, and T.K. Yang. 1971. Behavior of soil phenolic acids, p. 113–120. *In* Biochemical interaction among plants. National Academy of Science, Washington, DC.

Waychunas, G.A., C.S. Kim, and J.F. Banfield. 2005. Nanoparticulate iron oxide minerals in soils and sediments: Unique properties and contaminant scavenging mechanisms. J. Nanopart. Res. 7:409–433.

Weaver, R.M., M.F. Hochella, and E.S. Ilton. 2002. Dynamic processes occurring at the Cr-aq(III)–manganite (gamma-MnOOH) interface: Simultaneous adsorption, microprecipitation, oxidation/reduction, and dissolution. Geochim. Cosmochim. Acta 66:4119–4132.

Wehrli, B. 1990. Redox reactions of metal ions at mineral surfaces. *In* W. Stumm (ed.) Aquatic chemical kinetics. Wiley-Interscience, New York.

Wehrli, B., and W. Stumm. 1989. Vanadyl in natural waters: Adsorption and hydrolysis promote oxygenation. Geochim. Cosmochim. Acta 53:69–77.

Wehrli, B., E. Wieland, and G. Furrer. 1990. Chemical mechanism in the dissolution kinetics of minerals: The aspect of active sites. Aquat. Sci. 52:3–31.

Wei, J., G. Furrer, S. Kaufmann, and R. Schulin. 2001. Influence of clay minerals on the hydrolysis of carbamate pesticides. Environ. Sci. Technol. 35:2226–2232.

White, A.F., and M.L. Peterson. 1996. Reduction of aqueous transition metal species on the surfaces of Fe(II)-containing oxides. Geochim. Cosmochim. Acta 60:3799–3814.

Wigginton, N.S., K.L. Haus, and M.F. Hochella. 2007. Aquatic environmental nanoparticles. J. Environ. Monit. 9:1306–1316.

Wilk, P.A., D.A. Shaughnessy, R.E. Wilson, and H. Nitsche. 2005. Interfacial interactions between Np(V) and manganese oxide minerals manganite and hausmannite. Environ. Sci. Technol. 39:2608–2615.

Wilkins, R.G. 1991. Kinetics and mechanisms of reactions of transition metal complexes, 2nd Ed. VCH Publishers, Weinheim, Germany.

Williams, L.P. 1965. Michael Faraday. Chapman & Hall, London, U.K.

Williams, A.G.B., and M.M. Scherer. 2004. Spectroscopic evidence for Fe(II)–Fe(III) electron transfer at the iron oxide–water interface. Environ. Sci. Technol. 38:4782–4790.

Wilson, M.A., and K.M. Goh. 1977. Proton-decoupled pulse Fourier-transform ^{13}C nuclear magnetic resonance of soil organic matter. J. Soil Sci. 28:645–652.

Wilson, M.A., N.H. Tran, A.S. Milev, G.S.K. Kannangara, H. Volk, and G.Q.M. Lu. 2008. Nanomaterials in soils. Geoderma 146:291–302.

Wolff, S.P. 1996. Free radicals and glycation theory, p. 73–88. *In* R. Ikan (ed.) The Maillard reaction. Consequences for the chemical and life sciences. John Wiley & Sons, Chichester, U.K.

Wong, J.W., and T. Shibamoto. 1996. Genotoxicity of the Maillard reaction products, p. 129–159. *In* R. Ikan (ed.) The Maillard reaction. Consequences for the chemical and life sciences. John Wiley & Sons, Chichester, U.K.

Yan, L.B., and G.W. Bailey. 2001. Sorption and abiotic redox transformation of nitrobenzene at the smectite–water interface. J. Colloid Interface Sci. 241:142–153.

Yaylayan, V.A. 1997. Classification of the Maillard reaction: A conceptual approach. Trends Food Sci. Technol. 8:13–18.

Zhang, H.C., W.R. Chen, and C.H. Huang. 2008. Kinetic modeling of oxidation of antibacterial agents by manganese oxide. Environ. Sci. Technol. 42:5548–5554.

Zhang, H.C., and C.H. Huang. 2005a. Oxidative transformation of fluoroquinolone antibacterial agents and structurally related amines by manganese oxide. Environ. Sci. Technol. 39:4474–4483.

Zhang, H.C., and C.H. Huang. 2005b. Reactivity and transformation of antibacterial N-oxides in the presence of manganese oxide. Environ. Sci. Technol. 39:593–601.

Zhao, L., Z.Q. Yu, P.A. Peng et al. 2006. Oxidation kinetics of pentachlorophenol by manganese dioxide. Environ. Toxicol. Chem. 25:2912–2919.

Ziechmann, W. 1959. Die Darstellung von Huminsauren im heterogenen System mit neutraler Reaktion. Z. Pflanzenernahr. Dung. Bodenkd. 84:155–159.

<div style="text-align: right">

19

</div>

Soil pH and pH Buffering

Paul R. Bloom
University of Minnesota

Ulf Skyllberg
Swedish University of Agricultural Sciences

19.1 Introduction

Soil pH is a measure of soil acidity or alkalinity and is probably the single most important chemical characteristic of a soil. In the past, soil acidity or alkalinity, reported in pH units, was also referred to as "soil reaction" (Mason and Obenshain, 1939), and this term is occasionally still used in some publications. However, "soil reaction" is no longer used by soil chemists (SSSA, 2008). Because of the importance of acidity and alkalinity in soils, pH has been called a master variable (McBride, 1994), with knowledge of soil pH required to understand many chemical processes including ion mobility, precipitation and dissolution equilibria and kinetics, and oxidation–reduction equilibria. Soil pH also affects nutrient availability to plants and the negative response of many plant species to soil acidity.

Soils are a complex mixture of solid-phase components that react to yield a measured pH value. These components also buffer soils against pH changes caused by natural and anthropogenic inputs of acids and bases. Knowledge of the reactions that buffer soil pH is necessary for an understanding of natural soil weathering and the response of a soil to inputs of lime, acid-forming N fertilizers, acid-mine wastes, and acid rain. The discussion in

this chapter focuses on the reactions of components that buffer soil pH in upland soils. Factors affecting pH and buffering in flooded soils are discussed in Chapter 14.

19.2 Definition and Determination of Soil pH

Because pH is a term that is only defined for solutions, in a strict sense, it cannot be applied to a solid-phase material like soil. However, the chemical properties of the solid-phase components in soil define the pH of the soil solution, the water in soil pores (Chapter 12).

19.2.1 Definition of pH

The term pH was defined by Sorensen (1909) to provide a convenient way of representing the H$^+$ or OH$^-$ concentrations in aqueous solutions. As currently defined, pH is the negative logarithm, base 10, of H$^+$ activity, log (H$^+$), where activity is the concentration adjusted for nonideality caused by charge–charge interactions with other ions in solution. In all except saline soils, the ionic interaction correction is small and the difference between

H^+ concentration and activity is small. The determination of H^+ also provides a measurement of OH^- in solution. At 25°C in pure water, the relation between OH^- activity, (OH^-), and H^+ activity, (H^+), is given by

$$K_w = (H^+)(OH)^- = 1.0 \times 10^{-14} \qquad (19.1)$$

where K_w is the ionization constant for water. Expressing Equation 19.1 as negative logarithms yields

$$pH + pOH = 14 \qquad (19.2)$$

where pOH is the negative logarithm of the OH^- activity.

19.2.2 Determination of pH with Electrodes

Soil pH is determined by measuring the pH after equilibrating soil with pure water or a salt solution (SSSA, 2008). The pH of a solution in equilibrium with a soil varies with the composition and concentration of the salts in the solution because cations in solution displace H^+ and Al^{3+}, an acidic cation, from soil surfaces. Three common standard methods for determining pH in soils involve the suspension is distilled water, 0.01 M $CaCl_2$ or 1 M KCl solutions, and measurement of the solution pH in the solution (Thomas, 1996). In the United States, the most commonly reported pH measurements are in 1:1 (weight:volume) suspensions of air-dried soil in distilled water. A suspension of soil is prepared (e.g., 10 g of soil and 10 mL of distilled water), stirred vigorously, and then allowed to stand for 10 min. The suspension is stirred again and the pH is measured with a glass pH electrode and a pH meter (Thomas, 1996). The pH can also be measured with an ion-sensitive field-effect transistor (ISFET) electrode. These solid-state devices are more robust than glass electrodes, have a more rapid response, and are well adapted for making field measurements of soil pH (Rossel and Walter, 2004). However, the solid-state electrodes are more expensive than glass electrodes.

The determination of soil pH in a 1:1 water suspension is expected to give a higher pH value than the soil solution at field water contents. The quantity of water added results in a dilution of the salts in the soil solution and a lower-equilibrium salt concentration. This effect is illustrated by the 0.4 unit increase in pH obtained for an increase in water content in the soil suspension from 1:1 to 1:10 (Thomas, 1996). The 1:1 water determination of pH is influenced by the natural seasonal variation in soil solution salt concentrations and the effects of soil management; for example, recent fertilization. Soil pH measured in water is reported as $pH_{(H_2O)}$ or pH_w.

At low salt concentrations, an error due to the junction potential effect can occur in determination of $pH_{(H_2O)}$. A pH measurement requires contact with the solution by a glass electrode and a porous liquid junction of a reference electrode. This is necessary to complete the circuit between the pH-sensitive glass electrode and the reference electrode, which is often built in the same body as the glass electrode to form a combination electrode. The filling solution for the reference electrode is a concentrated KCl solution, which slowly leaks into the test solution through a porous liquid junction. Because soils are cation exchangers, they can have a disproportionate effect on the rate of K^+ diffusion compared to Cl^- diffusion into solution, which sets up a potential across the junction (Coleman et al., 1951). This can cause a reduction in the measured pH (Thomas, 1996).

The method for measurement of pH in 0.01 M $CaCl_2$ is the same as for measurement of pH in distilled water (Thomas, 1996). Because of the displacement of H^+ and Al^{3+} from soil materials by Ca^{2+}, pH in 0.01 M $CaCl_2$ [$pH_{(CaCl_2)}$] is on an average about 0.3–0.4 units lower than in water (Essington, 2004). The measurement of pH in 0.01 M $CaCl_2$ results in reduction or elimination of the problems due to variation in soil solution salt concentration and it eliminates the junction potential effect. Calcium is used in the determination of pH because it is the predominant soil solution cation in soils of temperate regions. The 0.01 M concentration, however, is somewhat higher than that found in most soil solutions.

The measurement of pH in 1 M KCl suspensions also generally yields lower pH values than in water. However, for some subsurface horizons of highly weathered soils of the tropics, pH in KCl can actually be greater than in water (van Raij and Peech, 1972). This occurs in subsoils low in permanent charge silicate clays and low in organic matter. Oxides and hydroxides of iron and aluminum are the predominant ion exchangers in these soils and at very acid pH values they are anion exchangers (Section 19.7). The addition of KCl results in a displacement of OH^- by Cl^- and an increase in pH (McBride, 1994).

19.2.3 Determination of Soil pH Using pH-Sensitive Dyes

Rapid determination of soil pH can be made colorimetrically with pH indicator dyes. Colorimetric methods are based on the change in color that takes place upon disassociation of a weak acid or a weak base organic dye. In solution, a weak acid dye molecule can dissociate:

$$HI = H^+ + I^-. \qquad (19.3)$$

An increase in pH will cause a change in color as the undissociated dye, HI, changes to the dissociated form, I^-. Typically a mixture of dyes is used to yield color changes that take place over a wide range of pH. A dye solution is mixed with soil to make a slurry, and the dye is decanted from the soil so the color can be compared with a color chart. The pH measured in this manner corresponds with $pH_{(H_2O)}$ measured with a glass electrode but is not as precise and can vary from the glass electrode results by as much as 0.4 units (Mason and Obenshain, 1939). Dye-impregnated pH papers also can be used to get quick pH measurement within 0.5 units of the true $pH_{(H_2O)}$ (Thomas, 1996).

19.3 Acids and Bases in Soil Solutions

Soil solutions contain a mixture of weak acids and weak bases that buffer the pH of the solution. This buffering is small compared to the quantity of acidity or basicity (alkalinity) associated with the solid phases in soil and has little impact on soil pH buffering. However, the acid and base components in soil solutions directly impact plant roots and soil microbes and thus directly influence the vitality of the biotic community in a soil. Also, the acids and bases in soil solution directly impact mineral weathering and biogeochemical reactions involved in soil genesis.

In this chapter, we will use Brønsted's definition of acids and bases where an acid is a hydrogen ion (H^+) donor and a base is an H^+ acceptor (Stumm and Morgan, 1996). Because an H^+ ion is just a proton, acids can be described as proton donors and bases as proton acceptors. According to this definition, reaction of an acid to lose a proton produces a conjugate base that can accept a proton; if the reaction is reversed. Conversely, a base that accepts a proton yields a conjugate acid. For example, when nitric acid donates protons to water, the base is H_2O. The reaction is

$$HNO_3 + H_2O = NO_3^- + H_3O^+. \quad (19.4)$$
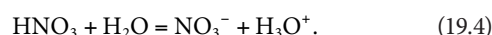
The nitrate anion is the conjugate base of nitric acid and the hydronium ion, H_3O^+, is the conjugate acid of water. Nitric acid is a strong acid because at all except very high H^+ concentrations Reaction 19.4 goes all the way to the right and all of the HNO_3 is ionized. The conjugate base of nitric acid, the nitrate anion, is a very weak base because it has very little tendency to accept protons. The above reaction in aqueous solution is often abbreviated to

$$HNO_3 = NO_3^- + H^+. \quad (19.5)$$
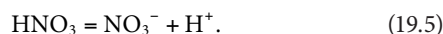
19.3.1 Weak Acids

The weak acids in soil solutions include organic carboxylic acids, H_2CO_3, and Al^{3+}, a metal ion that undergoes hydrolysis. Carboxylic acids are organic compounds with one or more –COOH groups that have the potential to ionize to form the carboxylate anion, $RCOO^-$, according to the reaction

$$RCOOH = RCOO^- + H^+ \quad (19.6)$$
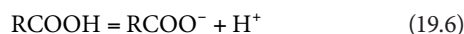
where
  RCOOH is any carboxylic acid
  R is an organic structure

Organic acids in soil solutions include simple acids like malic acid $(COOH)CH_2CH(OH)CH_2(COOH)$ and citric acid $(COOH)CH_2C(OH)(COOH)CH_2(COOH)$, which are exuded from plant roots or by soil microbes. These acids generally have a short lifetime in soils because they are easily consumed by soil microbes. In addition, soil solutions contain macromolecular humified components produced by degradation of plant debris that can be operationally separated into humic and fulvic acids

(Chapter 11; Stevenson, 1994), with fulvic acids predominating in soil solutions. These natural organic acids are complex mixtures that contain carboxylic acid groups.

The strength (ability to donate protons) of a carboxylic acid is defined by the equilibrium constant:

$$K_a = \frac{(RCOO^-)(H^+)}{RCOOH} \quad (19.7)$$

where K_a is the acidity constant. Stronger acids have greater acidity constants. Often, the strength of a weak acid is given in terms of the negative log of the K_a, the pK_a. The lower the pK_a the greater the strength of the acid. At the pH equal to the pK_a, an acid is half ionized. At higher pH, it is more than half ionized, while at lower pH it is less than half ionized. Thus, at solution pH values less than the pK_a, a weak acid has a greater ability to donate protons than at higher pH values. Carboxylic acids in soil solution have pK_a values generally in the range of 3–5.

The soluble fulvic acids also contain phenolic groups, ArOH, where Ar is an aromatic six-carbon ring structure. Phenolic groups are generally much less concentrated than the carboxyl groups and they are much less acidic having pK values ranging from 7 to 11.

The Al^{3+} ion can be an important weak acid component in soil solutions of highly acid soils. This ion contributes to soil solution acidity and causes the acid toxicity response experienced by many plants in acid soils (Bloom et al., 2005; Fageria and Baligar, 2008). In water, Al^{3+} undergoes a series of hydrolysis reactions, which produce protons (Nordstrom and May, 1996):

$$Al^{3+} + H_2O = AlOH^{2+} + H^+ \quad pK_{a1} = 5.00, \quad (19.8)$$

$$AlOH^{2+} + H_2O = Al(OH)_2^+ + H^+ \quad pK_{a2} = 5.1, \quad (19.9)$$

$$Al(OH)_2^+ + H_2O = Al(OH)_3^0 + H^+ \quad pK_{a3} = 6.7, \quad (19.10)$$

$$Al(OH)_3^0 + H_2O = Al(OH)_4^- + H^+ \quad pK_{a4} = 6.2. \quad (19.11)$$
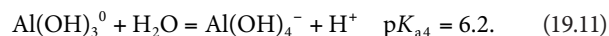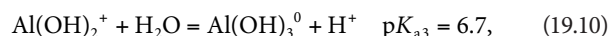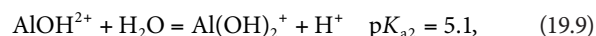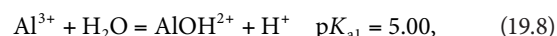
Although the second, third, and fourth hydrolysis reactions are important in determining the chemistry of Al, only the first hydrolysis reaction is a major source of protons in soil solution. The low solubility of solid $Al(OH)_3$ limits the quantity of Al ions. At a pH of 5.0, a solution in equilibrium with $Al(OH)_3$ has a total concentration of Al^{3+} of less than 10^{-6} M (Nordstrom and May, 1996).

In soils, carbonic acid is also an important weak acid (H_2CO_3) that is created by the dissolution of carbon dioxide:

$$CO_{2(g)} + H_2O = H_2CO_3 \quad pK = 1.51 \text{ (Stumm and Morgan, 1996).} \quad (19.12)$$
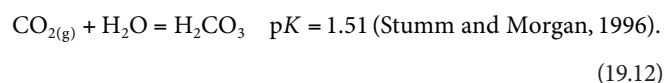
When pH is increased, this acid ionizes to produce bicarbonate and H^+:

$$H_2CO_3 = HCO_3^- + H^+. \tag{19.13}$$
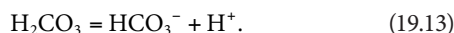
The pK_a for Reaction 19.13 is 6.35 (Stumm and Morgan, 1996) meaning that carbonic acid is a weaker acid than organic carboxylic acids or Al^{3+}. The concentrations of HCO^- in solution increase both with increase in pH and increase of concentration of CO_2 in the air of soil pores. At a pH value above 6, HCO_3^- is often the predominant anion in soil solution.

For solutions in equilibrium with the earth's atmosphere, H_2CO_3 is directly proportional to the partial pressure of atmospheric CO_2, which in 2007 was 3.83×10^{-4} atm. The concentration is increasing at the rate of 0.014×10^{-4} atm year^{-1} (Anon, 2008). However, in soils where root and microbial respiration are active, the partial pressure can easily be 100 times greater (Fernandez and Kosian, 1987; Magnusson, 1992). This has a great effect on the content of HCO_3^- in soils (Andrews and Schlesinger, 2001).

19.3.2 Weak Bases

All of the anions of the weak acids above, as well as the hydrolyzed Al cations, are weak bases that can accept protons in the pH range found in soils. Because of the low pK_a values for organic acids and $AlOH^{2+}$, these ions are only important proton acceptors in very acid soils. For most neutral and alkaline soils, the concentration of bicarbonate determines the alkalinity of soil solutions, where alkalinity is defined as the quantity of strong acid needed to lower the pH of a solution to the end point for complete protonation of bicarbonate (about pH = 4.8) (Stumm and Morgan, 1996). The carbonate ion, CO_3^{2-}, is significant only in very high pH soils because the pK_a for the ionization of HCO_3^- to CO_3^{2-} is 10.33 (Stumm and Morgan, 1996).

Under some conditions, solution ammonia, $NH_{3(aq)}$, is a significant base. These conditions are induced by the addition of fertilizer N as anhydrous ammonia $[NH_{3(g)}]$, or urea. When $NH_{3(g)}$ is injected into a soil, it reacts with the soil solution to lower the H^+ activity raising the pH. Ammonia is a conjugate base of NH_4^+ and if the reaction in water is written like the weak acid reactions above:

$$NH_4^+ = H^+ + NH_{3(aq)}. \tag{19.14}$$
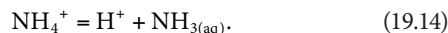
The pK_a of this reaction is 9.5 (Stumm and Morgan, 1996) and in all except the highest pH soils, the reaction goes to the left consuming protons. This reaction creates localized pH values near anhydrous ammonia injection bands that can be >9.5 (Havlin et al., 2005).

Dissolution and hydrolysis of urea fertilizer in soils can also produce localized high concentrations of $NH_{3(aq)}$ that can result in localized high pH. Fertilizer granules hydrolyze readily due to the presence of urease enzymes in soils, resulting in an overall reaction:

$$\underset{\text{Urea}}{(NH_2)_2CO} + 2H_2O = HCO_3^- + NH_4^+ + NH_{3(aq)}. \tag{19.15}$$
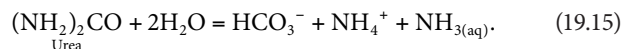
The basicity effect of the NH_3-producing fertilizers is short-lived because biological oxidation of ammonium to nitrates produces two H^+ ions, and the net effect is acidification.

19.4 Overview of Reactions Controlling pH and pH Buffering

Reactions of the solid-phase components control soil pH and are the reactions that buffer soil pH against rapid pH change with natural or anthropomorphic inputs of acids and bases (Bloom et al., 2005). Solid-phase soil materials have a much greater capacity to accept or donate protons than the soluble acids and bases in soil solution. Magdoff and Bartlett (1985) investigated the buffering in 51 surface soils of Vermont, USA by reacting these soils with H_2SO_4 or $CaCO_3$ for 30 days and determining the final pH in 0.01 M $CaCl_2$. They found maximal buffering (minimal slope of pH vs. added acid or $CaCO_3$) to occur at both pH 3.5 or less and pH 7.5 or greater (Figure 19.1). The upper bound of the pH in the buffering shown in Figure 19.1 is fixed by the solubility of $CaCO_3$ (Magdoff and Bartlett, 1985). They also showed that the buffer plot was generally linear in the pH range of 4.5–6.5, a range that encompasses most acid agricultural soils.

The pH-buffering reactions in soils include proton desorption and adsorption reactions with mineral and organic components as well as ion exchange, dissolution, and precipitation. Some of the soil components are effective in buffering over a wide range of pH values, while others are effective over a limited pH range (Table 19.1). The buffer ranges in Table 19.1 are those that were developed by Bloom et al. (2005) and are somewhat similar to those shown
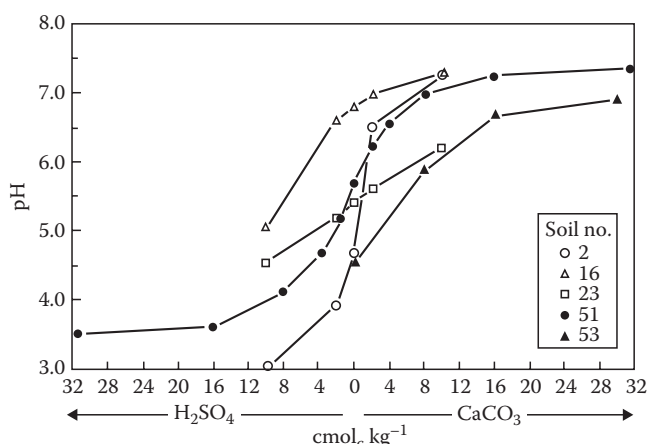
FIGURE 19.1 Titration of surface soils from Vermont, U.S.A, with H_2SO_4 and $CaCO_3$. The acid and base ($CaCO_3$) additions are reported per gram of soils. The pH_{CaCl_2} was measured after a 30-day reaction at field capacity moisture content. (From Magdoff, F.R., and R.J. Bartlett. 1985. Soil pH buffering revisited. Soil Sci. Soc. Am. J. 49:145–148. With permission of Soil Science Society of America.)

TABLE 19.1 Reactions of Solid-Phase Soil Components That Buffer pH in the Range of 3.5–9.5

pH Buffer Substance	pH Range	Proton Acceptor or Donation Reactions
Limestone, $CaCO_3$	7–9.5	Dissolution and precipitation
Oxides and hydroxides of Fe and Al; silicate clay edges	Whole pH range	H^+ adsorption and desorption on surface hydroxyl sites
H^+–SOM	Whole pH range	Dissociation and protonation of carboxyl and phenol groups
Al-SOM and $Al(OH)_{3(s)}$	5–8	Precipitation of organic bound Al^{3+} as $Al(OH)_3$ or dissolution of $Al(OH)_3$ by organic acids
Al^{3+}/H^+-exchange in SOM	Less than 4.5	H^+ exchange with Al^{3+} on carboxyl and phenol groups
$Al(OH)_3$	4–5.5	$Al(OH)_3$ dissolution/precipitation in soils with very low SOM content
Silicate clay interlayer $Al(OH)_n^{3-n}$ in 2:1 clays	4.2–7	Hydrolysis and precipitation or dissolution of interlayer $Al(OH)_3$
Permanent charge silicate clays	3.5–4.2	Ion exchange of H^+ and Al^{3+}
Irreversible dissolution of high-activity 2:1 silicate clays and poorly ordered aluminosilicates	Less than 3.5	Consumes H^+ upon release of Al^{3+} from reactive silicates
Very slow irreversible weathering primary silicate minerals	Whole pH range	Consumes H^+ upon dissolution of Ca^{2+}, Mg^{2+}, K^+, and Na^+ from primary minerals

Source: Bloom, P.R., U.L. Skyllberg, and M.E. Sumner. 2005. Soil acidity, p. 411–460. *In* A. Tabatabai and D. Sparks (eds.) Chemical processes in soils. SSSA, Madison, WI.

by Ulrich (1991), with modifications and additions. For example, Ulrich did not include soil organic matter (SOM) and reactions on oxide and hydrous oxide surfaces. The buffer reactions in Table 19.1 will be discussed in more detail in Sections 19.5 through 19.9.

19.5 Buffering by Soil Organic Matter

Soil organic matter is a very important component of pH buffering in surface soils, even in typical upland soils that contain only a few percent organic matter. Like the soluble organic components in soil solution discussed in Section 19.3.1, SOM contains carboxyl and phenolic groups that can donate protons. Most of the acidity in SOM is contributed by the humidified components, operationally defined as humic and fulvic acid. In soils, some of the weak acid sites exchange protons with Al^{3+}, which alters the proton donation behavior of SOM. Based on chemical analysis of humic and fulvic acids extracted from SOM (Chapter 11) various organic acids have been proposed to represent the acidity of carboxyl and phenolic sites in SOM. Most of the acidity is thought to arise from carboxyl and OH groups bound to aromatic ring structures, which can be represented by benzoic acid (Ar–COOH, pK_a = 4.2) and phenol (Ar–OH, pK_a = 10.0), where Ar is an aromatic ring (Smith et al., 2004).

Soil organic matter buffers pH over a wider range of pH values than predicted by a simple mixture of benzoic acid and phenol. The titration plots for two humic acids with an alkali metal hydroxide shown in Figure 19.2 illustrates that maximum buffering (minimal slope) is in the pH range of 4–6. At very low pH, the plots in Figure 19.2 have very low slopes because of the effect of dilution of acid on H^+ concentration as presented on a log (pH) scale. And at pH > 8.5, the buffering due to phenolic groups contributes to the low slope. A broadening of the buffering response occurs because the aromatic ring structures of the humic acid and fulvic acids are polysubstituted with carboxyl –COOH and phenolic –OH. This shifts the pK_a values, resulting in aromatic ring structures with wide ranges of individual pK_a values. Also, with the ionization of the acid sites, negative charges accumulate on
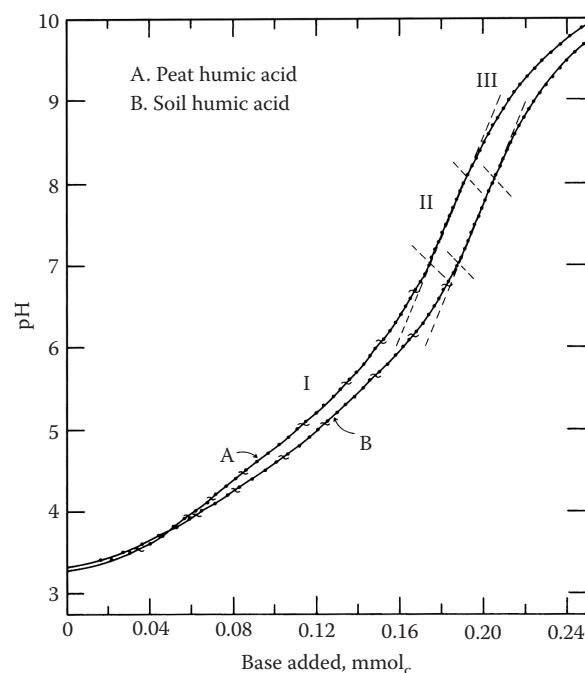
FIGURE 19.2 Titration of 2 humic acids with strong base. (From Stevenson, F.J. 1994. Humus chemistry: Genesis, composition, reactions, 2nd Ed. Wiley, New York. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. With permission.)

macromolecular structures and this charge reduces the proton donation ability of humic and fulvic acids at high pH values (McBride, 1994). The ionization of carboxyl groups is generally complete at pH 8 and the ionization of phenolic groups is complete at pH 11. Thus, common methods for determination of carboxyl acidity involve titrating to pH 8; and for the determination of the phenol titration to pH in excess of 11 (Stevenson, 1994). The negative charge sites produced by the ionization of the weak acid sites on humic and fulvic acids bind exchangeable cations. Organic matter contributes much of the pH-dependent cation exchange capacity (CEC) to soils (Chapter 16).
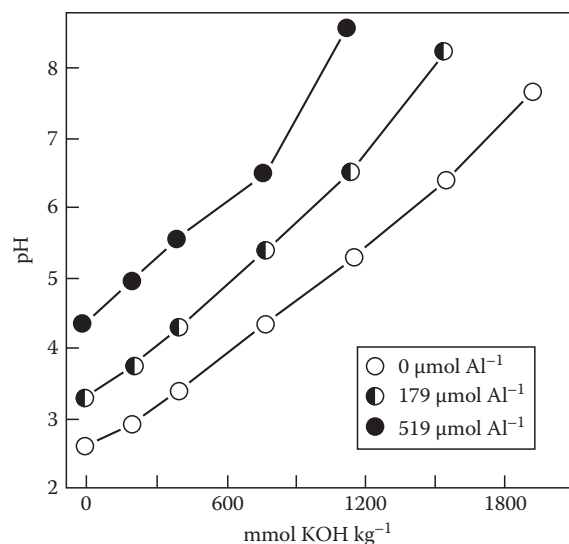
FIGURE 19.3 Titration of Al-SOM containing different contents of Al^{3+}. (From Hargrove, W.L., and G.W. Thomas. 1982. Titration properties of Al-organic matter. Soil Sci. 134:216–225. With permission of Wolters Kluwer Health.)

The binding of Al to acid functional groups in SOM is important both for Al chemistry and for pH buffering. In acid mineral soils, many of the carboxylate sites in SOM are bound to Al^{3+}. This bonding is strong enough that only a small fraction of the Al can be extracted with 1 M KCl (Bloom et al., 1979), and most authors consider Al (and H^+) bound to SOM to be nonexchangeable (Bertsch and Bloom, 1996). This strongly bound Al has a large effect on buffering. Hargrove and Thomas (1982) titrated SOM with varying Al^{3+} contents and found that the pH buffering of the Al^{3+}-treated SOM occurred at higher pH than for H^+-saturated SOM (Figure 19.3). Aluminum-saturated SOM is a weaker acid than H^+-saturated SOM.

As the pH of Al-SOM is raised to greater than about 5, the Al^{3+} in solution exceeds the solubility of $Al(OH)_3$ and precipitation of amorphous $Al(OH)_3$ occurs (Walker et al., 1990). This can be illustrated by

$$2Al(RCOO)_3 + H_2O = 3RCOO^- + 3H^+ + Al(OH)_{3(s)}. \qquad (19.16)$$
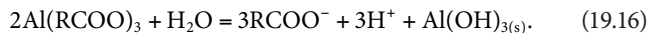
This reaction is reversible and when the pH is lowered $Al(OH)_3$ dissolves and Al again becomes bound to organic matter. As with H^+-SOM, increasing of pH by adding a base increases surface charge and the organic matter becomes more saturated with the cation of the base. In humic materials, except at very low pH, the pH is positively correlated with the saturation of cation exchange sites with the Ca^{2+}, Mg^{2+}, Na^+, and K^+; all cations of strong bases. In soil science, these are called "base cations" and a base saturation (BS) value, which is positively correlated to pH, can be calculated as a sum of charges of base cations divided by the CEC:

$$BS = \frac{2Ca^{2+} + 2Mg^{2+} + Na^+ + K^+}{CEC}. \qquad (19.17)$$

In this calculation, the CEC value is determined at a reference pH of 7 or 8 (Bloom and Grigal, 1985; Ross et al., 2008). In less-acidic and neutral soils, the most abundant base cation is Ca^{2+}.

As soils are acidified, base cations are displaced and the BS value decreases. At pH values less than 5, the BS is low and protons can displace soil adsorbed Al^{3+} ions. This reaction results in a decrease in pH without a decrease in BS because H^+-saturated SOM is a stronger acid than Al-saturated SOM. At pH less than 4.5, this exchange can add very significant quantities of Al^{3+} to soil solution. At these low pH values, where hydrolysis of Al^{3+} is not very significant Al^{3+} is acting more like a base cation. In the high acid A and O horizons typical of many forest soils of northern temperate and humid tropical regions, plots of BS vs. pH show a better positive correlation with BS when Al^{3+} is included as a base cation than if it is included as an acid cation (Skyllberg, 1994, 1999; Ross et al., 2008).

SOM provides much of the pH buffering in most surface soils and much of the CEC. Helling et al. (1964) estimated the mean CEC of SOM in 60 mineral soils of Wisconsin, USA, at pH 8, to be 2000 mmol$_c$ kg^{-1} of SOM. This value is the equivalent to the value for the titration of the acid form of organic matter with $Ca(OH)_2$ to pH 8 and provides a reasonable estimate for the capacity of SOM in mineral soils to buffer pH in the range of 3–8. With the exception of $CaCO_3$, the pH-buffering capacity of SOM is equal or greater than other components in the soil (Table 19.2). Another indication of the importance of SOM for buffering in soils is the results after adding H_2SO_4 and $CaCO_3$ to 51 surface soils (Magdoff and Bartlett, 1985). They took the data, including the four soils in Figure 19.1, and replotted it normalized to equivalent SOM contents. This yielded a single plot (Figure 19.4) for soils of differing organic matter contents suggesting that in these northern temperate region surface soils, the pH buffering is a function mostly of organic matter content.

TABLE 19.2 Approximate Maximum Proton Donation or Adsorption Capacity of Soil Materials in the pH Range of 3.5–8

Soil Material	Capacity, mmol kg^{-1}	References
Silicate clays		
Smectites	800–1,500	McBride (1994)
Vermiculite	1,500–2,000	McBride (1994)
Illite	200–400	McBride (1994)
Kaolinite	10–50	Thomas and Hargrove (1984)
SOM	2,000	Helling et al. (1964)
Allophane and imogolite	200–500	Wada (1989)
Hydroxides and oxides of Fe and Al[a]	50–400	Borggaard (1983)
$CaCO_3$[b]	20,000	Stumm and Morgan (1996)

[a] Based on linear extrapolation between pH 8 and 3.5 of the data for hematite and goethite.

[b] The precipitation and dissolution of $CaCO_3$ takes place at pH 7 or greater.
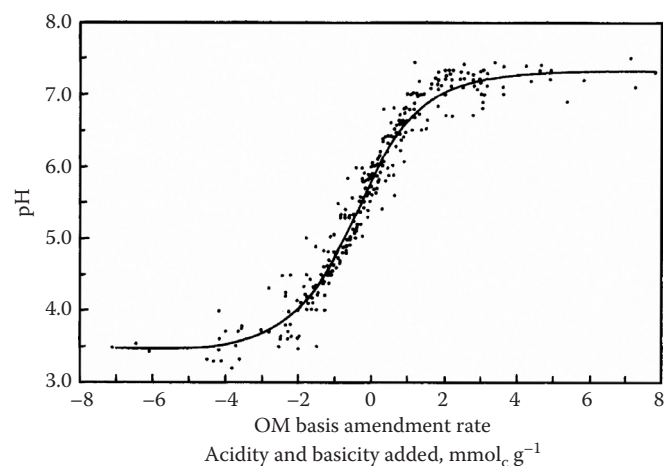
FIGURE 19.4 pH_{CaCl2} after a 30-day reaction of 51 soils from Vermont, U.S.A, with varying additions of H_2SO_4 and $CaCO_3$. The acid and base addition are reported per gram of SOM. The data for all of the soils were adjusted along the x-axis to coincide with the zero addition pH for soil 51 (see Figure 19.1). (From Magdoff, F.R., and R.J. Bartlett. 1985. Soil pH buffering revisited. Soil Sci. Soc. Am. J. 49:145–148. With permission of Soil Science Society of America.)

19.6 Proton and Al Exchange in Silicate Clays

Ion exchange reactions on permanent charge sites in silicate clays can contribute to pH buffering (for a detailed discussion of ion exchange in clays see Chapter 16). The pH buffering by permanent charge on clays can involve exchange of hydronium ions, H_3O^+, but the exchange of Al^{3+} is much more significant.

High-activity clays; smectites, vermiculites, and illites, when saturated with H^+ have very low pH values and are not stable. When montmorillonite, a smectite, is saturated with H^+ the pH is less than 3.5. After 24 h of aging, acid dissolution of the clay structure produces Al^{3+} ions resulting in an increase in pH and replacement of the exchangeable H^+ ions with Al^{3+} (Coleman and Craig, 1961). Aluminum clays are proton donors because of the ability of Al^{3+} on clay exchange sites to hydrolyze and to precipitate as $Al(OH)_3$, but they are weaker proton donors than H^+ clays. Extraction of acid soils with neutral salt solutions, for example, 1 M KCl, yields mostly Al^{3+} not H^+ (Thomas and Hargrove, 1984).

Titration of Al^{3+} montmorillonite (Figure 19.5) shows that at $pH_{(H_2O)}$ values in the range of 5–6, the Al^{3+} on permanent charge sites strongly buffers pH (Turner and Nichol, 1962). Data in Figure 19.5 also show that as in soil, the measured pH in $CaCl_2$ is lower than the pH in water.

When base is added to an Al^{3+}-saturated smectite or vermiculite, the initial reaction is the hydrolysis of Al cations on exchange sites (Bloom et al., 1977), but with sufficient addition of strong base, the Al^{3+} will eventually precipitate as $Al(OH)_3$ in the interlayer space in the clay, accounting for much of the clay-mediated buffering at pH > 5. Because of OH^- deficiencies in interlayer $Al(OH)_3$, the interlayer can neutralize the negative charge on the clay. In the case of smectite, this produces hydroxy interlayer
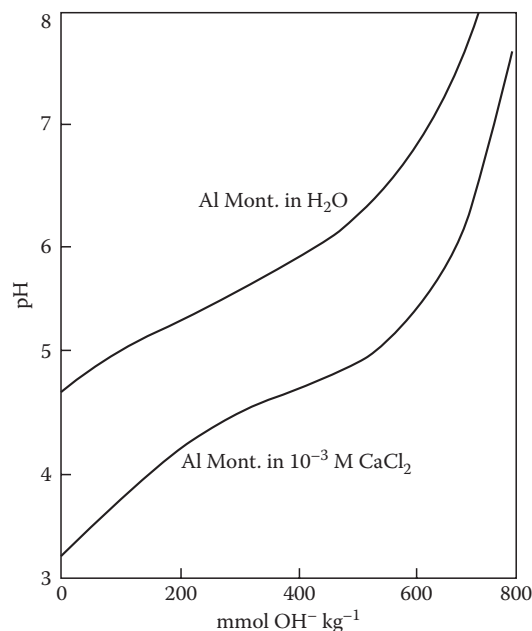
FIGURE 19.5 Titration of an Al saturated montmorillonitic clay with NaOH with and with without the addition 0.001 M $CaCl_2$. (From Turner, R.C., and W.E. Nichol. 1962. A study of the lime potential: 2. Relation between the lime potential and percent base saturation of negatively charged clays in aqueous salt suspensions. Soil Sci. 94:58–63. With permission of Wolters Kluwer Health.)

smectite (HIS), and in the case of vermiculite, this produces hydroxy interlayer vermiculite (HIV) (Chapter 21). Because the precipitation of interlayer $Al(OH)_3$ blocks Al^{3+} in the interlayer from reacting with the base, the titration end point in Figure 19.5 is less than the CEC of the clay, which is 800 $mmol_c$ kg^{-1}.

As with SOM, buffering in Al^{3+}-saturated clays can be described as function of BS, assuming that both H^+ and Al^{3+} acid cations and that Al^{3+} cations represent 3 mol of acidity per mole of Al^{3+}. This is a usable generalization at pH values high enough for Al^{3+} to hydrolyze and act like an acid cation. In contrast, studies of very acidic surface soils from northern temperate forests have shown that at pH values less than 4.5, Al^{3+} should be treated as a base cation (Skyllberg, 1994; Ross et al., 2008).

The capacity of silicate clays to buffer pH varies widely because of the large difference in the CEC of clays (Table 19.2). Kaolinite, a low-activity clay, has little or no permanent charge and has little capacity to buffer pH (Chapter 21). Illite has a much greater CEC and is a much more effective buffer of pH. Smectites and vermiculites are even more effective in buffering pH. Among the silicate clays, only the highest charge vermiculites approach the buffer capacity of SOM.

19.7 pH-Dependent Charge Buffering by Mineral Components

Like SOM, oxides and hydroxides of Fe(III) and Al, and poorly crystalline aluminosilicates, contribute to the pH-dependent charges in soil and to pH buffering. In addition, the edges of

crystalline silicate clays make a minor contribution to pH-dependent charge (Chapter 16.) Unlike SOM, these materials have surfaces that are both proton donors and proton acceptors.

19.7.1 Oxides and Hydroxides of Iron and Aluminum

The oxides and hydroxides of Fe(III) and Al that accumulate in response to weathering are particularly important for pH buffering highly weathered soils (Uehara and Gillman, 1982). Common iron minerals in soil include hematite (α-Fe_2O_3), goethite (α-FeOOH), and ferrihydrite ($Fe_5O_7 \cdot 4H_2O$) (Chapter 22). By far the most common aluminum hydroxide mineral is gibbsite, $Al(OH)_3$. In water, the hydrated surfaces of these oxides and hydroxides contain strongly bound OH groups that can act both as proton donors and proton acceptors. For hematite, the surface hydration reaction is

$$\text{Surface } Fe_2O_3 + 3H_2O = \text{Surface } 2Fe(OH)_3. \qquad (19.18)$$
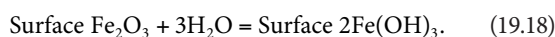
The strong bonding of the O atom to Fe(III) and Al weakens the bond of the proton to the O atom, making the surface-OH groups weakly acidic and at high pH, the metal-bound hydroxide groups can donate proton creating negative surface charges (McBride, 1994):

$$\text{Surface FeOH} = \text{Surface FeO}^- + H^+. \qquad (19.19)$$

This reaction can also be written as an adsorption of OH^-:

$$\text{Surface FeOH} + OH^- = \text{Surface FeO}^- + H_2O. \qquad (19.20)$$
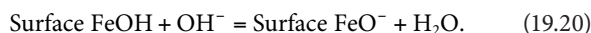
Also these surface-OH groups can accept protons at low pH to become positively charged:

$$\text{Surface FeOH} + H^+ = \text{Surface FeOH}_2^+. \qquad (19.21)$$
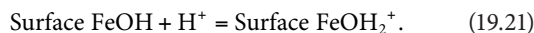
Surface adsorption of H^+ and OH^- is quantified by titrating oxide suspensions in Na^+ or K^+ salts of anions that do strongly adsorb to the oxide surfaces (e.g., NO_3^-). The data of Parks and de Bruyn (1962) illustrate typical titrations for oxide suspensions (Figure 19.6). These data show the effects of increasing KNO_3 concentration on the quantity of H^+ or OH^- adsorbed. McBride (1994) explains this effect, in a KNO_3 solution, as due to K^+ displacement of H^+ at high pH and the NO_3^- displacement of OH^- at low pH. At pH values less than 7, the oxides and hydroxides of Al and Fe are positively charged, and these surfaces are anion exchangers while at pH greater than 9 they are cation exchangers. At pH values in the range of 7–9, a point is found where the concentration of salt in solution has no effect on the pH. This is the point at which the net charge on the surface is zero; the point of zero net charge (pznc, also called the point of zero salt effect). For oxides and hydroxides of iron and aluminum, almost all measured pznc values are in the range of 7–9 (Chapter 15), meaning that
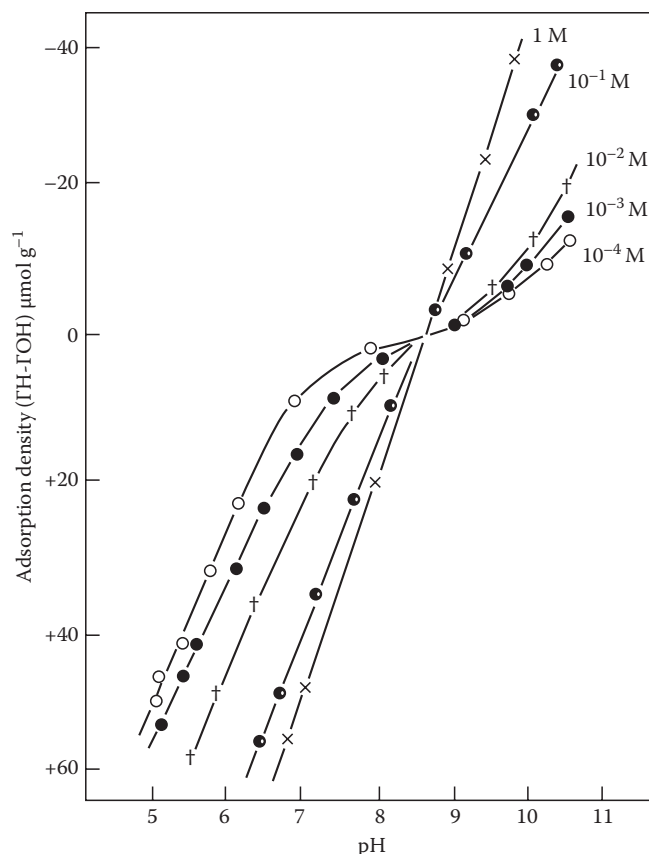
FIGURE 19.6 Titration of Fe_2O_3 with KOH and HNO_3 in suspensions containing different concentrations of KNO_3. (With permission from Parks, G.A., and P.D. de Bruyn. 1962. The zero point of charge of oxides. J. Phys. Chem. 66:967–963. Copyright 1962 American Chemical Society.)

unlike silicate clays and organic matter the net surface charge of these materials is positive in all except alkaline pH soils.

19.7.2 pH-Dependent Charges on Silicate Clay Edges

The layer structure of silicate clays includes an AlOH layer with exposed hydroxyl groups on the edges of the structure (McBride, 1994). These Al hydroxyl groups can adsorb or desorb protons, but because of the influence of the Si^{4+} ions in the structure, the pznc is lower than for gibbsite. For example, a reported value for the pznc for the edge of kaolinite is 2.9 (Chapter 15). This means the edge charge can contribute to the negative charge in soils. For high-charged clays like smectites and vermiculite, this edge charge is small compared to the permanent structural charge. For example, for a smectite at pH 8, the pH-dependent negative charge is about 50 mmol$_c$ kg^{-1} compared to a permanent charge of about 750 mmol$_c$ kg^{-1} (McBride, 1994).

19.7.3 Imogolite and Allophane

Allophane and imogolite are aluminosilicates that like silicate clay edges have pH-dependent charge surfaces. Allophane is noncrystalline, not having long-range ordering in the structure,

and imogolite, which forms long tubes (Chapter 23; Wada, 1989) is a paracrystalline mineral with some disorder in the structure. Both of these materials are found in abundance in young volcanic soils. They are also found in some young soils in glacial parent material (e.g., spodosols in northern forests). Imogolite and allophane have pznc values in the range of 5.5–7 (Clark and McBride, 1984; Wada, 1980).

19.8 Buffering by Dissolution and Precipitation of Carbonates

Carbonates are very important for pH buffering in nonacid soils. Calcite, $CaCO_3$, and dolomite, $CaMg(CO_3)_2$, are common components of soil parent materials in many soils. With sufficient rainfall and good drainage, these minerals weather from soils. In arid regions and in some areas in humid regions with poor drainage, however, $CaCO_3$ can accumulate in soils (Chapter 30; Doner and Lynn, 1989). Under some conditions in arid regions, $MgCO_3$, which is much more soluble than calcite, can accumulate. Calcite acts as proton acceptor according to

$$CaCO_3 + 2H^+ = Ca^{2+} + H_2O + CO_2$$

$$\log K = 9.73 \quad \text{(Stumm and Morgan, 1996)}. \quad (19.22)$$
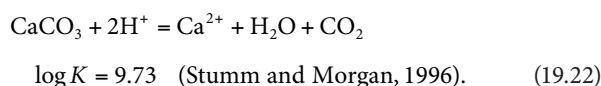
The equilibrium equation for this reaction is

$$K = (Ca^{2+})\frac{(P_{CO_2})}{(H^+)_2}. \quad (19.23)$$

Calculation of this equilibrium requires knowledge of the CO_2 content in the soil atmosphere, expressed in units of partial pressure (P). At 1 atm total pressure, the value of partial pressure is equivalent to the volume fraction in the air. This equilibrium, rewritten in a form that shows how $CaCO_3$ weathers from calcareous soils, is

$$CaCO_3 + H_2CO_3 = Ca^{2+} + 2HCO_3^- \quad \log K = -7.42. \quad (19.24)$$

In soils, the CO_2 produced by soil microbial and root respiration dissolves $CaCO_3$ to produce calcium bicarbonate and over time $CaCO_3$ will weather out of well-drained soils.

The pH in soils that contain calcite varies widely. In a $CaCO_3$ suspension in distilled water in equilibrium with the ambient atmosphere ($P_{CO_2} \approx 3.8 \times 10^{-4}$ atm) will contain two bicarbonate ions in solution for every Ca^{2+} ion and the pH will be 8.3 (Stumm and Morgan, 1996). Because the P_{CO_2} in soils is greatly increased due to microbial and root respiration, soil solution in equilibrium with calcite will have pH values considerably lower than 8.3 (Inskeep and Bloom, 1986). The effect of P_{CO_2} on soil pH varies due to differences in the rate of soil respiration and the rate of gas exchange with the atmosphere. Gas exchange varies with depth and quantity of air-filled pores in a soil, and P_{CO_2} increases with depth and after rainfall fills soil pores with water (Flechard et al., 2007).

Inskeep and Bloom (1986) showed that in pots with growing soybean plants the P_{CO_2} ranged from 0.0006 atm in drier soils to 0.031 atm at moisture contents approaching saturation. By the principle of Le Chatelier (the mass action principle), Reaction 19.22 will be shifted to the left when P_{CO_2} is raised, resulting in an increase in (H^+) and a decrease in pH.

In soils, bicarbonate concentrations are commonly not equal to twice the concentration of Ca^{2+} because there can be other sources of HCO_3^- and Ca^{2+} than dissolution of calcite by carbonic acid. When the chemistry of a calcareous soil is dominated by calcium salts other than bicarbonate, $2[Ca^{2+}]$ can be much greater than $[HCO_3^-]$ and pH is much less than 8.3, even at low P_{CO_2}. Thus, calcareous soils containing gypsum ($CaSO_4 \cdot 2H_2O$), where $[Ca^{2+}]$ is elevated by gypsum solubility, pH can be as low as 7.0. In some arid regions, input of Mg^{2+}, Na^+, and HCO_3^- results in concentrations of HCO_3^- much greater than twice $[Ca^{2+}]$ and calcareous soils can have pH values as high as 9.5. In arid regions, where Mg accumulates, and hydrated $MgCO_3$ forms, pH is greater than 8.5 because magnesium carbonate is much more soluble than $CaCO_3$.

Calcareous soils are very strongly buffered against acidification. One kilogram of $CaCO_3$ can neutralize 20,000 mmol of H^+ and hold pH at a value greater than 7 until all of $CaCO_3$ is dissolved. Thus, $CaCO_3$ is a much more effective buffer than SOM or any other component in soil (Table 19.2). If $CaCO_3$ is compared with SOM on the basis of buffering against a decrease in pH from 8.0 to 7.0, the difference is even greater than shown in Table 19.2. This pH drop would result in the total dissolution of $CaCO_3$ but would only neutralize about 200 mmol$_c$ kg^{-1} of SOM, about one-tenth of the total buffer capacity of SOM. Because of the high buffer capacity of the $CaCO_3$, calcareous soils are considered to be unaffected by acid rain and it is generally considered impractical to artificially lower the pH of calcareous soils to allow for the growth of acid-loving plants.

19.9 H$^+$ Consumption by Irreversible Weathering of Aluminous Minerals

The buffering reactions discussed above are all reversible and can to some degree be described using equilibrium equations. Some reactions that consume H^+ in soils, however, are irreversible. These reactions include the weathering of primary aluminosilicate minerals and the destruction of high-activity clays in very low pH soils.

19.9.1 Weathering of Primary Silicates

Over the very long term, the weathering of primary aluminosilicate minerals from soil parent material consumes protons (Chapter 20). These reactions are irreversible for most primary minerals because these minerals only form at the high temperatures involved in rock-forming processes. An example of the weathering of a primary mineral is the reaction of the feldspar

anorthite to form kaolinite in acid soils. This reaction consumes two protons for every mole of calcium mobilized:

$$\underset{\text{anorthite}}{CaAlSiO_8} + H_2O + 2H^+ \rightarrow \tfrac{1}{2}\underset{\text{kaolinite}}{Al_2Si_2O_5(OH)_4} + Ca^{2+}. \quad (19.25)$$
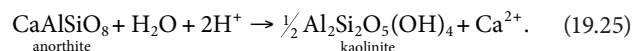
The rates of weathering reactions depend on the type of primary mineral and the surface area exposed to soil solutions (Bloom and Nater, 1991). This means that, per unit mass, silt-sized particles weather much more rapidly than sand-sized particles. The rates of aluminosilicate dissolution reactions are pH dependent and at pH values typical in soils, aluminosilicates dissolve more rapidly at lower pH (Bloom and Nater, 1991; Sposito, 2004). Over time periods of less than decades, weathering of primary silicate minerals is generally not significant in buffering pH.

19.9.2 H⁺ Consumption by the Destruction of High-Activity Clay

When the $pH_{(H_2O)}$ of a soil is decreased to values less than about 3.5, high-activity clays can dissolve quite rapidly consuming protons and releasing Al^{3+} into solution. As discussed in Section 19.6, H⁺-saturated high-activity clays have pH values less than 3.5 and dissolve to release Al^{3+}. At this pH, $Al(OH)_3$ is too soluble to form in soils and kaolinite forms slowly, so high-activity clays tend to dissolve congruently (no solid phase is formed in the reaction). For a Ca^{2+} montmorillonite

$$Ca_{0.167}(Al_{1.67}Mg_{0.333})Si_4O_{10}(OH)_2 + 6H^+ + 4H_2O$$

$$\rightarrow 0.167Ca^{2+} + 0.167Mg^{2+} + 1.67Al^{3+} + 4Si(OH)_4 \quad (19.26)$$
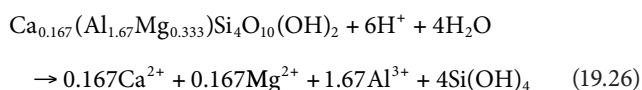
where five of the six protons consumed are due to the release of Al^{3+} from the structure of the clay. This reaction likely accounts for much of the very strong buffering at pH 3.5 that was observed by Magdoff and Bartlett (1985) when they added H_2SO_4 to soils (Figures 19.1 and 19.3).

19.10 Determination of Buffer Capacities

The determination of the capacity of soils to buffer pH has long been an interest of soil chemists. Many crops respond positively to the addition of lime to acid soils, but because of the differences in buffer capacity, soils of similar pH may require vastly different quantities of lime to yield the same increase in pH. Horticulturists have also been concerned with decreasing the pH of nonacid soils and they use aluminum sulfate or elemental S to allow for the vigorous growth of acidophilic plants (e.g., azaleas, rhododendron, and blueberries). As with liming, the quantity of acidifying agent necessary to lower the pH of a soil by a given amount depends on the buffer capacity. Also, acid rain can negatively impact poorly buffered forest soils and the acidifying effects depend on the capacity of soils to buffer the long-term input of very dilute HNO_3 and H_2SO_4 (Bloom and Grigal, 1985; Ulrich, 1991; Watmough and Dillon, 2004).

19.10.1 Total Titratable Acidity

The total acidity in acid soils is operationally defined as consisting of two components, salt-replaceable (exchangeable) and residual (nonexchangeable) acidity (SSSA, 2008). Salt-replaceable acidity is the H⁺ and Al^{3+} extractable with 1 M KCl, while the residual acidity is the acidity that is titratable in a soil suspension but is not easily replaceable with KCl (Bertsch and Bloom, 1996). In practice, the residual acidity is determined by the difference between the total acidity neutralized by raising the pH of a soil to a reference pH, usually 7.0 or 8.0, and the salt-replaceable acidity.

Total acidity can be determined by the titration of a soil suspension in a salt solution to the reference pH using a strong base or by the addition of increments of $CaCO_3$ as shown in Figure 19.1. This produces a plot of the pH vs. the quantity of base consumed, but the response to titration with strong base and to $CaCO_3$ additions is slow, and other laboratory methods are usually used. A standard method for relatively rapid determination of total acidity is to react a soil for several hours or overnight with a solution containing 0.5 M $BaCl_2$ plus a triethanolamine (TEA) buffer adjusted to pH 8.0 or 8.2 (Thomas, 1982). TEA is well buffered at pH 8 and the solution contains Ba^{2+} to displace acidity from soil components. A reference pH of 8.0 or 8.2 is used to represent the pH attained when a soil is limed with excess lime. From a more fundamental point of view, pH 8 is a good choice because this is the end point for the titration of organic carboxyl acidity, as well as the Al^{3+} acidity bound to clays. With the $BaCl_2$–TEA method, the total acidity is calculated as the difference in the quantity of H⁺ needed to titrate the $BaCl_2$–TEA solution to a pH of about 5, using a pH indicator dye, before and after reaction with a soil. This difference represents the quantity of acidity that reacts with the TEA buffer.

An alternative method for the calculation of total acidity is the difference between the CEC determined at pH 7.0 with ammonium acetate and the sum of exchangeable bases. Determination of CEC at pH 7.0 is the method most often used in the classification of soils (Burt, 2004). In this method, soil is reacted with an ammonium acetate solution at pH 7.0, the excess ammonium acetate is leached from the soil, and the NH_4^+ ions bound to cation exchange sites are extracted with a salt solution and quantified (Sumner and Miller, 1996; Burt, 2004). Subtraction of the exchangeable bases from the pH 7.0 CEC represents the total acidity referenced to pH 7.0.

19.10.2 Lime Requirement

Development of the recommendations to growers for the liming of an acid soil with ground calcitic or dolomitic limestone (agricultural lime) requires a quick and easy method for the

determination of the quantity of base needed to raise the soil pH to an optimum value for a given crop (Bloom et al., 2005; Fageria and Baligar, 2008). Because crops differ in sensitivity to soil acidity, recommendations for liming may differ with crop. The most common methods used in the United States involve the reaction of a mixed buffer solution with a soil and after a period of reaction, measurement of the equilibrium pH (Godsey et al., 2007). The commonly used Mehlich method uses a mixed buffer solution containing acetate, glycerophosphate and TEA, in $BaCl_2$, that provides for a linear buffer response to acidity at pH values ranging from 4 to 6.5 (Mehlich, 1976; Hoskins and Erich, 2008). This solution has been modified by replacing the Ba with Ca to avoid the use of toxic substances (Godsey et al., 2007; Hoskins and Erich, 2008). The pH of the buffer mixture is adjusted to 6.5 and the lime requirement is determined from the pH after reaction with a soil using regression equations for different target pH values. At lower equilibrium pH, the lime requirement is greater. The regression equations are developed using the pH response of acid soils to known additions of $CaCO_3$ (Hoskins and Erich, 2008). Hoskins and Erich (2008) showed that the inclusion of a soil $pH_{(H_2O)}$ term increased the correlation coefficient of the regression. The Shoemaker–McLean–Pratt (SMP) mixed buffer method is also commonly used (McLean et al., 1977), but this method is decreasing in acceptance because it contains two toxic buffers, *p*-nitrophenol and chromate, and tends to cause plugging of liquid junctions of pH electrodes (Hoskins and Erich, 2008). Other mixed buffer methods for determination of lime requirement include those of Adams and Evans (1962), Woodruff (1948), and Nômmik (1983).

In areas of the world where highly weathered soils are dominant, it is common to base lime recommendations on the reduction of Al^{3+} saturation of the effective CEC, calculated as the sum of bases plus 1 M KCl extractable acidity (Fageria and Baligar, 2008). When lime is added, Al^{3+} precipitates to form insoluble $Al(OH)_3$, decreasing Al^{3+} toxicity. To determine the lime requirement, the 1 M KCl replaceable Al^{3+} acidity is calculated as three times the molar quantity of extracted Al and this acidity, expressed in units of $cmol_c$ kg^{-1}, is multiplied by a factor to calculate the lime requirement in Mg ha^{-1}. The factor varies from 1, or less, in sandy soils to greater than 3 in very clayey soils. Also an adjustment for exchangeable Ca^{2+} and Mg^{2+} is used to yield a lime requirement, even for low acidity soils, if a soil is low in Ca and Mg. These two ions are often deficient for plant growth in highly weathered soils (Fageria and Baligar, 2008).

19.10.3 Calcium Carbonate Concentrations in Soil

The quantity of calcium carbonate in soils is generally determined by methods that measure the CO_2 evolved after adding acid or the weight loss due to CO_2 loss. In the weight loss method, carefully weighed soil samples are added to weighed flasks containing 3 M HCl and loosely covered with a rubber stopper (Loeppert and Suarez, 1996). The flask is swirled and uncovered periodically until all of the CO_2 is evolved and the

flask is carefully reweighed. The content of carbonate is calculated assuming all of the carbonate is $CaCO_3$ with a formula weight of 100 g mol^{-1}. For soils that have less than 2% or 3% carbonate, the weight loss method is not accurate (U.S. Salinity Laboratory Staff, 1954; Loeppert and Suarez, 1996).

Determination of the evolved CO_2 provides a more accurate and precise measure of calcium carbonate (Loeppert and Suarez, 1996). Commonly used methods involve the determination of the pressure increase in a closed cell using an Hg manometer or pressure transducer CO_2 (Williams, 1949; Martin and Reeve, 1955; Skinner et al., 1959; Evangelou et al., 1984) or the gravimetric determination of CO_2 collected in a magnesium perchlorate trap (Loeppert and Suarez, 1996).

19.10.4 Determination of the Buffering of H+ Inputs in Noncalcareous Soils

Routine laboratory methods have not been developed for the determination of the capacity of soils to buffer against acidification. However, addition of increments of acid with long-term incubation, as used by Magdoff and Bartlett (1985), can be used to determine the response of soils to acid additions (Figures 19.1 and 19.4). An alternative method is to equilibrate soil suspensions repeatedly with dilute acid. Bloom and Grigal (1985) suspended soils in 2.5×10^{-4} M H_2SO_4 for 24 h decanted and repeated 13 times. By measuring the pH of the equilibrated solutions, they were able to determine the quantity of protons adsorbed due to pH buffering in the soil. They were also able to determine the Al^{3+} and base cation losses in each treatment. This method much more closely simulates the response to acidic deposition (acid rain) with the resultant loss of bases and Al^{3+}. Bloom and Grigal (1985) used their plots of pH response to acid to test a numerical model for the response of soils to acidic deposition.

19.11 Soil Acidification

Soil acidification occurs as a natural process during soil formation. With sufficient inputs of acid rain to poorly buffered soils, natural acidification can be very significantly accelerated (Bloom and Grigal, 1985; Ulrich, 1991). Much more rapid acidification is produced by the addition of ammonium-forming nitrogen fertilizers. Even more rapid acidification is produced by the oxidation of sulfur and sulfur compounds added to soils or of sulfur compounds deposited by natural processes in coastal wetland soils that are exposed to oxidation when the water table is lowered by artificial drainage.

19.11.1 Natural Acidification

Under conditions of good drainage and rainfall sufficient to produce leaching, H_2CO_3 and soluble organic acids cause the very slow acidification of soils that occurs during natural soil weathering. At pH values of about 6 and greater, the donation of protons by the reaction of H_2CO_3 to produce HCO_3^- (Reaction 19.13) is the predominant acidifying reaction. As the pH is

decreased to values much lower than the pK_a (6.35), H_2CO_3 is a less effective proton donor and organic acids, having lower pK_a values, are the predominant proton donors. In calcareous soils, $CaCO_3$ dissolution buffers the pH at values greater than 7 until it is all leached from the soil.

19.11.2 Acid Rain

The dry aerosol and wet acid deposition that makes up what is popularly called acid rain contains sulfuric and nitric acid plus NH_4^+ that can be oxidized in soil to produce nitric acid (Sparks, 2003). Under preindustrial conditions, these acids were of minor importance in determining the composition of soil solutions. Industrialization has resulted in elevated strong-acid acidity in rainfall, which can have long-term negative impacts on poorly buffered soils in forests and other natural areas (Bloom and Grigal, 1985; Ulrich, 1991). Since the concentration of acidic components of acidic deposition is small compared to soil buffering, even in poorly buffered soils, soil acidification is difficult to document by field-based observations and contradictory results have been reported (Courchesne et al., 2005). One of the major difficulties is the spatial variability in soil properties (Likens et al., 1998). Studies that have shown significant soil acidification that appears to be due to acid rain include the work of Courchesne et al. (2005), who reported a statistically significant decrease in pH in the humus layer of soil collected from 1993 to 2002 and Watmough and Dillon (2004), who reported a depletion of exchangeable Ca for soil collected from 1983 to 1999. Both of these studies were conducted in forests of eastern Canada.

19.11.3 Acid-Producing Fertilizers

Nitrogen fertilizers that produce NH_4^+ in soils; ammonium sulfate, ammonium nitrate, anhydrous ammonia, and urea, acidify soils. In soil, bacteria oxidize NH_4^+ to nitrate producing H^+ (Chapter 23; Havlin et al., 2005):

$$NH_4^+ + O_2 \rightarrow NO_4^- + 2H^+ + 2H_2O. \tag{19.27}$$
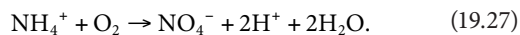
This reaction can produce sufficient acidification to require periodic addition of lime to cropland (Adams, 1984).

19.11.4 Acidification by Sulfur Compounds

Sulfuric acid produced in acid mine wastes or when the soils of coastal swamps are drained, can have a large impact on soils. When pyrite or iron sulfide associated with a seam of coal or a coastal organic soil is exposed to air, bacterial oxidation results in the production of sulfuric acid (McElnea et al., 2002; Sparks, 2003):

$$\underset{\text{pyrite}}{FeS_2} + {}^{15}\!/_{4}O_2 + {}^{7}\!/_{2}H_2O = Fe(OH)_3 + 2H_2SO_4. \tag{19.28}$$
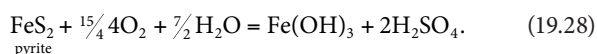
Under these conditions, soil pH values of less than 3 are possible. This reaction has produced large areas of acid sulfate soils in Vietnam, Malaysia, and Indonesia (Sanchez, 1976).

Acidification of nonacid soils for the growth of acid-loving (acidophilic) plants can be accomplished using elemental S. Bacterial oxidation produces a similar effect as for the oxidation of pyrite (Slaton, 2001):

$$S + {}^{3}\!/_{2}O_2 + H_2O = H_2SO_4. \tag{19.29}$$

Growers using elemental S must be very careful because pH values of 3.5 are easily attainable when excess sulfur produces too much sulfuric acid (Leitzke and Peterson, 1987) (Figure 19.4).

An alternative method is to add aluminum sulfate, $Al_2(SO_4)_3$, which produces acidity according to the reaction:

$$Al_2(SO_4)_3 + 6H_2O = 2Al(OH)_3 + 3H_2SO_4. \tag{19.30}$$
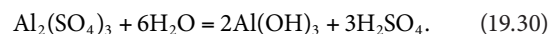
The lower pH generated by aluminum sulfate is limited by the solubility of $Al(OH)_3$ and pH will not be decreased below about 4.5 but excessive aluminum sulfate can induce Al^{3+} toxicity problems for plant growth (Leitzke and Peterson, 1987).

References

Adams, F. 1984. Crop response to lime in the southern United States, p. 211–265. *In* F. Adams (ed.) Soil acidity and liming. No. 12 Agronomy. ASA, Madison, WI.

Adams, F., and C.E. Evans. 1962. A rapid method for measuring of lime requirement of red–yellow podzolic soils. Soil Sci. Soc. Am. Proc. 26:355–357.

Andrews, J.A., and W.H. Schlesinger. 2001. Soil CO_2 dynamics, acidification, and chemical weathering in a temperate forest with experimental CO_2 enrichment. Global Biogeochem. Cy. 15:149–162.

Anon. 2008. State of the climate in 2007. Bull. Am. Meteorol. Soc. 89:S6–S102.

Bertsch, P.M., and P.R. Bloom. 1996. Aluminum, p. 517–549. *In* D.L. Sparks (ed.) Methods of soil analysis. Part 3. Chemical methods. SSSA, Madison, WI.

Bloom, P.R., and D.F. Grigal. 1985. Modeling soil response to acidic deposition in non-sulfate adsorbing soils. J. Environ. Qual. 14:481–495.

Bloom, P.R., M.B. McBride, and B. Chadbourne. 1977. Adsorption of aluminum by smectite: I. Surface hydrolysis during Ca^{2+}–Al^{3+} exchange. Soil Sci. Soc. Am. J. 41:1068–1073.

Bloom, P.R., M.B. McBride, and R.M. Weaver. 1979. Aluminum organic matter in acid soils: Salt extractable aluminum. Soil Sci. Soc. Am. J. 43:813–815.

Bloom, P.R., and E.A. Nater. 1991. Kinetics of dissolution of oxide and silicate minerals, p. 151–189. *In* D.L. Sparks and D.L. Suarez (eds.) The kinetics of physicochemical processes in soils. SSSA, Madison, WI.

Bloom, P.R., U.L. Skyllberg, and M.E. Sumner. 2005. Soil acidity, p. 411–460. *In* A. Tabatabai and D. Sparks (eds.) Chemical processes in soils. SSSA, Madison, WI.

Borggaard, O.K. 1983. Effect of surface area and mineralogy of iron oxides on their surface. Clay. Clay Miner. 31:230–232.

Burt, R. (ed.). 2004. Soil survey laboratory methods manual. Soil survey investigations. Report No. 42. Version 4.0. Available at ftp://ftp-fc.sc.egov.usda.gov/NSSC/Lab_Methods_Manual/SSIR42_2004_view.pdf (accessed on April 26, 2011)

Clark, C.J., and M.B. McBride. 1984. Cation and anion retention by natural and synthetic allophane and imogolite. Clay. Clay Miner. 32:291–299.

Coleman, N.T., and D. Craig. 1961. The spontaneous alteration of hydrogen clay. Soil Sci. 91:14–18.

Coleman, N.T., D.E. Williams, T.R. Nielson, and H. Jenny. 1951. On the validity of interpretations of potentiometrically measured soil pH. Soil Sci. Soc. Am. Proc. 15:106–110.

Courchesne, F., B. Cote, J.W. Fyles, W.H. Hendershot, P.M. Biron, A.G. Roy, and M.C. Turmel. 2005. Recent changes in soil chemistry in a forested ecosystem of Southern Quebec, Canada. Soil Sci. Soc. Am. J. 69:1298–1313.

Doner, H.E., and Lynn, W.E. 1989. Carbonate, halide, sulfate, and sulfide minerals, p. 279–330. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. Soil Sci. Soc. Am., Madison, WI.

Essington, M.E. 2004. Soil and water chemistry. CRC press, Boca Raton, FL.

Evangelou, V.P., L.D. Whittig, and K.K. Tanji. 1984. An automated manometric method for the quantitative determination of calcite and dolomite. Soil Sci. Soc. Am. J. 48:1236–1239.

Fageria, N.K., and V.C. Baligar. 2008. Ameliorating soil acidity of tropical oxisols by liming for sustainable crop production. Adv. Agron. 99:345–399.

Fernandez, I.J., and P.A. Kosian. 1987. Soil air carbon dioxide concentrations in a New England spruce-fir forest. Soil Sci. Soc. Am. J. 51:261–263.

Flechard, C.R., A. Neftel, M. Jocher, C. Ammann, J. Leifeld, and J. Fuhrer. 2007. Temporal changes in soil pore space CO_2 concentration and storage under permanent grassland. Agric. For. Meteorol. 142:66–84.

Godsey, C.B., G.M. Pierzynski, D.B. Mengel, and R.E. Lamond. 2007. Evaluation of common lime requirement methods. Soil Sci. Soc. Am. J. 71:843–850.

Hargrove, W.L., and G.W. Thomas. 1982. Titration properties of Al-organic matter. Soil Sci. 134:216–225.

Havlin, J.L., J.D. Beaton, S.L. Tisdale, and W.L. Nelson. 2005. Soil fertility and fertilizers, 7th Ed. Prentice Hall, Upper Saddle River, NJ.

Helling, C.S., G. Chesters, and R.B. Corey. 1964. Contribution of organic matter and clay to soil cation-exchange capacity as affected by the pH of the saturating solution. Soil Sci. Soc. Am. Proc. 28:517–520.

Hoskins, B.R., and M.S. Erich. 2008. Modification of the Mehlich lime buffer test. Commun. Soil Sci. Plant Anal. 39:2270–2281.

Inskeep, W.P., and P.R. Bloom. 1986. Kinetics of calcite precipitation in the presence of water soluble organic ligands. Soil Sci. Soc. Am. J. 50:1167–1172.

Leitzke, D.A., and D.V. Peterson. 1987. Effects of soil acidification and chemical and mineralogical properties of a limed soil. Soil Sci. Soc. Am. J. 51:620–625.

Likens, G.E., C.T. Driscoll, D.C. Buso, T.G. Siccama, C.E. Johnson, G.M. Lovett, W.A. Reiners, D.F. Ryan, C.W. Martin, and S.W. Bailey. 1998. The biogeochemistry of calcium at Hubbard Brook. Biogeochemistry 41:89–173.

Loeppert, R.H., and D.L. Suarez. 1996. Carbonate and gypsum, p. 437–474. *In* D.L. Sparks (ed.) Methods of soil analysis. Part 3. Chemical methods. SSSA, Madison, WI.

Magdoff, F.R., and R.J. Bartlett. 1985. Soil pH buffering revisited. Soil Sci. Soc. Am. J. 49:145–148.

Magnusson, T. 1992. Studies of the soil atmosphere and related physical site characteristics in mineral forest soils. J. Soil Sci. 43:767–790.

Martin, S.E., and R. Reeve. 1955. A rapid manometric method for the determining soil carbonate. Soil Sci. 79:187–197.

Mason, D.D., and S.S. Obenshain. 1939. A comparison of methods for the determination of soil reaction. Soil Sci. Soc. Am. Proc. 3:129–137.

McBride, M.B. 1994. Environmental chemistry of soils. Oxford University Press, New York.

McElnea, A.E., C.R. Ahern, and N.W. Menzies. 2002. Improvements to peroxide oxidation methods for analysing sulfur in acid sulfate soils. Aust. J. Soil Res. 40:1115–1132.

McLean, E.O., J.F. Terwiiler, and D.J. Eckert. 1977. Improved SMP buffer method for determination of lime requirement of acid soils. Commun. Soil Sci. Plant Anal. 8:667–675.

Mehlich, A. 1976. A new buffer pH method for rapid estimation of exchangeable acidity and lime requirement of soils. Commun. Soil Sci. Plant Anal. 7:637–652.

Nômmik, H. 1983. A modified procedure for the rapid determination of titratable acidity and lime requirements in soils. Acta Agric. Scand. 33:337–348.

Nordstrom, D.K., and H.M. May. 1996. Aqueous equilibrium data for mononuclear aluminum species, p. 30–80. *In* G. Sposito (ed.) The environmental chemistry of aluminum. Lewis Publishers, Boca Raton, FL.

Parks, G.A., and P.D. de Bruyn. 1962. The zero point of charge of oxides. J. Phys. Chem. 66:967–963.

Ross, D.S., G. Matschonat, and U. Skyllberg. 2008. Cation exchange in forest soils: The need for a new perspective. Eur. J. Soil Sci. 59:1141–1159.

Rossel, R.A.V., and C. Walter. 2004. Rapid, quantitative and spatial field measurements of soil pH using an ion sensitive field effect transistor. Geoderma 119:9–20.

Sanchez, P.A. 1976. Properties and management of soils in the tropics. Wiley, New York.

Skinner, S.I.M., R.L. Halstead, and J.E. Brydon. 1959. Quantitative manometric determination of calcite and dolomite in soils and limestones. Can. J. Soil Sci. 39:197–204.

Skyllberg, U. 1994. Aluminum associated with a pH-increase in the humus layer of a boreal haplic podzol. Interciencia 19:356–365.

Skyllberg, U. 1999. pH and solubility of aluminium in acidic soils: A consequence of reactions between organic acidity and aluminium alkalinity. Eur. J. Soil Sci. 50:95–106.

Slaton, N.A., R.J. Norman, and J.T. Gilmour. 2001. Oxidation rates of commercial elemental sulfur products applied to an alkaline silt loam from Arkansas. Soil Sci. Soc. Am. J. 65:239–243.

Smith, R.M., A.E. Martell, and R.J. Motekaitis. 2004. NIST critically selected stability constants of metal complexes database. National Institute of Standards and Technology, U.S. Department of Commerce, Washington, DC.

Sorensen, S.P.L. 1909. Enzyme studies: II. The measurement and importance of the hydrogen ion concentration in enzyme reaction. Compt. Rend. Trav. Lab. (Carlsberg) 8:1.

Sparks, D.L. 2003. Environmental soil chemistry. Academic Press, New York.

Sposito, G. 2004. The surface chemistry of natural particles. Oxford University Press, New York.

SSSA (Soil Science Society of America). 2008. Glossary of soil science terms. SSSA, Madison, WI.

Stevenson, F.J. 1994. Humus chemistry: Genesis, composition, reactions, 2nd Ed. Wiley, New York.

Stumm, W., and J.J. Morgan. 1996. Aquatic chemistry: Chemical equilibria and rates in natural waters, 3rd Ed. Wiley, New York.

Sumner, M.E., and W.P. Miller. 1996. Cation exchange capacity and exchange coefficients, p. 1201–1229. *In* D.L. Sparks (ed.) Methods of soil analysis. Part 3. Chemical methods. SSSA Book Series 5. SSSA, Madison, WI.

Thomas, G.W. 1982. Exchangeable acidity, p. 159–165. *In* Methods of soil analysis. Part 2. Chemical methods. Chemical and microbiological properties. ASA Monograph 9. ASA, Madison, WI.

Thomas, G.W. 1996. Soil pH and soil acidity, p. 475–490. *In* Methods of soil analysis. Part 3. Chemical methods. SSSA, Madison, WI.

Thomas, G.W., and W.L. Hargrove. 1984. The chemistry of soil acidity, p. 3–56. *In* F. Adams (ed.) Soil acidity and liming. 2nd Ed. ASA, Madison, WI.

Turner, R.C., and W.E. Nichol. 1962. A study of the lime potential: 2. Relation between the lime potential and percent base saturation of negatively charged clays in aqueous salt suspensions. Soil Sci. 94:58–63.

Uehara, G., and G. Gillman. 1982. The mineralogy, chemistry and physics of tropical soils with variable charge clays. Westview Press, Boulder, CO.

Ulrich, B. 1991. An ecosystem approach to soil acidification, p. 28–79. *In* B. Ulrich and M.E. Sumner (eds.) Soil acidity. Springer-Verlag, Berlin, Germany.

U.S. Salinity Laboratory Staff. 1954. Diagnosis and improvements of saline and alkali soils. USDA Handbook 60. U.S. Government Printing Office, Washington, DC.

Van Raij, B., and M. Peech. 1972. Electrochemical properties of some oxisols and alfisols of the tropics. Soil Sci. Soc. Am. Proc. 36:446–451.

Wada, K. 1980. Mineralogical characteristics of andosols, p. 87–107. *In* B.K.G. Theng (ed.) Soils with variable charge. New Zealand Soil Science Society, Lower Hutt, New Zealand.

Wada, K. 1989. Allophane and imogolite, p. 1051–1087. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Walker, W.J., C.S. Cronan, and P.R. Bloom. 1990. Aluminum solubility in organic soil horizons from northern and southern forested watersheds. Soil Sci. Soc. Am. J. 54:369–374.

Watmough, S.A., and P.J. Dillon. 2004. Major element fluxes from a coniferous catchment in central Ontario, 1983–1999. Biogeochemistry 67:369–398.

Williams, D.E. 1949. A rapid manometric method of the determination of carbonate in soil. Soil Sci. Soc. Am. Proc. 13:127–129.

Woodruff, C.M. 1948. Testing of lime requirement by means of a buffer solution and the glass electrode. Soil Sci. 66:53–63.

III

Soil Mineralogy

Joseph W. Stucki
University of Illinois

FUNDAMENTAL TO THE EXISTENCE of soil is its inorganic mineral fraction. Indeed, of all soil constituents only this one is required. Without it, soil ceases to be soil. Organic matter, bacteria, and fertilizer chemicals greatly enhance various properties of the soil, but without the inorganic minerals, it is reduced to a soilless medium incapable of the diverse range of functions that are often taken for granted. A correct and complete understanding of the soil requires, therefore, an understanding of soil mineralogy.

A common misconception is that soil minerals are rather inert, unchanging, and nonlabile. The chapters in this part quickly dispel this view by painting a very different picture. They will show that the soil minerals are not only diverse in their origins, structures, and chemistry, but they provide the backbone for active chemical surfaces where many reactions are catalyzed and basic behaviors of the soil are born and changed. As described here and elsewhere in this book, these active surfaces

of the so-called inert fraction react with organic matter, invoking mutual alterations in the behavior of both the minerals and the organic matter. Life-sustaining liquid water is held at their surfaces and in their interstitial pores, and even retained in the liquid state at very low temperatures to preserve plant life during hard winter months. Oxidation–reduction reactions alter the chemical and physical properties of the constituent Fe- and Mn-bearing minerals, thus creating real-time transformations that are critical to many soil processes. Soil minerals come in a wide range of particle sizes and morphologies, which add texture and body to the soil, and vary in their degree of crystallinity. Their colors are also diverse, ranging from intense reddish brown to very light gray, or even green, yellow, or blue. They are dynamic in their properties, constantly changing with climate, time, and other environmental conditions. Some changes are rapid; others, slow. These attributes bring the soil to life, as it were, and create the framework within which the plant and

animal kingdoms spring forth and are sustained through every season and in every clime.

Minerals, the heart of the soil, are a dynamic and essential resource, classified according to their properties. Presented in the following chapters are descriptions of how minerals are formed and transformed. Specifics are given regarding the properties and behavior of the phyllosilicates, the plate-shaped minerals; iron, aluminum, and other (oxyhydr)oxides, which represent the more highly weathered mineral constituents; and amorphous minerals, which display active chemical properties but lack high order in their crystals. These are among the most important of the soil minerals. As such, knowledge of their characteristics will provide the basis from which a more complete understanding of soil physical and chemical properties may be acquired.

Alteration, Formation, and Occurrence of Minerals in Soils

G. Jock Churchman
University of Adelaide

David J. Lowe
University of Waikato

20.1 Introduction

This chapter, like Churchman (2000), seeks to bring readers up to date with information and understanding about the alteration of minerals and the nature of their products in the context of the formation and development of soils. It complements the articles by Bergaya et al. (2006), and recent books by Velde and Meunier (2008) and Velde and Barré (2010). This current chapter differs from Churchman (2000) in that it discusses the manner in which minerals, and especially secondary minerals, actually occur in soils, that is, it deals (in Section 20.3) with the occurrence, as well as the alteration and formation, of minerals in soils.

Secondary minerals are the most reactive inorganic materials in soils. Furthermore, they occur commonly in association with the most reactive organic materials in soils. It has often been inferred that knowledge of the properties of these reactive materials should enable close predictions of the useful soil properties, whether for growing plants, for filtering and partitioning water flow, for immobilizing contaminants, for supporting human-made structures, or for other purposes. However, most of the information about the formation of secondary minerals through the alteration of primary minerals derives from studies of largely inorganic processes taking place in relatively "clean" environments. The soil, by contrast, is a heterogeneous milieu that is a dynamic part of the biosphere continually changing in response to climatic variations over all scales of time and space. It is not at all surprising that the minerals in the soil can be quite different in their chemical and physical characteristics from those of the "type" minerals formed in more "geological" environments with which they may share a name and ideal crystalline structure (Churchman, 2010). Chadwick and Chorover (2001) made the point that the crisp boundaries in stability diagrams are, in the real world of natural soils, not so clear-cut. However, the effort involved in understanding the real nature of minerals in soils is worthwhile in view of the potential reward of being able to explain and predict properties across different sorts of soils from the nature of their constituents

TABLE 20.1   Definitions of Terms

Term	Definitions[a]	References	Issues with Definitions
Mineral	An element or compound that is normally crystalline and that has formed as a result of geological processes	Gaines et al. (1997)	Crystallinity implies a regularity of structure, but this is dependent on type of analysis. Some materials (e.g., volcanic glass) lack a fixed structure and are referred to as mineraloids; others such as allophane, earlier thought to be amorphous and technically thus not minerals, are now known to be nanocrystalline and hence do qualify as minerals
Primary minerals	Mineral deposited or formed at the same time as the rock containing it	Lapidus (1987)	Minerals in sedimentary and partially metamorphosed rocks may contain secondary minerals from earlier cycles
Secondary minerals	Mineral formed as a result of alteration of preexisting minerals	Adapted from Lapidus (1987)	Minerals commonly altered in several steps; hence all products of 1° mineral alteration are 2° minerals Often referred to as "clay minerals"
Weathering	Weathering is the breakdown of geological deposits by physical, chemical, and biochemical processes • Physical weathering includes breakdown by thermal and mechanical action, generating smaller particles (comminution); chemical weathering includes breakdown by hydrolysis, carbonation, hydration, dissolution, oxidation, and reduction, changing the physical and chemical properties of the original geological deposits	Ashman and Puri (2002) and Buol et al. (2003)	Some regard weathering as encompassing both the breakdown of primary minerals and the synthesis of new secondary minerals, but others distinguish between two phases, *decomposition* and *argillization*, the latter being the synthesis of secondary minerals (mainly clays)
Soil	The natural, 3D body (comprising pedons), typically up to ~1 to 2 m thick (thicknesses can vary markedly), of unconsolidated mineral and organic material mantling the land surface that can support rooted land plants, and which • Is characterized by one or more soil horizons ("layers") that have evolved through additions, losses, transfers, and transformations of energy and matter by interactions of climate, biota, relief, and parent materials through time, and which differ from the material(s) from which they derive • Comprises solids (inorganic and organic matter), liquids, and gases • Is essential to life through recycling of nutrients, carbon, and oxygen • Is nonrenewable in human timescales	Schaetzl and Anderson (2005)	Most definitions of soil fail to encompass and recognize its complexity as the most complex ecosystem on Earth and as a biological habitat and critically important repository for genes. Soil is also being considered increasingly in policy making, for example, as a provider of "ecosystem services" (e.g., Blum et al., 2006) and as "natural capital" (the stock of biotic and abiotic mass that contains energy and organization) (see Sparling et al., 2006, and Robinson et al., 2009)

[a] Working definition for this chapter.

and that of their associations. As discussed in Section 20.4, soil mineralogy has to some extent been in decline as a subdiscipline of soil science and it may be that this is due, at least in part, to the often-unjustified assumption that soil minerals are similar in their properties to those of their well-characterized "types" formed in nonsoil environments. The implications of this erroneous or only partly true assumption, and new directions for soil clay mineralogy, are also discussed in Section 20.4.

This chapter deals with minerals, both primary and secondary, with the processes of weathering and associated clay formation, and with soils. We define each of these terms in Table 20.1, which also includes some alternative definitions, and some issues involved in these definitions.

20.2 Alteration of Primary Minerals

20.2.1 Common Primary Minerals in Soils

The most common primary minerals found in soils are listed in Table 20.2 within their chemical (compositional) types and their mineralogical groupings together with the types of (crystal) structures for these groupings. Table 20.2 also lists chemical formulae for the minerals or groups of minerals and summarizes the general sorts of soils in which they occur and often also their relative abundance in these soils. Other properties of the minerals listed are given in Churchman (2006, Table 3.1). In soils, primary minerals, as residuals of physical and chemical weathering processes, most commonly occur in coarser particles

TABLE 20.2 Structure, Composition, and Soil Occurrences of Common Primary[a] Minerals in Soils

Chemical Type	Group	Structural Type	Common Soil Minerals	Chemical Formula	Related Phases	Soils of Main Occurrence
Silicates[b]	Silica	Tectosilicate	Quartz	SiO_2	Cristobalite Tridymite Opal (biogenic)	Quartz in almost all soils, Nil/minor from basalt
	Feldspar	Tectosilicate	Orthoclase	$KAlSi_3O_8$	Microcline Sanidine (all K-feldspars)	Feldspars occur in many soils, absent from highly weathered soils
			Albite	$NaAlSi_3O_8$	Oligoclase	
			Oligoclase	$Na_xCa_yAlSi_zO_8$ $(x \gg y)$, $(z = 2\text{–}3)$	Andesine	
			Anorthite	$CaAl_2Si_2O_8$		
			Labradorite	$Ca_yNa_xAlSi_zO_8$ $(y \gg x)$, $(z = 2\text{–}3)$	Bytownite (all plagioclases)	
	Zeolite	Tectosilicate	Clinoptilolite	$Na_3K_3(Al_6Si_{30}O_{72}) \cdot 24H_2O$	Erionite Faujasite Hollandite	Rare in soils; analcime formed in high pH saline soils
			Analcime	$Na_{16}Al_{16}Si_{32}O_{96} \cdot 16H_2O$	Laumontite Mordenite	
	Mica	Phyllosilicate (dioctahedral)	Muscovite	$KAl_2AlSi_3O_{10}(OH)_2$	Paragonite Maragarite Glauconite	Widespread
		Phyllosilicate (trioctahedral)	Biotite	$K(MgFe^{II})_3AlSi_3O_{10}(OH)_2$	Phlogopite Clintonite Lepidolite	Only in slightly weathered soils
	Chlorite	Phyllosilicate	Chlorite	$(Fe,Mg,Al)_6(Si,Al)_4O_{10}(OH)_8$	Cookeite Sudoite Bonbassite	Only in very slightly weathered ("raw") soils
	Serpentine	Phyllosilicate	Chrysotile	$Mg_3Si_2O_5(OH)_4$	Antigorite Lizardite Amesite Berthierine	Rare in soils
	Amphibole	Inosilicate (double chain)	Hornblende	$(Ca,Na)_{2.3}(Mg,Fe,Al)_5(Si,Al)_8O_{22}(OH)_2$	Tremolite Actinolite Cummingtonite Glaucophane	Quite widespread, but absent from highly weathered soils
	Pyroxene	Inosilicate (single chain)	Augite	$(Ca,Na)(Mg,Al,Fe)(Si,Al)_2O_6$	Enstatite Hypersthene Diopside Pigeonite Jadeite Spodumene Hedenbergite	Relatively rare except in some volcanic soils (especially intermediate-basic)
	(Cyclosilicate)	Cyclosilicate	Tourmaline	$(Na,Ca)(Li,Mg,Al)(Li,Fe,Mn)_6(BO_3)_3Si_6O_{18}(OH)_4$	Beryl	Rare
	(Sorosilicate)	Sorosilicate	Epidote	$Ca_2(Al,Fe)Al_2O(Si_2O_7)SiO_4(OH)_4$	Zoisite	Uncommon
	Olivine	Nesosilicate	Forsterite	Mg_2SiO_4	Fayalite Tephroite Monticellite	Rare
	Garnet	Nesosilicate	Almandine	$Fe_3Al_2(SiO_4)_3$		Rare
	Zircon	Nesosilicate	Zircon	$ZrSiO_4$	Baddeleyite (ZrO_2)	Widespread

TABLE 20.2 (continued)   Structure, Composition, and Soil Occurrences of Common Primary[a] Minerals in Soils

Chemical Type	Group	Structural Type	Common Soil Minerals	Chemical Formula	Related Phases	Soils of Main Occurrence
Nonsilicates	Phosphates	Insular, hexagonal	Apatite	$Ca_5(PO_4)_3(OH,F,Cl)$	Fluorapatite Hydroxyapatite Variscite Wavellite Monazite	Some in "raw" soils
	Titanium oxide	Tetragonal	Rutile	TiO_2	Brookite Sphene ($CaTiSi_5$)	Widespread in small amounts
	Carbonate	Rhombohedral	Calcite	$CaCO_3$	Aragonite Siderite ($FeCO_3$)	Common in arid soils
	Carbonate	Rhombohedral	Dolomite	$CaMg(CO_3)_2$	Ankerite Mg calcite	From dolomitic rocks
	Iron oxide	Cubic	Magnetite	Fe_3O_4	Titanomagnetite	Rare except in some volcanic soils and in soils derived from detrital volcanic deposits

[a] Some are also formed in soils, hence may also be regarded as secondary minerals.

[b] Volcanic glass (including obsidian) is an amorphous solid with a poorly ordered internal structure comprising loosely linked SiO_4 tetrahedra (Fisher and Schminke, 1984). Of widely varying composition, it is a common primary mineraloid in many volcanic soils and derivatives.

(of sand- and silt-size), although the phyllosilicates such as the micas, chlorite, vermiculite, and talc, and some oxides, such as those of titanium and also apatite, and volcanic glass, can also be found in reasonable proportions in the clay-size factions of some soils.

20.2.2 Observations of Relative Stabilities of Primary Minerals in Soils

In most cases, weathering involves the reaction of minerals (or mineraloids) with water, or, at least, an aqueous solution. There are exceptions where alteration of primary minerals occurs but where liquid water is either absent or is not the primary agent effecting alteration. This is the case in cold (and polar) and also hot desert zones on Earth (Shoji et al., 2006). It is probably also the case in some extraterrestrial environments, such as the Moon, but not in others, such as Mars, where there is abundant evidence for liquid water, at least in the past (Certini and Scalenghe, 2006). In environments where liquid water does not provide the main weathering milieu, the dominant agents of alteration nonetheless involve climatic factors; hence, they constitute weathering. These include mainly physical actions such as frost wedging and successive freezing and thawing in cold desert zones on Earth (Shoji et al., 2006), processes such as wind ablation in both hot and cold terrestrial desert zones, and such processes as solar wind irradiation on Moon (Certini and Scalenghe, 2006), and perhaps also on other dry extraterrestrial bodies and regions. Other agents besides the purely climatic can also alter primary minerals, given enough time. Among these are some resulting from various human activities (Certini and Scalenghe, 2006), including agriculture (Velde and Meunier, 2008), mining (Murakami et al., 1996; Banfield and Murakami, 1998), landfills

(Batchelder et al., 1998), industrial pollution (Rampazzo and Blum, 1992; Yaron et al., 2010), and probably also global climate change (Berg and Banwart, 2000; Amundson, 2001; Leifeld, 2006). Most of these agents affect mineral alteration via aqueous solutions, although in some cases, for example, the oxidation of sulfides, particularly iron pyrite, alteration of minerals occurs when water is absent (Dent and Pons, 1995). Nevertheless, alteration by weathering most often occurs in an aqueous environment and the most important determinant of the relative stability of primary minerals to weathering is their relative solubility in water.

All minerals have some solubility in water. Only rarely, however, for example, for simple salts like halite (NaCl) found, for example, in arid and saline soils, do they dissolve congruently in water. Instead, hydrolysis commonly occurs with the more soluble components being removed selectively from the mineral. This removal may leave a solid residue, which differs in composition from the original mineral, or a solid alteration product that precipitates out of the solution. The solid precipitate will most likely have a different composition and perhaps also a different crystal structure from that of its antecedent (partly or completely) dissolved mineral. Some texts refer to the synthesis or precipitation of such products as "argillization," meaning the formation of clay; in this case, the term "weathering" encompasses only the decompositional phase (Figure 20.1) (e.g., Buol et al., 2003). Another term widely used is neogenesis, effectively the formation or synthesis of new clays and other secondary material from the breakdown products of antecedent minerals.

A simple chemical analogy of the hydrolysis undergone by silicates, the most common type of primary mineral, in water, is realized when they are considered as salts of silicic acid H_4SiO_4 and bases of the appropriate cation (Carroll, 1970; Chamley, 1989).
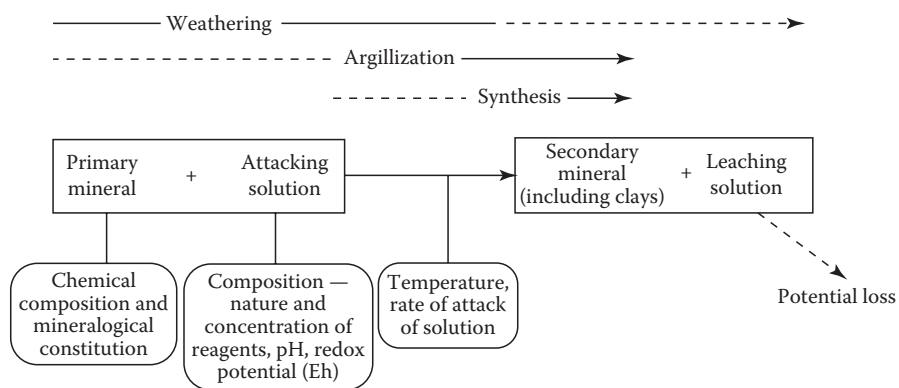
FIGURE 20.1 Weathering (and associated processes of argillization and synthesis) in relation to the factors governing the genesis of secondary minerals including clays in soils. (Modified after Percival, H.J. 1985. Soil solutions, minerals, and equilibria. New Zealand Soil Bureau Scientific Report 69. Department of Scientific and Industrial Research, Lower Hutt, New Zealand. With permission.)

Box 20.1   Hydrolysis of K-Feldspar, Interpreted as a Salt, to Form Kaolinite

Silicate minerals ≡ salt of silicic acid plus base of appropriate cation

$$\therefore \quad \text{K-feldspar} \equiv \text{silicic acid} + \text{hydroxyl-Al } (+K^+)$$

$$\therefore \quad 2KAlSi_3O_8 + 11H_2O \rightarrow Al_2Si_2O_5(OH)_4 + 4H_4SiO_4 + 2K^+ + 2OH^- \tag{20.1}$$
$$\text{K-feldspar} \text{kaolinite} \text{silicic acid}$$

or $\qquad\qquad 2KAlSi_3O_8 + 9H_2O + 2H^+ \rightarrow Al_2Si_2O_5(OH)_4 + 4H_4SiO_4 + 2K^+ \tag{20.1a}$
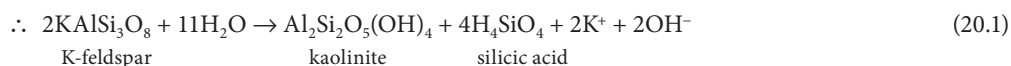
The hydrolysis of silicates in water results in an acid, namely, silicic acid, and a base, namely, the hydroxy-Al–Si, kaolinite (see Box 20.1). Because most aqueous solutions in which weathering takes place are acidic because water is charged with dissolved CO_2, giving rise to a continuous supply of carbonic acid H_2CO_3 and hence H^+ ions, hydrogen ions are important weathering agents. Therefore, Equation 20.1a represents the natural situation more realistically than Equation 20.1.

The relative stabilities of the principal primary minerals in soils to alteration have long been studied. More than 70 years ago, Goldich (1938) drew up a stability series for these minerals (Figure 20.2) that has generally stood the test of time and further experimentation, for example, by Franke and Teschner-Steinhardt (1994).

20.2.3 Thermodynamic and Structural Explanations of Relative Stabilities

Goldich's series (Figure 20.2) makes thermodynamic sense insofar as the minerals therein are in the identical (but reversed) order to those in Bowen's classic reaction series describing the order in which the minerals crystallized out of a magma on cooling (Bowen, 1922). Goldich (1938) reasoned that the higher the temperature at which a mineral crystallized from magma, the greater the extent to which it was out of equilibrium with the surface temperature of Earth and, therefore, the more susceptible it would be to breakdown by weathering at the Earth's surface.

The alteration of a particular mineral or mineraloid begins with the disruption of its weakest bond. Structural explanations of the relative stabilities of minerals have followed this generalization. Bonds in all silicates are based on silica tetrahedra. Their strength depends on the following: (1) the nature of the links between tetrahedra, (2) the extent of substitution of four-valency Si within tetrahedra by three-valency Al and hence gain of negative charge (isomorphous substitution), and (3) the extent of incorporation of charge-balancing cations, and their location, in the structure. In Table 20.3, we show the nature of bonding in the different structural types of silicates (see Table 20.2 to identify the relevant minerals).

Sposito (1989) pointed to an apparent correspondence between the ordering of the molar Si:O ratios in the silicate units, namely, olivines < pyroxenes < amphiboles < micas < feldspars = quartz
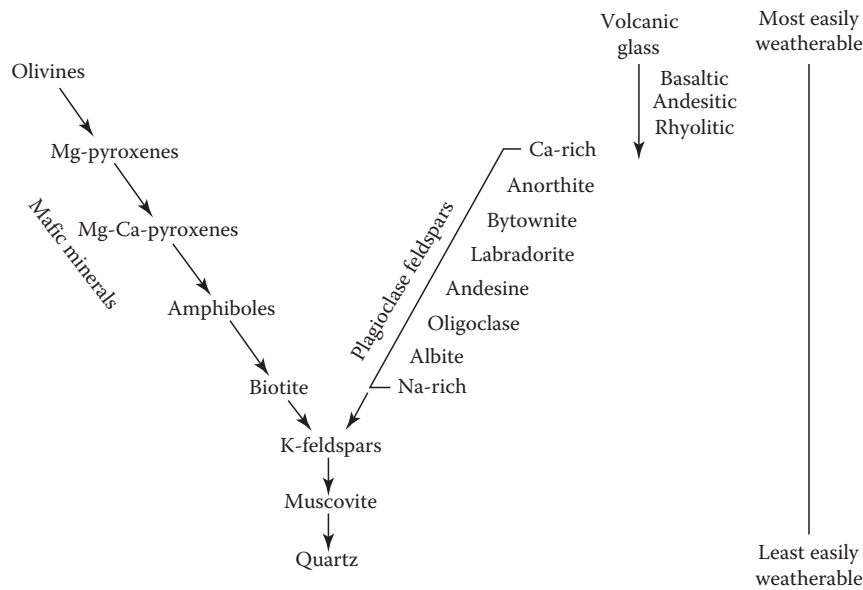
FIGURE 20.2 Stability series for the common primary minerals (after Goldich, S.S. 1938. A study in rock weathering. J. Geol. 46:17–58.) and also volcanic glass (not part of the Goldich series). Basaltic and other glasses, and olivines, are normally the first phases altered by weathering (Wolff-Boenisch, D. et al., 2004).

TABLE 20.3 The Nature of Bonding and the Weakest Bonds in Different Types of Silicates That Are Common in Soils (in Order of Decreasing Stability)[a]

| Structural Type | Group Name | Silicate Unit | | | Examples |
		Formula, Structure	Shared O per Si	Weakest Bonds	
Framework	Tectosilicates	SiO_2	4	Si–O	Quartz
	Tectosilicates	SiO_2, with Al substitution	4	With cations (K^+, Na^+, and Ca^{2+})	Feldspars
Layer	Phyllosilicates	$Si_2O_5^{2-}$, with Al substitution, as a sheet joined to Al-, Mg-, Fe-OH sheet	3	With interlayer cations, usually K^+	Micas
Single chains	Inosilicates	SiO_3^{2-}, with Al substitution	2.5	With divalent, and other cations	Pyroxenes
Double chains	Inosilicates	$Si_4O_{11}^{6-}$, with Al substitution	2	With divalent, and other cations	Amphiboles
Isolated tetrahedra	Nesosilicates	SiO_4^{2-}	0	With divalent cations	Olivines

[a] Cyclosilicates (e.g., tourmaline) have two O bonds per Si and sorosilicates (e.g., epidote) have one O bond per Si, but neither is common in soils (Table 20.2).

and that in the Goldich series. Nonetheless, reference to Table 20.3 may lead to oversimplification concerning possible structural control on the relative stability of silicates to weathering, especially when other silicates besides those in the broad groups represented by both types of ordering are considered. Following Loughnan (1969), for example, we note that both zircon ($ZrSiO_4$) and andalusite (Al_2SiO_5) are neosilicates like olivine and have $(Si:O)_{molar}$ ratios of 0.25 and 0.2, respectively, which are equal or less than those of olivine and yet are both very stable (i.e., resistant) to breakdown by weathering, in contrast to olivine, which is notably unstable according to Goldich's (1938) series (Figure 20.2). Furthermore, the broad classifications shown in Table 20.3 hide the observed differences in the stabilities between individual minerals within structural types. Within the phyllosilicate mica group, biotites are generally much less stable than muscovites, whereas the tectosilicate feldspars include albites and anorthites, each with quite different stabilities.

Another danger of oversimplification arises because weathering (and argillization) is not simply a matter of dissolution of the silicate minerals in water. Instead, weathering often occurs in an acid environment and, commonly in soils, also involves a range of organic compounds, which can complex the ions in the silicate structure and can both enhance their breakdown and affect its course and the products formed (e.g., Velde and Barré, 2010).

The rate at which hydrolysis affects cations released is related to the strength of the bonds formed by the cations within the minerals, according to Pauling's electrostatic valency principle (Pauling, 1929). The ratio of the valency (the charge on the ion) to the coordination number (the number of ions of opposite charge surrounding an ion) is known as the electrostatic valency. The smaller its value, the greater the ease of hydrolysis (Paton et al., 1995). The hydrolysis, or replacement of a cation in a crystal structure by H^+, of K is relatively easy because it is univalent (K^+).

It is commonly found in feldspars, where its coordination number is 9, and in micas, where it has a coordination number of 12. Hence, the electrostatic valency of K in these minerals is 1:9 and 1:12, respectively (Paton et al., 1995). The hydrolysis of Al is more difficult because it is trivalent (Al^{3+}), with coordination numbers of 6 or 4, so the electrostatic valency of Al in primary minerals is either 1:2 or 3:4. These differences between cations, and also between the same cations in different minerals, are reflected in the relative responses of the four main groups of silicates to weathering pressures (Paton et al., 1995). The solubility or relative mobility of various ions such as Ca, K, and Mg relates to ionic potential, which is the ionic charge (valency) divided by the ionic radius (size), so that ions with a low ionic potential such as Na^+, K^+, Ca^{2+}, and Mn^{2+} are more easily leached (lost) in solution, although temperature and pH affect solubility as well (Schaetzl and Anderson, 2005).

Some components of the silicate framework can greatly influence the rate and nature of their breakdown. Chief among these is Fe. According to Millot (1970), the most important weathering reactions involve Fe, a component of many of the primary minerals involved in soil formation (Table 20.2). Along with Mn, which is less abundant in the primary soil-forming minerals (Table 20.2), Fe generally occurs in its reduced form in primary minerals. Fe occurs as Fe(II) and Mn as Mn(II). These are both easily converted to their oxidized forms—Fe(III) and Mn(IV)—when oxidation occurs, for example, when soils dry. This change in valency sets up a charge imbalance in the appropriate minerals, leading to a loss of other ions from the mineral structures, which can destabilize the minerals, enhancing their further breakdown by hydrolysis, which is assisted by the hydrogen ions produced during oxidation. The oxidized forms of Fe and Mn occur as oxides and oxyhydroxides and those of Fe, in particular, are almost ubiquitous in soils. Seasonal wetting and drying cause alternate oxidation and reduction of Mn(II) and Mn(IV) and lead to the common precipitation of the blue-black mineral pyrolusite (MnO_2)—as well as other forms of MnO_2 such as todorokite (Churchman et al., 2010)—as redox segregations (also known as redox concentrations) in the form of coatings (mangans), nodules, or concretions (Vepraskas, 1992; Birkeland, 1999). Manganese oxides in most conditions form preferably to those of Fe during wetting and drying because of the greater ease of acceptance of electrons by Mn than by Fe. The reduction of Fe(III) to Fe(II), commonly by organisms during phases in which a soil is waterlogged, and then its oxidation on drying, leads to Fe oxides being precipitated as mottles, coatings, or concretions. The persistence of reducing conditions causes soil matrix materials to have pale, low-chroma colors (described as redox depletions), and these grayish colors characterize continuously reduced soils that, therefore, have few or no redox segregations (Birkeland, 1999; Vepraskas et al., 2004; Morgan and Stolt, 2006). Brinkman (1970) characterized the clay decomposition brought about by cyclical alternation of redox conditions by the name of "ferrolysis." Even so, a reexamination of some texture-contrast soils having coarser topsoils than subsoils by Van Ranst and De Coninck (2002) found that such coarsening with depth could be more easily attributed to the translocation of clay than to its destruction.

Weathering and argillization, together with soil formation, generally occur in a dynamic system that is open to the wider environment. Leaching—the removal or loss of material in solution—commonly occurs. The driving force that provides the chemical potential energy for mineral alteration is the difference in composition between the "attacking" solution and the solid mineral. However, it is the rate of leaching, which governs the rate of removal of solutes, which most influences the rate of mineral alteration and any subsequent precipitation (Figure 20.1). It also influences the course that the alteration takes. Garrels and Mackenzie (1971) stated that, given sufficient leaching, almost all rocks or unconsolidated deposits will leave a residue that is largely composed of the relatively insoluble oxides of Al, Fe, and Ti.

Hence, the origin of primary minerals as given by their temperature of crystallization from magma influences their thermodynamic stability against breakdown by weathering in an aqueous environment, as does their structure and their particular chemical composition (Figure 20.1). However, such environmental factors as the particular composition of the solutions bathing the minerals, the dynamics of the redox conditions, and the rate of throughflow of water past the minerals can play a decisive role in the kinetics, course, and ultimate products of their alteration.

20.2.4 Processes and Products of Alteration of the Main Types of Primary Minerals by Weathering

Early work exploring either the fate of primary minerals or the origin of secondary minerals on weathering (e.g., Jackson et al., 1948) tended to take the stance that specific primary minerals have led to particular secondary minerals as products. A 1:1 correspondence between altered rock-forming mineral and phyllosilicate product appears justified when one dominant type of primary mineral is altered more than others in the parent material and the overlying soil, saprolite, or regolith more generally is dominated by just one type of secondary mineral (clay). Some modern studies have also concluded that each type of primary mineral generally leads to one particular type of secondary mineral. Thus, according to a generalization by Wilson (2006), the weathering of granite rocks leads most commonly to kaolin minerals (kaolinite and halloysite), largely deriving from the alteration of feldspars, and vermiculite minerals from the alteration of micas. Furthermore, many clay minerals appear to be found in close proximity to particular types of altered primary minerals, for example, halloysite tubes on altered feldspar (Figure 20.3). Clay minerals may also have a morphology that appears to mimic that of altering or altered primary minerals. The similarity in shapes between vermicular books of kaolinite and degraded micas, often occurring as some form of vermiculite (Figure 20.4), can be suggestive of a micaceous origin for the particular kaolinite. To examine whether there is always a 1:1 relationship between particular
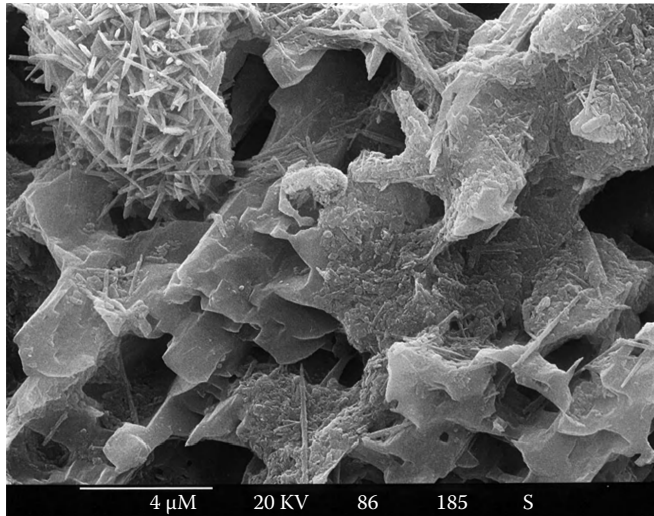
FIGURE 20.3 Scanning electron micrograph of tubular halloysite occurring in close proximity to weathered feldspar in granitic saprolite from Hong Kong. (Photo by Stuart McClure. Reproduced with the permission of the Geotechnical Engineering Office, Civil Engineering Office, Hong Kong.)
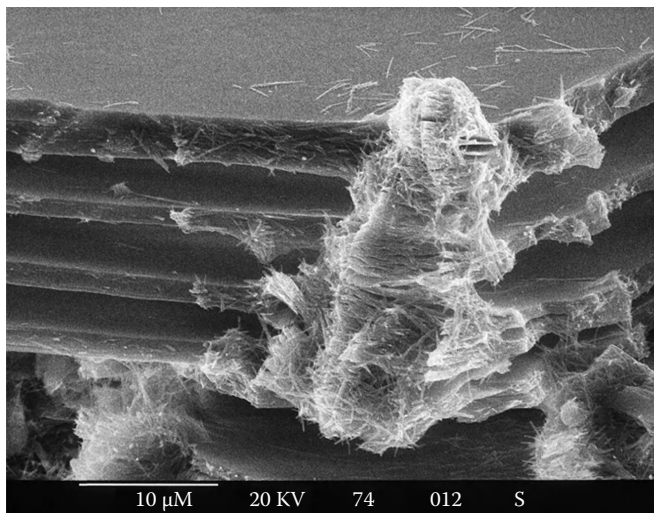
FIGURE 20.4 Scanning electron micrograph of vermicular kaolinite (with some halloysite tubes) occurring in close proximity to weathered mica in granitic saprolite in Hong Kong. (Photo by Stuart McClure. Reproduced with the permission of the Geotechnical Engineering Office, Civil Engineering Office, Hong Kong.)

primary minerals and specific secondary minerals, the following subsections summarize studies that have concentrated on the fate of individual types of primary minerals to alteration by weathering. In these subsections, the generalizations have often been drawn from studies of the weathering of rocks or other geological deposits, the minerals derived from these, and at depth within soil profiles. The influence of particular soil factors on the alterations and products of weathering is discussed in Section 20.2.5.

20.2.4.1 Processes and Products of Alteration of Olivines, Pyroxenes, and Amphiboles by Weathering

In an open environment, generally unstable olivines easily lose Mg^{2+}, and Fe^{2+} (or Mn^{2+} if manganous) to give, first serpentine, then a smectite (commonly saponite), and also goethite, which together have been characterized as "iddingsite" (Loughnan, 1969; Nahon et al., 1982; Eggleton, 1984). Both optical (Craig and Loughnan, 1964) and electron microscopy (Eggleton, 1984) have enabled identification of iddingsite as a mixture of saponite and goethite. Nahon et al. (1982) made the important observation that an aluminous dioctahedral smectite had formed from the breakdown of the Mg-olivine, forsterite (Mg_2SiO_4), which contains very little Al. They suggested that Al probably derived from the simultaneous breakdown of a pyroxene, enstatite. Secondary minerals could, therefore, derive their elemental constituents from the breakdown of more than one primary mineral. Indeed, olivines could give rise to a variety of secondary mineral products, including smectite, kaolinite, halloysite, and also various oxides, hydroxides, and oxyhydroxides of Fe, and, if present, also Mn, upon leaching (Huang, 1989).

Pyroxenes and amphiboles, although slightly more stable than olivines, tend to break down in a similar fashion to the olivines in the field, with an initial loss of their divalent cations, typically Mg, Ca, and Fe(II). Chlorite and/or smectite often result, although talc may also be formed. Calcite can also form as one of the products if Ca is released in abundance (Loughnan, 1969; Eggleton and Boland, 1982; Huang, 1989). Examination at the fine scale of the mineral grain revealed that amphibole weathering led to a dioctahedral (montmorillonite) and a trioctahedral (saponite) smectite, and also kaolinite–smectite simultaneously at the early stages of alteration, with halloysite developing with time (Proust et al., 2006). The different products were associated with different crystallographic faces of the host amphibole (see Section 20.2.5.1). As with olivines, pyroxenes and amphiboles could give rise to products with simpler structures, such as kaolinite, Fe(III) oxides, and anatase upon strong leaching (see also Wilson, 2004). The common pathway of alteration of this group of minerals by weathering is shown in Box 20.2.

Many laboratory studies have been designed to help understand the initial stages of alteration of these ferromagnesian minerals. At low pH, divalent ions are lost relative to silica, whereas the converse occurs at high pH with a preferential loss of Si (Schott and Berner, 1985). It has usually been inferred that leached layers, depleted in some of their original constituents, develop as a result on mineral surfaces (Schott and Berner, 1985; Banfield et al., 1995). Experimental studies (e.g., Velbel, 1985) have often indicated weathering rates that far exceed those observed in nature by up to 5 orders of magnitude (Banfield and Barker, 1994).

However, by following evidence from electron microscopy particularly, weathering of olivines, pyroxenes, and amphiboles has been seen to involve the formation, enlargement, and coalescence of etch pits developed on crystal dislocations

<div style="border:1px solid black;">

Box 20.2 Initial Alteration of Olivines, Pyroxenes, and Amphiboles as Commonly Reported

Olivines		Serpentine,		Smectite
Pyroxenes	$-Mg^{2+}, Fe^{2+} (Mn^{2+})$	chlorite,	$\longrightarrow$	(trioctahedral + goethite,
Amphiboles	$\longrightarrow$	and/or talc		e.g., saponite)

</div>
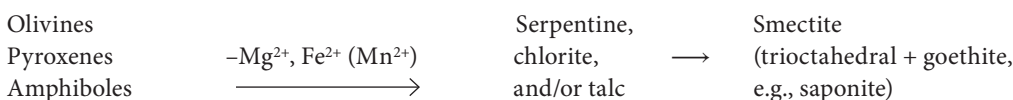
(Huang, 1989; Hodder et al., 1991; Seyama et al., 1996; Wilson, 2004) rather than through an enleached layer. Velbel and Barker (2008) have acknowledged that surface tension upon air-drying samples for conventional scanning electron microscopy (SEM) may affect surface structure and have examined partially weathered pyroxenes using high-pressure cryofixation (HPF) and cryo-field emission gun SEM, both of which avoid the surface tension effects of air-drying. They were thus able to view the products formed in small-scale etch pits on the pyroxenes and have concluded that pyroxene weathers to smectite by a multistage process. Velbel and Barker (2008) proposed that the process includes initial dissolution, topotactic reactions, mechanical disruption from wetting and drying cycles, further dissolution and, eventually, microbial colonization of the pores, with weathering finally enhanced by microbial extracellular polymers. As for the alteration of feldspars (see Section 20.2.4.2), the findings from laboratory studies may not reflect the processes occurring in the field because of oversimplification of the factors involved in natural weathering.

Nonetheless, Mogk and Locke (1988), using Auger electron spectroscopy (AES), found that naturally weathered hornblende showed systematic cation depletion to depths of ~120 nm. They considered that the dissolution of minerals generally occurs from both their surfaces and their bulk. Furthermore, Welch and Banfield (2002) were able to produce microstructural channeling effects in the olivine, fayalite, from laboratory dissolution experiments with an initial pH of 2.0, which were identical to those seen in the naturally weathered mineral (e.g., Banfield et al., 1991). X-ray photoelectron spectroscopy (XPS) had shown earlier that surfaces are coated by films of Fe(III) oxides both at low and high pH (Schott and Berner, 1985; Casey et al., 1993; Seyama et al., 1996), and Welch and Banfield (2002) found that when either microbes that enzymatically oxidize iron or instead ferric iron was added to the system, the minerals degraded at a greatly reduced rate. They identified the formation of laihunite, a ferric–ferrous olivine-like mineral containing vacant atomic sites, in the region of the mineral surface. This laihunite would protect the mineral against further dissolution.

Generally, alteration of primary minerals to give secondary products involves either transformation within the solid state or neoformation, with either partial or complete dissolution of solutes, then their reprecipitation as new secondary phases. Structural considerations appear to dictate that neither the neosilicate olivines, with independent silica tetrahedra, nor the inosilicate pyroxenes and amphiboles with tetrahedra in chains, can transform directly within the solid state to phyllosilicate clay minerals with flat sheets. Nevertheless, high-resolution transmission electron microscopy (HRTEM) observations of weathering amphiboles led Banfield and Barker (1994) to conclude that a type of transformation of amphiboles to smectites can occur within the solid state. They observed these reactions occurring in interstitial spaces within the primary amphiboles, with secondary smectites formed in a topotactic relationship to the altering amphibole. The change was described as a partial depolymerization of the amphibole structure leading to a repolymerization to that of smectite. There is no need for bulk water to bring about this change, with the water of hydration of the smectite product being able to supply the water required for the transformation reaction. Although Banfield and Barker (1994) pointed out that this type of transformation is not representative of all weathering reactions, it enables episodic alteration to take place within rocks or geological deposits, where it is limited by an intermittent supply of water and, therefore, much slower than that occurring on the surfaces of rocks or rock minerals immersed in aqueous solution. Banfield et al. (1995) found that smectites could form within weathering pyroxenoids (rhodonite and pyroxmangite, both $MnSiO_3$), which were studied by a similar type of in situ transformation. Periodic drying, which often occurs in this situation, may promote the repolymerization of phyllosilicates rather than depolymerized silica (Banfield et al., 1995). The reaction has little significance for soil formation, however, because the smectitic products are "quite ephemeral" and disappear when transported from the altered rock into the more open soil environment (Wilson, 2004).

Not all workers have considered that the primary ferromagnesian minerals are altered directly to phyllosilicates such as smectites. Nahon and Colin (1982) and Singh and Gilkes (1993) both identified amorphous or noncrystalline intermediates in the alteration of (different) pyroxenes. In comparing this observation with those of Banfield and coworkers, among others, who

observed direct formation of phyllosilicates on altering ferro-magnesian minerals, Wilson (2004) cautioned that the course of formation of phyllosilicates may have been established by prior hydrothermal or deuteric alteration in the latter cases and alteration by weathering had continued along the predisposed course.

20.2.4.2 Processes and Products of Alteration of Feldspars by Weathering

Feldspars have been thought to give rise to many different types of secondary minerals, encompassing the range of structural complexity from smectites through to gibbsite and quartz. Allen and Hajek's (1989) review referred to several studies where smectites have been identified as products of the alteration of feldspars. These authors surmised that the resultant smectites would be beidellitic because the parent feldspars lacked Mg or Fe, and Nettleton et al. (1970) characterized a smectite seen to form directly from a feldspar crystal as a beidellite–montmorillonite. Reports of the alteration of feldspars to gibbsite and quartz, at the lower end of the scale of structural complexity, were given by Allen and Hajek (1989) and Estoule-Choux et al. (1995). Otherwise, feldspars have been considered as the source minerals for micas (Carroll, 1970; Millot, 1970) and very often also for kaolinites and halloysites (Eswaran and Wong, 1978; Calvert et al., 1980; Anand et al., 1985). The micaceous minerals formed from feldspars, both K-feldspars and plagioclases, have often been described as "sericites," with the process of their formation being described as sericitization (Carroll, 1970; Millot, 1970). However, the term sericite has also been used to describe white mica formed from the reaction of biotite with plagioclase during metamorphism (Meunier and Velde, 1976). Furthermore, orthoclase (a K-feldspar) has been considered to hydrolyze to muscovite (Hemley, 1959). On the contrary, Meunier and Velde (1979) found that the micaceous phase, which they considered to be an illite, developed at a boundary between muscovite and orthoclase within weathering granite, and contained more Fe and Mg than either the muscovite or orthoclase. Hence, the illite formed (precipitated) out of solution. Using electron microscopy, Eggleton and Buseck (1980) detected micaceous phases forming within vacuoles within microcline. They considered these to be either illite or interstratified illite–smectite. Bétard et al. (2009) found that an illite formed within nonalkali feldspar (a plagioclase) by neoformation in Luvisols in northeast Brazil, and, more generally, they considered this particular weathering reaction to be typical of semiarid tropical and subtropical climates.

Studies of the artificial weathering of feldspars in controlled conditions in the laboratory have largely focused on the possibility of the formation of a leached layer on feldspar surfaces during their alteration. On the one hand, different types of surface-sensitive analyses—including secondary ion mass spectrometry, SIMS (also known as ion microprobe) (Muir and Nesbitt, 1997); elastic recoil detection, ERD, and Rutherford backscattering, RBS (Casey et al., 1988, 1989); and XPS (Muir et al., 1989, 1990; Hellmann et al., 1990)—indicated the formation of a dealkalized leached layer of up to 1 μm deep, especially at low pH. The leached product from an albite was identified as an amorphous phase with some of the characteristics of nanocrystalline allophane, which later became detached (Kawano and Tomita, 1994). The fate of the leached layer was found to be dependent upon the Si:Al ratio in the parent feldspar (Oelkers and Schott, 1995). Where this ratio was 1, as in plagioclases (anorthite was studied), the removal of Al led to completely detached silica tetrahedra, but where it was 3, as in alkali feldspars, Si tetrahedra that were still partially linked remained.

On the other hand, some measurements by XPS (Berner and Holdren, 1979) showed no leached layer forming, and when the artificial alteration was carried out at near-neutral pH values (between 5 and 8), rather than at low pH, the layer formed was only thin (Blum and Stillings, 1995). Huang (1989) concluded that, in any case, leached layers are not thick enough to inhibit transport and also that there were no continuous layers of secondary precipitates on weathered feldspars. As in the case of olivines, pyroxenes, and amphiboles, dissolution of feldspars probably occurred at sites of excess energy such as dislocations (Huang, 1989).

HRTEM studies have shown that unweathered feldspars are often turbid as a result of minute vesicles being filled with fluid (Hochella and Banfield, 1995), consistent with the observation that they can have a high microporosity (Worden et al., 1990). HRTEM has also shown that their alteration often occurs preferentially at crystal defects (Wilson and McHardy, 1980; Holdren and Speyer, 1987) and that secondary minerals form throughout the primary mineral, not just at grain boundaries (Banfield and Eggleton, 1989; Banfield et al., 1991). Application of SEM and associated energy-dispersive x-ray (EDX) analyses showed etch pits and secondary coatings (of a kaolin mineral) developing as the weathering alteration of feldspars progressed (Inskeep et al., 1993). These authors saw the formation of secondary coatings on feldspars as an integral part of their alteration. These were not observed in artificial weathering studies. Essential features of weathering in the field include those of seasonal wetting and drying, a generally high solid:solution ratio, and long residence times for water (Inskeep et al., 1993). The net effect is a weathering rate, determined for plagioclases, that is several orders of magnitude slower than the experimental dissolution rate (White et al., 1996, 2005; White and Brantley, 2003). Maher et al. (2009) have shown that other factors besides mineral dissolution rate and the rate of aqueous transport of solutes, notably the rate of secondary mineral precipitation, ensure a much slower rate of weathering of minerals, including feldspars, in the field than in the laboratory. Furthermore, fungi have been shown to contribute to the breakdown of feldspars on weathering, partly, at least, through their tunneling by organic anions exuded at the tips of the hyphae of (presumably ectomycorrhizal) fungae (Smits et al., 2005; see also Section 20.2.5.3 on podzolization). Failure to reproduce these features in necessarily hastened artificial weathering studies under closely controlled conditions almost certainly means that their results cannot be applied readily to the understanding of natural weathering.

20.2.4.3 Processes and Products of Alteration of Micas by Weathering

Changes from any of the neosilicate olivines, with independent silica tetrahedra, the inosilicate pyroxenes and amphiboles with tetrahedra in chains, and the tectosilicate feldspars with continuous frameworks of silica and alumina tetrahedral, to yield phyllosilicate clay minerals with flat layers, all involve some dissolution and recrystallization. These processes usually occur within the solution phase, when they can confidently be labeled as neogenesis, although, as we have seen from Banfield and coworkers, descriptions (see Section 20.2.4.1), neogenesis can also take place by a type of transformation within the solid phase that involves successive depolymerization and repolymerization. By contrast, phyllosilicate micas can transform easily to phyllosilicate clay minerals in the solid state.

The main changes that occur in these transformations are exchange of the interlayer cations and reduction in the charge of the layers. Potassium ions occupy the interlayer regions of the common micas, biotite and muscovite. The replacement of K^+ by hydrated cations such as Mg or Ca leads to a loss of strength of binding between adjacent layers. The interlayers expand, so that, with complete replacement by hydrated divalent cations, the basal spacing increases from 1.0 to 1.4 nm and vermiculite is formed. The concomitant reduction in layer charge has been supposed to occur by a wide variety of mechanisms. These include incorporation of protons into the layers (Raman and Jackson, 1966; Leonard and Weed, 1967), exchange of Si for Al in the tetrahedral sheet (Jackson, 1964; Sridhar and Jackson, 1974; Vicente et al., 1977), loss of hydroxyls (Stucki and Roth, 1977), and the deprotonation of hydroxyl groups and the loss of octahedral Fe, both occurring together (Farmer et al., 1971; Douglas, 1989; Fanning et al., 1989).

In soils particularly, micas rarely transform completely to vermiculite. In soil science, the term "illite" is most commonly used to describe clay-sized micaceous minerals. According to Grim et al.'s (1937) original definition of illite, it was "a general term for the mica-like clay minerals occurring in argillaceous sediments," which also "showed substantially no expanding lattice characteristics." It tends to be used more widely in practice. Illites commonly show a deficit of K and an excess of water in comparison with muscovite or biotite. This deficit reflects the greater ease of loss of K from micas that are fine grained rather than coarse, together with a paradoxical especially strong retention of some K in the fine-grained micas against replacement, both in nature and in the laboratory (Fanning et al., 1989).

Fine-grained micaceous minerals that include some expanding layers are sometimes also called illite (Grim, 1968; Wentworth, 1970; Norrish and Pickering, 1983; Weaver, 1989; Laird and Nater, 1993). Strictly speaking, however, these are interstratified illite (or mica)–vermiculites (or smectites). The interstratification may be either random or regular. If they are regular, with close to a 1:1 mix of the two constituent layer types, these two types alternate and the basal spacing is a sum of the spacings of the unaltered mica layers (1.0 nm) and the hydrous vermiculite layers (1.4 nm), giving a resultant spacing of 2.4 nm.

Alternating layer types in regular 1:1 interstratifications are explained by the replacement of the K ions in one interlayer by hydrated divalent ions, leading to the expansion of this interlayer, causing, in turn, a strengthening of the bond between K^+ and the aluminosilicate layers in the two adjacent interlayers (Bassett, 1959). These bonds probably strengthen because of a shift of structural hydroxyl groups toward the opened interlayer (Norrish, 1973). Regular interstratifications of micas with vermiculite (and also with smectites, as products of further transformation of micas) tend to occur in soils in colder climates, such as the upland regions of the former Yugoslavia (Gjems, 1970), Scotland (Wilson, 1970), Scandinavia (Kapoor, 1973), and the South Island of New Zealand (Churchman, 1980). In these cold climates, the displacement of interlayer K^+ occurs more slowly and in more discrete steps, that is, via alternate interlayer regions, than in warmer climates, where the transformation occurs more rapidly and in a relatively haphazard fashion to give irregular layer stacking. The release of K^+ from micas is considered to occur by either layer weathering, as proposed by Jackson et al. (1952), or edge weathering, as proposed by Mortland (1958), or both. In the former, most, if not all, of the K^+ in a particular interlayer is released virtually simultaneously. This release commonly occurs in the alteration of clay-sized micas. In edge weathering, K ions are released by diffusion from edges and fractures. This loss by diffusion commonly occurs in the alteration of larger mica flakes and has been observed as fraying in artificially altered micas, but it can also take place alongside layer weathering in clay-size micas (Fanning et al., 1989).

Trioctahedral micas, among them biotite and phlogopite, weather more readily than dioctahedral micas, including muscovite and most illites. Plants can bring about the transformation of biotite to vermiculite through their extraction of K from the mineral (Section 20.2.5.3). Among vermiculites in soils, the dioctahedral forms are more common than their trioctahedral counterparts (Jackson, 1959). Not only are the dioctahedral vermiculites more stable, but also they can form from biotites at the expense of the trioctahedral varieties as a result of a structural rearrangement whereby some octahedral cations are lost from the structure only to be replaced by some Al cations that are lost from tetrahedral sites (Douglas, 1989).

Transformation of micas in soils to expandable phases often occurs under acidic conditions. When pHs are relatively low (ca. 5 ± 0.5) and organic matter contents are low, Al is mobile in the aqueous phase as hydroxyl cations, and, where wetting and drying frequently occur, the conditions favor the deposition of Al-hydroxy species in the interlayers of vermiculites (and also smectites; Rich, 1968). The products, which are nonexpandable or only poorly expandable, are known by a variety of names including dioctahedral chlorite, pedogenic chlorite, 2:1–2:2 intergrade, chloritized vermiculite, and hydroxyl-interlayered vermiculite (HIV), and also their smectitic counterparts, such as hydroxyl-interlayered smectite (HIS) (Barnhisel and Bertsch, 1989).

Aluminous interlayers can protect vermiculites from further breakdown (Douglas, 1989) but can be destroyed by chemicals that dissolve Al hydroxyl species, including citrates. Citrates were once used routinely prior to x-ray diffraction (XRD) analyses (Mehra and Jackson, 1960) but their use can lead to the loss of potentially useful genetic information from the occurrence of the aluminous interlayers (Churchman and Bruce, 1988).

Vermiculitization of micas/illites has been found to occur by the expansion of the interlayer region into a wedge shape, known as the "frayed edge site" (Sawhney, 1972; Nakao et al., 2009). Through the use of a measurement known as "radiocesium interception potential" (RIP), which gives the amount of these frayed edge sites, Nakao et al. (2009) have compared minerals in soils that have formed from parent materials containing micas in various parts of Asia for their vermiculitic character in relation to the moisture regimes in which they have formed. RIP measures the selectivity of the frayed edge sites for Cs^+ in comparison with hydrated ions. A high RIP value, indicating a high selectivity for Cs, indicates a high concentration of vermiculite per se, whereas a low value can indicate either that little transformation of illite/mica has occurred or else that hydroxyl-Al interlayering has blocked the frayed edge sites that develop with vermiculitization. Nakao et al. (2009) discovered that, in subtropical Thailand, with an ustic moisture regime in which there is a distinct dry season, illites showed little alteration and, consequently, RIP was low. However, in a strongly leaching udic moisture regime and hyperthermic temperature regime in Indonesia, an advanced degree of vermiculitization was indicated by a high RIP. By contrast, an udic and mesic moisture regime in temperate Japan, while also producing a high degree of vermiculitization, nonetheless gave a low RIP value. This result arose because hydroxyl-Al interlayers blocked frayed edges. Nakao et al. (2009) noted that an index of weathering that is based on the degree of vermiculitization showed no relationship with one based on the oxidation status of Fe, which implied that the Thai soil was highly weathered despite its mica/illite showing little or no alteration. The factors that affect the development of Fe oxides in soils and those that affect loss of K^+ from the interlayers of micas and illites need bear no relation to one another. In particular, Nakao et al.'s (2009) results showed that the balance between the time spent by a soil in a dry season and the extent of leaching that occurs plays an important role in determining the extent of vermiculitization of its constituent micaceous phases.

Under low pH conditions, commonly where organic matter contents are high, as they often are under forests, micas may transform to smectites in soils as a result of acid leaching. These conditions typify podzolization and smectites often form through the transformation of micas in the eluviated horizons of podzols. The resulting smectites are generally Al-rich and hence beidellitic (Ross and Mortland, 1966; Churchman, 1980; Borchardt, 1989), but may also be montmorillonitic (Aragoneses and García-González, 1991). Strong weathering without podzolization can also transform micas to smectites (Stoch and Sikora, 1976; Egashira and Tsuda, 1983; Senkayi et al., 1983; Singh and Gilkes, 1991; van Wesemael et al., 1995), even in Antarctic soils, where smectites were among the transformation products of micas (Boyer, 1975). Smectites formed this way have generally been identified as beidellites. Their genesis contrasts strongly with those of most smectites, which form at high pH under poor drainage (see also Section 20.3.5).

Under cool climates, and, with sufficient rainfall, vermiculitic layers that are interstratified with mica layers can themselves be further transformed while the adjacent mica layers remain essentially unaltered. Thus, stepwise transformations of (muscovite) mica, first to regularly interstratified mica–vermiculite, then to regularly interstratified mica–beidellite, have occurred over a climosequence on increasing rainfall on mica–chlorite schist in South Island, New Zealand (Churchman, 1978, 1980). These changes to alternate layers have occurred under a current grassland vegetation, rather than under native forest. Indeed, soils within the area that have formed under forest, but with the same precipitation regime as one of the wetter sites in the climosequence under grassland, showed advanced transformation of the mica to a discrete beidellite phase. It appears that the often more acidic exudates from trees (e.g., Courchesne, 2006) impose a stronger weathering regime than that under grassland, so that the driving force exceeds that leading to transformation of alternate layers alone.

An advantage of many cool-climate studies is that not all primary minerals alter at the same rate, and the origin of secondary minerals can, therefore, be traced to the few, if not single, primary minerals that have altered. In the bulk of Churchman's (1978, 1980) studies, feldspars were fresh and only mica and chlorite showed signs of alteration. In warmer climates and, with sufficient throughflow of water and passage of time, micas break down further to form 1:1 minerals, both kaolinite and halloysite. Although the direct formation of kaolin minerals from micas alone cannot be concluded when other primary minerals have also altered, many workers have established that micas have altered to form kaolins mainly because of the appearance of the secondary minerals as pseudomorphs after the micas. This apparent pseudomorphic replacement applies for kaolinites formed in tropical Nigeria (Ojanuga, 1973), from tertiary weathering in Europe (Stoch and Sikora, 1976), in ferralitic (lateritic) soils in Western Australia (Gilkes and Suddhiprakarn, 1979), and in deeply weathered soils in the continental United States (Rebertus et al., 1986), and also for both halloysite and kaolinite formed alongside each other in close association with biotite in tropical Malaysia (Eswaran and Yeow, 1976).

Mainly with the help of electron microscopy, some workers have even found that kaolin minerals can form from micas early in the weathering process including from both biotite (Ahn and Peacor, 1987; Banfield and Eggleton, 1988) and muscovite (Robertson and Eggleton, 1991). Ahn and Peacor (1987) showed that kaolinite can also form as (irregularly) intercalated layers between biotite layers, presumably by a dissolution/crystallization process whereby one biotite layer gives rise to two kaolinite layers. This change takes place within "plasmic microsystems" (Velde and Meunier, 2008) involving small differences in chemical potentials, which arise because of connections with a major

passageway for water outside a rock or geological deposit. Such an occurrence—the kaolinization of biotite—may have occurred in weathered tephra beds aged ca. 350,000 years in northern New Zealand and which contain biotite in the unweathered parent tephra (Rangitawa tephra). The kaolinized mineral occurs as a sand-sized golden platy mineral. It comprises (using SEM) nearly nine primary (6–8 μm thick), four secondary (1–1.5 μm), and five to seven tertiary (0.7 μm) lamellae units, and was characterized via XRD as a K-depleted, partially random interstratified micaceous kaolinite intergrade containing <50% of 1 M trioctahedral mica and >50% of partially disordered dioctahedral kaolinite (Shepherd, 1984; Lowe and Percival, 1993).

Conversely, some (Eswaran and Yeow, 1976; Gilkes and Suddhiprakarn, 1979) have found that vermiculite—often considered to be an unstable intermediate phase (Kittrick, 1973)—nonetheless occurred alongside kaolin minerals in soils formed under strong leaching. One of the likely reasons for the appearance of apparently early weathering products such as vermiculite alongside apparently late stage products such as kaolin minerals is that iron (oxyhydr)oxides are very common products of the weathering of micas. The secondary iron phases coat other minerals such as vermiculites and preserve them against further breakdown. They include goethite (Sousa and Eswaran, 1975; Eswaran and Yeow, 1976; Gilkes and Suddhiprakarn, 1979; Banfield and Eggleton, 1988) as well as hematite (Gilkes and Suddhiprakarn, 1979). Several studies of the products of natural weathering of micas (Rice and Williams, 1969; Aldridge and Churchman, 1991; Aoudjit et al., 1996; Seyama et al., 1996), using variously Mössbauer and XPS spectroscopy, as well as the laboratory study of Farmer et al. (1971), have shown that total Fe and Fe(II) were depleted from micas on weathering while Fe(III), Fe gels, and Fe(oxyhydr)oxides built up outside the micas. Oxides and hydroxides of other metals, including Al, as gibbsite (Gilkes and Suddhiprakarn, 1979), and also titanium dioxide (Milnes and Fitzpatrick, 1989), may also have formed, at least partially, from micas.

20.2.4.4 Processes and Products of Alteration of Chlorites by Weathering

Chlorites occur most commonly as trioctahedral minerals in parent materials for soils. They originate mainly from low-grade metamorphic rocks and as products of the early alteration of Fe- and Mg-containing primary minerals such as augite, hornblende, biotite, and serpentines (see Section 20.2.4.5). They are not common in soils mainly because of their low stability with regard to weathering. The initial stages of their alteration are similar to those of micas and involve the loss of their interlayer species, most often hydroxides of Mg (i.e., brucite structures), but also those of Fe and other cations. Products of the earliest stages of weathering of chlorites have been identified variously as a randomly interstratified hydrous phase (Churchman, 1980), swelling chlorite, which is a type of chlorite depleted of some of its interlayer hydroxides (Stephen and MacEwan, 1951; Bain and Russell, 1981), and sometimes also regular interstratifications of chlorite and swelling chlorite (Churchman, 1980), as well as a

chlorite/vermiculite intergrade (Murakami et al., 1996). There is also a strong tendency to form 1:1 regular, or "semiregular," chlorite–vermiculite interstratifications at the next early stage of weathering. This mineral type has been observed in weathering by Johnson (1964), Herbillon and Makumbi (1975), Churchman (1980), Banfield and Murakami (1998), and Aspandiar and Eggleton (2002a, 2002b) among others. A study with atomic-resolution transmission electron microscopy (TEM) led Banfield and Murakami (1998) to propose that there was a tendency toward regular 1:1 interstratification because interlayer Mg and Fe (oxyhydr)oxides are removed from every second interlayer on account of a layer shift (of ~a/3) occurring after the removal of one interlayer (oxyhydr)oxide that thereby stabilizes the adjacent interlayer (oxyhydr)oxide. The stabilization means that the (oxyhydr)oxide in the next interlayer is more labile to replacement. However, Wilson (2004) warned that observations of regular chlorite–vermiculite interstratifications should be treated with caution as products of weathering alone because some, at least, of the parent rocks may have undergone prior hydrothermal alteration. Interstratified chlorites and also vermiculite-like phases have been found to occur in some metamorphosed rocks. Nonetheless, not all sequences of alteration of chlorites are the same, with several authors (e.g., McKeague and Brydon, 1970; Bain, 1977; Churchman, 1980; Ross et al., 1982; Righi et al., 1993; Carnicelli et al., 1997) finding that the chlorite component, often of chlorite–mica schists, disappeared as a result of acid weathering in a podzolized soil, or spodosol. Chlorite dissolved, typically leaving a solid residue of iron oxides and oxyhydroxides, for example, goethite (Bain, 1977; Ross et al., 1982). Frequently, however, there is further development of the interstratifications of chlorite with vermiculite to discrete vermiculite (Loveland and Bullock, 1975; Murakami et al., 1996). Often also, smectites are considered to form from chlorite (Herbillon and Makumbi, 1975; Carnicelli et al., 1997). Although the primary chlorites are trioctahedral, the resulting vermiculites and smectites are mostly dioctahedral (Wilson, 2004). Ultimately, kaolin minerals, both kaolinite (Herbillon and Makumbi, 1975; Murakami et al., 1996) and also halloysite (Cho and Mermut, 1992), can result from the strong weathering of chlorites.

20.2.4.5 Processes and Products of Alteration of Serpentines by Weathering

Serpentinite rocks, which are dominated by serpentines (mainly chrysotile, antigorite, and minor lizardite), together with iron oxides and such minerals as amphiboles, pyroxenes, and talc (e.g., Bonifacio et al., 1997), are generally unstable in soils. In central California, this rock type gave rise to virtually pure Fe-rich smectite in the fine clay (<0.2 μm) fraction (Wildman et al., 1968). These authors attributed the mineralogical change to the loss of the more mobile elements Mg and Si and relative enrichment of Fe and Al. In the Massif Central in France, alteration of serpentinite in a poorly drained B horizon led to a trioctahedral chlorite, which is normally regarded as a primary mineral, and also to its alteration products: a regularly interstratified chlorite–vermiculite and a nontronitic smectite

(Ducloux et al., 1976). This change, while involving some loss of Mg and Si—as expected in an open soil system—is more typical of a closed system (Ducloux et al., 1976). The weathering products of a serpentinite in northwest Italy were found to depend strongly on drainage conditions (Bonifacio et al., 1997). Low-charge vermiculite sometimes formed but either dissolved or was transformed to smectite if drainage was poor. An aluminous chlorite could result from either the preformed vermiculite or smectite as a result of interlayering by hydroxy-Al. Lee et al. (2003) in California and Caillaud et al. (2004) in France also found that vermiculite and smectite were formed in the course of weathering of serpentinites. In general, the smectites formed from serpentinites per se are complex and heterogeneous according to Caillaud et al. (2004). Caillaud et al. (2004) focused on a microsystem in the serpentinite weathering system. The types of vermiculite and/or smectite formed varied, with both trioctahedral and dioctahedral structural types covering a range of layer charges, depending upon their originating primary minerals. For instance, a thin lizardite bastite gave rise to Al-poor trioctahedral saponite, whereas a dioctahedral Fe-rich montmorillonite appeared to derive directly from chrysotile. Using a toposequence in Taiwan on serpentinite containing chrysotile, antigorite, and lizardite, and also chlorite and talc, Hseu et al. (2007) were able to identify the products of weathering as, first, smectite that was dominantly trioctahedral, then interstratified chlorite–vermiculite, and, finally, kaolinite and quartz.

20.2.4.6 Processes and Products of Alteration of Volcanic Parent Materials by Weathering

One outstanding feature of volcanic materials as parent materials for clay-size minerals is that they usually contain glass, which is a fast-weathering source of Si and Al (and other elements) for mineral neogenesis. They can also contain other weatherable minerals, depending on their origin, nature, and composition. Broadly, volcanic materials may originate from effusive eruptions, forming lavas, or explosive eruptions that generate fragmental, unconsolidated deposits called pyroclastic materials, or tephra, which may be distributed widely by the wind. Among the latter, there are three main types, based on composition. These are (1) basaltic, which are rich in Fe and Mg, reflecting usually high contents of ferromagnesian, or mafic, minerals, namely, olivines, amphiboles, and pyroxenes, as well as feldspars and brownish-colored basaltic glass low in Si and relatively high in Al; (2) rhyolitic, which have abundant Si-rich glass comparatively low in Al, and which may contain minor amounts of mafic minerals including biotite, and also feldspars; and (3) andesitic or dacitic, which are intermediate in composition between basaltic and rhyolitic eruptives (see Section 33.3, for more details on their composition, occurrences, and properties).

20.2.4.6.1 Glass

The alteration of glass is a special case. Glass is an amorphous solid with a poorly ordered internal structure comprising loosely linked SiO_4 tetrahedra with considerable intermolecular space

(in which cations such as sodium occur) (Fisher and Schminke, 1984). The hydration and breakdown of glass results in fluxes of some elements from the glass into interstitial pore waters, and the very rapid precipitation of secondary minerals from such solutions as well as replacement of glass shards by new minerals (e.g., Daux et al., 1994). Dissolution of basaltic glass, and probably also other glass types, as determined by Oelkers and Gislason (2001) and Gislason and Oelkers (2003), proceeds in effectively three steps: (1) relatively rapid and largely complete removal of univalent and divalent cations from the near-surface glass structure via the breaking of metal–oxygen bonds and their replacement with proton–oxygen bonds, (2) Al-releasing exchange reactions between three aqueous H^+ ions and Al in the glass structure, and (3) relatively slow detachment of partially liberated silica. The tetrahedrally coordinated Si–O bonds are the most stable bonds in the glass framework. Si tetrahedra on the glass surface may be connected to the glass framework via one to three bridging oxygens, and the rate of release of any Si atom at the surface decreases markedly as the number of bridging oxygen bonds increases (Gislason and Oelkers, 2003). The breaking of Al–O bonds does not destroy the glass framework but instead only partially liberates the silica tetrahedral chains by removing adjoining Al atoms (Gislason and Oelkers, 2003). Glass dissolution rates demonstrably increase with decreasing Si content (Wolff-Boenisch et al., 2004), and, hence, basaltic glass (low in Si) usually dissolves faster than andesitic glass (intermediate Si content) or rhyolitic glass (high in Si) under similar environmental conditions (Neall, 1977; Hodder et al., 1990; Shoji et al., 1993b; De Vleeschouwer et al., 2008; Sigfusson et al., 2008; Figure 20.2). In all cases, fragmental and vesicular glass components, such as those which occur in tephra deposits, have high surface areas and are very porous and so break down very quickly, and at rates closely proportional to geometric surface areas (e.g., Dahlgren et al., 2004; Wolff-Boenisch et al., 2004). The initial alteration of basaltic glass, described widely as "palagonitization," is described as a special case in Section 20.2.4.6.3.

20.2.4.6.2 Formation of Allophane and Halloysite from Volcanic Parent Materials

Allophane and halloysite are the most common secondary minerals that are formed in soils developed from loose volcanic material or tephra, which may be ash, pumice, and cinders (the last also referred to as scoria).

Using ^{27}Al and ^{29}Si NMR, Hiradate and Wada (2005) have deduced a mechanism for the formation of allophane from glass as another step in the weathering-synthesis process (Figure 20.5). This mechanism involves (1) dissolution of Al mainly from volcanic glass (via Al-releasing exchange reactions with protons) and its concomitant transformation from its tetrahedral to octahedral state, from ^{IV}Al to ^{VI}Al; (2) hydrolysis of the Al released into solution to give a gibbsite-like sheet; (3) dissolution (via breaking of Si–O bonds) of Si from volcanic glass, where it occurs as a silica gel-like polymer, to give monosilicic acid in solution; and (4) reaction between the gibbsite-like sheet and monosilicic acid to generate allophane.

Volcanic glass

^{IV}Al
$Q^40Al, Q^30Al,$
Q^41Al, Q^31Al

Dissolution of Al →

Formation of gibbsite-like sheet

^{VI}Al

Dissolution of Si
(breaking of Si–O–Si bonds) ↓

↓

$Si(OH)_4$

Monosilicic acid

→

^{VI}Al
Q^3Al

Formation of allophane
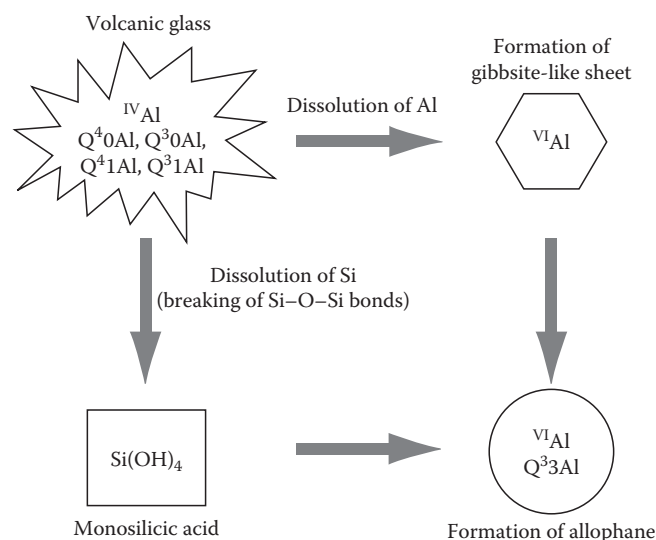
FIGURE 20.5 Model depicting the formation of allophane. Q notation describes the connectivity of a Q unit—that is, an Si atom surrounded by four oxygen atoms—to other Si atoms (Wright and Sommerdijk, 2001): Q4 indicates that the Q unit is fully interconnected with silicate groups, and Q3 indicates that the Q unit is linked in three places (referred to as branching points). (Redrawn from Hiradate, S., and S.-I. Wada. 2005. Weathering processes of volcanic glass to allophane determined by ^{27}Al and ^{29}Si solid-state NMR. Clays Clay Miner. 53:401–408, Fig. 4. With permission of The Clay Minerals Society.)

This process is consistent with modern understanding about the crystal structure of allophane. Its structure is regarded as being based on that of imogolite, another aluminosilicate that is sometimes found (typically in small amounts) in soils derived from tephra and in some other soils including podzols. Imogolite has an ideal chemical formula of $SiO_2 \cdot Al_2O_3 \cdot 2H_2O$ and, therefore, an Al:Si ratio of 2:1. The imogolite structure, established by Cradwick et al. (1972), consists of a defect gibbsitic Al-octahedral sheet framework. Si in each silica tetrahedron is attached to this framework by sharing three of its O atoms (hence the Q3 notation shown in Figure 20.5) with an octahedrally coordinated Al on the gibbsitic sheet while the remaining O atom on the tetrahedrally coordinated Si acquires an H atom to form a Si–OH bond pointing away from the gibbsitic sheet. Although imogolite, which comprises long, thin hollow nanotubes, always in bundles, has a definite XRD pattern (Yoshinaga and Aomine, 1962), and, hence, some long-range order, allophane gives extremely broad, low-intensity humps in XRD and is regarded as a mineral with short-range order (SRO), or a poorly ordered aluminosilicate, as well as a "structured nanomineral" (Hochella, 2008; Theng and Yuan, 2008; McDaniel et al., Section 33.3). The best term for it is "nanocrystalline," meaning structured at the nanometer scale, that is, 1–100 nm. (Another nanocrystalline mineral that occurs commonly but usually in relatively small amounts in weathered tephras and other materials, and which is analogous to allophane, is ferrihydrite, described later herein.) Allophane composition varies over a much greater range than that of imogolite, although morphologically all unit particles

comprise tiny hollow spherules or "nanoballs" (Parfitt, 2009). In soils, there are two main types: (1) Al-rich, with Al:Si ~ 2:1—these are imogolite-like or proto-imogolite allophanes—and (2) Si-rich, with Al:Si ~ 1:1—these are halloysite-like allophanes. The imogolite-like nanocrystalline structure applies only to Al-rich allophane, which is the most common type (Parfitt, 2009). Generally, it has an XRD pattern that is often indefinite, even if distinct peaks appear for some samples (Parfitt, 1990), but the pattern is likely to be indistinct when allophane is associated with long-range-ordered crystalline minerals. Therefore, a number of other sophisticated instrumental methods have been proposed for its identification. Earlier, these included infrared spectroscopy, differential thermal analysis, and TEM (Fieldes, 1955). More recently, electron diffraction (Wada and Yoshinaga, 1969), small-angle neutron scattering (Hall et al., 1985), ^{27}Al and ^{29}Si NMR (Goodman et al., 1985; Hiradate and Wada, 2005; Hiradate et al., 2006), and XPS (He et al., 2008) have also been used. Yet, it is a chemical method, for the extraction of Al and Si, among other metals, using acidified ammonium oxalate and developed long ago by Tamm (1922), which has become the standard (and critically important) procedure for both identifying and quantifying allophane (Parfitt and Henmi, 1980; Parfitt and Wilson, 1985; Wada, 1987, 1989).

Halloysite often also forms by the weathering of tephra. Although halloysite, as a 1:1 Si:Al mineral with the same aluminosilicate composition as kaolinite, has often been identified by a peculiar shape in electron micrographs (see Section 20.3.2), it is only the occurrence, or else evidence for prior occurrence, of interlayer water that distinguishes halloysites unequivocally from kaolinites (Churchman and Carr, 1975; Churchman, 2000). With the knowledge that allophane is a fast-forming SRO (nanocrystalline) product from the dissolution mainly of glass whereas halloysites, being crystalline with long-range order, generally give distinct peaks in XRD, there has been much debate over the question of whether allophane alters to halloysite and, if so, how this change occurs. There have been two main schools of thought regarding this question. One of these, proposed by Fieldes (1955) and which was the predominant idea until around the 1980s, held that allophane altered to halloysite with the passage of time. This mineralogical change was considered to occur by a solid-phase transformation involving dehydration and "crystallization" from an "amorphous" material.

According to the alternative view, which was formulated in Parfitt et al. (1983) and subsequent papers (e.g., Parfitt et al., 1984; Parfitt, 1990, 2009), it is the concentration of Si in soil solutions [Si], and availability of Al, which largely determine the nature of the aluminosilicate secondary minerals that form from volcanic parent materials by weathering. In particular, allophane is favored by a lower [Si], whereas halloysite tends to result when [Si] is relatively high. This theory was based on mineralogical and soil solution data from a rainfall sequence of soils formed on tephra-derived deposits in northern New Zealand. It was supported by Singleton et al. (1989) who measured in detail modern solution [Si] in soils forming a drainage sequence on tephric materials of similar composition and age.

Parfitt's theory helped to explain some anomalies concerning the application of Fieldes' (1955) hitherto pervasive theory that allophane seemed to inevitably alter or "transform" to halloysite with time. Fieldes' theory was based on the worldwide recognition that halloysite tended to occur at depth in many tephra-derived soil sequences whereas allophane predominated in surface horizons. Because such sequences become stratigraphically older with increasing depth, the assumption was made that allophane formed first and then, after ca. 10,000–15,000 years, it apparently transformed to halloysite (see review by Lowe, 1986). However, in northern New Zealand, McIntosh (1979) showed that authigenic halloysite (both spheroidal and tubular) had formed by recent processes in tephra deposits only ~1800 years old. He demonstrated that such halloysite formation was a consequence of resilication from a Si-rich soil solution—the resilication was indicated by Si:Al ratios of ~2 at depths of ≥ 2 m, in contrast to values of ≤1 in surface horizons, and by modern lysimeter leachate compositions (McIntosh, 1980). Other reports had previously demonstrated the seemingly "anomalous" occurrence of halloysite in young soils (Hay, 1960; Bates, 1962; Bleeker and Parfitt, 1974), and electron micrographs showed a close association of halloysite with the surfaces of parent minerals/mineraloids, both glass (Dixon and McKee, 1974) and feldspars (Tazaki, 1979; see Lowe, 1986, and also Figure 20.3 herein). Similarly, Ogura et al. (2008) showed in recent, proximal scoriaceous tephras (deposited since AD 800) near Mt. Fuji, Japan, that the coarse particle sizes and large pores facilitated rapid drying, aiding the concentration of Si and hence neoformation of tiny spherical halloysite particles. Thus, it became clear that halloysite, rather than allophane, was able to form directly from tephra materials in young soils under certain conditions where [Si] was relatively high. In reality, [Si] may vary seasonally, and kinetic considerations can lead to the formation of halloysite if the periods of high [Si] greatly exceed those when [Si] is low. The use of oxygen isotopes to trace the temperature of formation of minerals, among other techniques, led Ziegler et al. (2003) to conclude that halloysite has formed continuously from the early stages of formation from basalt in an arid zone in Hawaii as the result of prolonged extremely dry seasons following short periods of intense rainfall. The release of Al in the latter had been followed by a prolonged buildup of Si and hence halloysite formation. In addition, it had long been known that allophane was present in some very old weathered-tephra sequences (Ward, 1967; Tonkin, 1970), which had always been a puzzle in view of the Fieldes' model (although the amounts of allophane were not well quantified at that time). A key study was that by Stevens and Vucetich (1985) who used ammonium oxalate dissolution methods to quantify allophane content in a 10 m high tephra weathering sequence in northern New Zealand dating back ca. 350,000 years (the basal tephra being reidentified more recently as the Rangitawa tephra: Lowe et al., 2001). They demonstrated unequivocally that abundant allophane was present in many beds in the sequence. Moreover, through tephrochronology, Stevens and Vucetich (1985) were able to show that allophane (as well as subordinate gibbsite) was predominant during warm,

wet interglacial periods but that halloysite predominated during cool, dry glacial periods, implying that changing environmental conditions had led to changing [Si] and hence different clays. A similar finding was recorded by Bakker et al. (1996). Thus, the weathering of tephra followed separate pathways leading to the formation either of allophane or halloysite, rather than following a single pathway governed by time (Lowe, 1986).

The idea that a high [Si] favored halloysite also made sense of reports of the formation of halloysite instead of allophane when there was a thick overburden (Mejia et al., 1968; Aomine and Mizota, 1973; Wada, 1987; Cronin et al., 1996), and also where drainage was impeded or poor (Aomine and Wada, 1962; Dudas and Harward, 1975; Stevens and Vucetich, 1985; Cronin et al., 1996). Both these situations would give rise to a buildup of Si that, it was then thought, would react with preformed allophane to form halloysite (Aomine and Wada, 1962; Dudas and Harward, 1975; Saigusa et al., 1978).

Zehetner et al. (2003), in Andean Ecuador, Rasmussen et al. (2007), in the Sierra Nevada in California, and Chadwick et al. (2003), on Kohala Mountain in Hawaii, studied the mineralogy of soils formed at different altitudes but from a common volcanic source (andesitic/dacitic ash, andesitic lahar, and basaltic lava, respectively). In each case, it was found that allophane (with imogolite and ferrihydrite in Hawaii) was dominant in soils at higher altitudes and halloysite in soils at lower altitudes, confirming the trends discovered by Nizeyimana et al. (1997) in a similar study on volcanic materials in Rwanda. At higher altitudes, the precipitation is greater; hence, leaching is stronger, solution [Si] is lower, and allophane prevails in the soils formed there. Conversely, precipitation and degree of leaching are lower, and [Si] in solution is higher, at lower altitudes and so halloysite dominates in these soils. Chadwick et al. (2003) emphasized in their Hawaiian study, however, that as moisture increased along the sampling transect, different sets of secondary minerals were favored in response to conditions controlled by arid conditions, by rapid and intense cycles of wetting and drying, or by essentially continuously wet conditions. They developed a leaching index as the ratio of water balance to the integrated porosity of the top meter of soil on an annual basis. The index reached 1 (total filling of the pore space each year) where the mean annual precipitation (MAP) was ~1400 mm. Index values >1 indicated intense leaching conditions because of pore water replacement; leaching losses of soluble base cations and Si were nearly complete at such index values, whereas only 60% of Al had been lost. Where index values were <1, leaching losses were progressively lower with the lowest rainfall sites having lost 10%–20% of the original base cations and Si, but none of the Al.

Rasmussen et al. (2007) found further that soils in intermediate zones of altitude contained both allophane and halloysite and also that soils at extremely high altitudes had neither allophane nor halloysite but instead were dominated by interlayered 2:1 Si:Al layer silicates, which they considered to be inherited from the parent material. These authors considered that climate controlled the clay minerals formed from the volcanic materials. Rainfall explained the secondary phases formed except at the
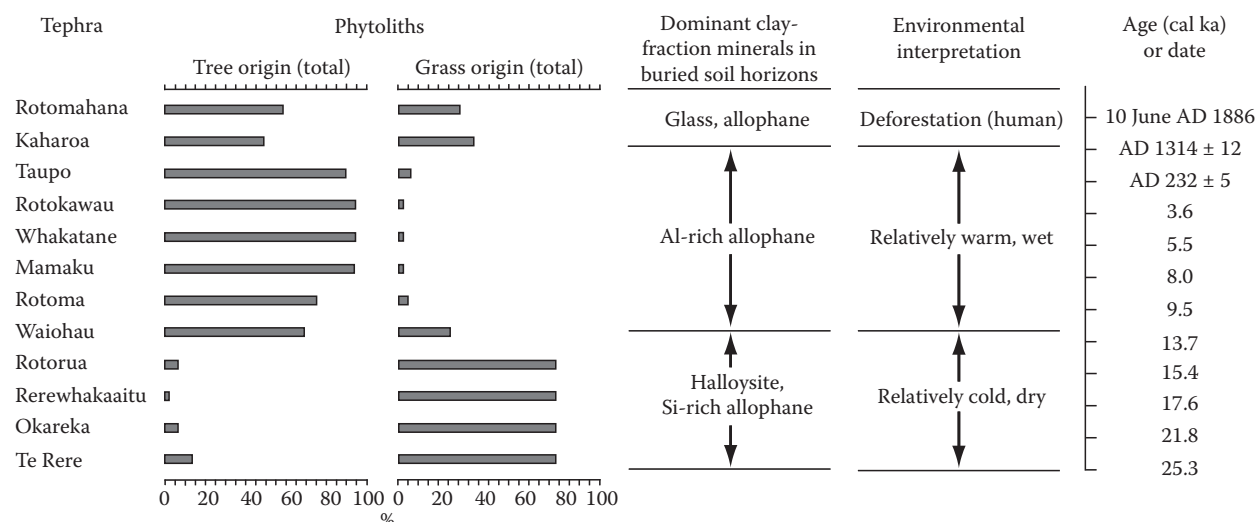
FIGURE 20.6 Dominant clay minerals and phytoliths identified in buried soil horizons in a sequence of 11 rhyolitic tephras and one basaltic tephra (Rotokawau) deposited near Rotorua, New Zealand, since ~25,000 calendar (cal) years ago showing the relationship between the clay mineral assemblages and environment (based on Green, 1987; Sase et al., 1988; Lowe and Percival, 1993; Kondo et al., 1994). ka, thousands of years before present. Dates for Kaharoa and Taupo eruptions are from Hogg et al. (2003, 2011); other ages are from Lowe et al. (2008). (Modified after Newnham, R.M., D.J. Lowe, and P.W. Williams. 1999. Quaternary environmental change in New Zealand: A review. Prog. Phys. Geog. 23:567–610. With permission of SAGE Publications.)

higher, colder sites, where temperature inhibited mineral alteration and the formation of new mineral phases. These altitudinal (hence rainfall) studies have further demonstrated the validity of Parfitt et al.'s (1983) model for the effect of [Si] in controlling the formation of secondary minerals from the weathering of volcanic parent materials (where Al availability effectively is unlimited). Further evidence for the validity of this model was demonstrated in a succession of buried soil horizons formed on rhyolitic tephras dating back ~25,000 years near Rotorua in New Zealand (Figure 20.6; Newnham et al., 1999). Cold and dry conditions and grassland vegetation before ~13,000 years ago resulted in halloysite (with minor Si-rich allophane) being formed because of limited Si leaching under lowered rainfall. Al-rich allophane was formed after ~13,000 years ago when Si loss increased through stronger leaching as rainfall and temperatures increased and forest replaced the grassland. In addition, halloysite formation was likely enhanced through enrichment by SiO_2 in leachates carried down the section as the tephras accumulated above, and by a concomitant reduction in permeability in the halloysitic horizons, which have a greater bulk density (Parfitt et al., 1983; Bakker et al., 1996). The paleoenvironmental interpretation was established by analyses of phytoliths (siliceous plant cell remains) at the same site together with palynological and other evidence from sites elsewhere in the region (Sase et al., 1988; Kondo et al., 1994; Newnham et al., 1999). A similar conclusion was reached in a parallel study in Mexico (Sedov et al., 2003).

The model was further supported by a study of seven pedons within the caldera of a volcano in Italy, which showed that the secondary minerals reflected the hydraulic properties, namely, drainage, of the parent materials (Vacca et al., 2003). In this case,

allophane had formed in soils developed in younger, porous, permeable ash deposits, while soils formed in older, less porous and less permeable scoria and consolidated tuffs contained crystalline minerals, especially halloysite, but no allophane.

Other aluminosilicate minerals can also form from volcanic parent materials by weathering. Possible products include 2:1 Si:Al aluminosilicates, which are most likely to occur where biotite is present as a product of the transformation of this mineral through the replacement of interlayer potassium ions, most probably by hydrated divalent cations (see Section 20.2.4.3). Various workers in the past (reviewed by Lowe, 1986) have proposed that a wide variety of 2:1 Si:Al aluminosilicates have formed in soils from volcanic parent materials, including volcanic glass, but several more recent publications (e.g., Nieuwenhuyse et al., 2000; Kautz and Ryan, 2003; Mirabella et al., 2005; Rasmussen et al., 2007) have confirmed that these types of phyllosilicates, including illite, smectite, vermiculite, chlorite, kaolinite, and their interstratifications with one another, may be inherited either from the parent pyroclastic material, where they may have formed by hydrothermal processes prior to weathering, or else as a contaminant of other, nonvolcanic origin such as aeolian dust.

Nonetheless, some aluminosilicate, and also more aluminous minerals, do form from volcanic parent materials either alongside, or instead of, allophane and halloysite (Theng et al., 1982). Parfitt et al. (1983) noted that gibbsite was recognized in a 1968 report as a product of the weathering of volcanic ash that was deposited in root channels in soils from Japan while it has also been found to occur in the weathering products of volcaniclastic materials in northern California that have rhyolitic, andesitic, and basaltic inputs (Takahashi et al., 1993). In these latter soils, gibbsite occurs alongside imogolite (or imogolite-like 2:1 Al:Si

allophane) and also halloysite. This concurrence suggests that gibbsite formation may occur during the wet season (winter and early spring) of the Mediterranean (or xeric) climate in northern California, while, as soils dry out in the approach to summer (the dry season), first imogolite/allophane, and then halloysite, is formed. In humid tropical conditions in Cameroon, the occurrence of minor gibbsite in a soil developed on hydrothermally altered nephenelitic materials (Si-poor, alkali-rich lavas) was attributed to the development of a strongly developed microporosity that facilitated the elimination of silica through leaching (Etame et al., 2009).

There appears to be an annual cycle among the secondary minerals resulting from the continually dissolving volcanic parent materials, particularly glass. Fieldes (1968) had proposed that halloysite formation occurred from allophane precursors as a result of seasonal drying (a process invoked also by Chadwick et al., 2003, whereby rapid wetting and drying cycles in arid zones were said to destabilize allophane, forcing it to "dewater and transform into halloysite"). Lilienfein et al. (2003) studied a chronosequence of quite young soils on andesitic mudflows in northern California nearby those in Takahashi et al.'s (1993) study and found that the amount of allophane increased linearly and quite rapidly in soils up to 600 years of age, although the oldest of these soils had less than half as much allophane as those studied by Takahashi et al. (1993) and hence were probably less well developed than the latter soils. Lilienfein et al. (2003) did not record any gibbsite in their soils.

By contrast, Nieuwenhuyse et al. (2000), who also studied a chronosequence on andesitic parent materials (lava, in this case) but in a humid tropical climate in Costa Rica, found all of allophane, kaolin minerals—kaolinite and halloysite—and gibbsite (as well as Al- and Fe-humus complexes and ferrihydrite) in soils from all ages of the parent lavas, ranging from 2,000 to ~450,000 years. Currently, mean monthly precipitation exceeds potential evaporation every month of the year in the present climate for these soils, so there is no dry season to explain the possible formation of relatively Si-rich kaolin minerals along with Si-free gibbsite and Si-poor (i.e., Al-rich) allophane. To explain the simultaneous appearance of these different minerals, Nieuwenhuyse et al. (2000), like Newnham et al. (1999), pointed to palynological evidence of a past drier climate, which evidently occurred in this region around the last glacial maximum. This drier climate, as at Rotorua in New Zealand (Figure 20.6), would have enabled Si-rich halloysite to form at that time, and some of it, at least, has persisted into the current strong leaching regime, with an average annual precipitation of 4500 mm year^{-1}. Nieuwenhuyse et al. (2000) proposed that gibbsite is formed in the soils as a result of the disintegration of the kaolin minerals under strong leaching. This mechanism explains the dominance of gibbsite among secondary minerals after 450,000 years of soil formation. By way of contrast, Certini et al. (2006) found that soils formed from 200,000 year old trachyandesite pyroclastic materials in Italy showed only well-formed gibbsite and an "embryonic" halloysite in their clay fractions. These authors

considered that both are products of the early stages of weathering of volcanic glass. In a hot, wet (perudic) climate in Guadeloupe similar to those studied by Nieuwenhuyse et al. (2000), Ndayiragije and Delvaux (2003) found that gibbsite, allophane, kaolinite, and hydroxyl-Al-interlayered smectite/vermiculite coexisted in a soil from andesitic–dacitic ash, but there was no halloysite in this soil. This soil appears to be similar mineralogically to those studied by Nieuwenhuyse et al. (2000), which occurs in a similar climate, except for the hydroxy-Al 2:1 Si:Al aluminosilicate. This last mineral was seen by Ndayiragije and Delvaux (2003) as an inheritance from the parent tephra and its aluminous interlayers were regarded as performing an antigibbsite (Jackson, 1963) or analogous antiallophane effect by preferentially sequestering Al, at least until the available interlayers became saturated.

Although activity of Si is clearly critically important in governing the formation of halloysite, allophane, or gibbsite in soils from volcanic parent materials (especially tephra) (see, e.g., a stability diagram for these minerals in Churchman, 2000, Figure 1.11), there may also be competitors for Al, such as hydroxyl-Al-interlayered smectite/vermiculite, as discussed by Ndayiragije and Delvaux (2003) (see also Kleber et al., 2007). Quite often competition for Al also arises from organic matter (especially large quantities are derived from, e.g., pampas grass *Miscanthus sinensis* in Japan), which gives rise to fast-forming, resistant Al-humus complexes in soils with pH < 5 (Shoji et al., 1993a; Dahlgren et al., 2004; Hiradate et al., 2004; Parfitt, 2009). Carboxyl groups of humic materials and the 2:1 layer silicates effectively compete for dissolved Al, leaving little Al available for coprecipitation with Si to form allophane or imogolite (Dahlgren et al., 2004; Theng and Yuan, 2008). The preferential incorporation of Al into Al-humus complexes, as with hydroxyl-Al interlayers of 2:1 layer silicates as noted above, is another example of the "antiallophane" effect (McDaniel et al., Section 33.3). Thus, preformed allophane disintegrated and was replaced by Al-humus after only 30 years under bracken fern, consistent with a fall in pH from 5.2 to 4.6 (Johnson-Maynard et al., 1997). Furthermore, weathering of tephra from the 1980 eruption of Mt. St. Helens led to Al-humus and also hydroxy-Al interlayers in 2:1 Si:Al aluminosilicates rather than to any of allophane, imogolite, or opaline silica (Dahlgren et al., 1997). However, opaline silica can form alongside Al-humus complexes, and Fe may also become incorporated into complexes with humus (Wada, 1989). Recent studies in Ecuador (Poulenard et al., 2003) and Japan (Yagasaki et al., 2006) have shown, respectively, that Al-humus complexes form rather than allophane when the content of organic C is particularly high or that allophane is dissolved in favor of Al-humus complexes when organic C content increases. As well, Al-humus complexing predominates at sites more distal to volcanoes because such sites infrequently receive a "top up" of Al through the deposition of weatherable tephras in comparison with sites closer to volcanic sources that are more regularly dusted with tephra.
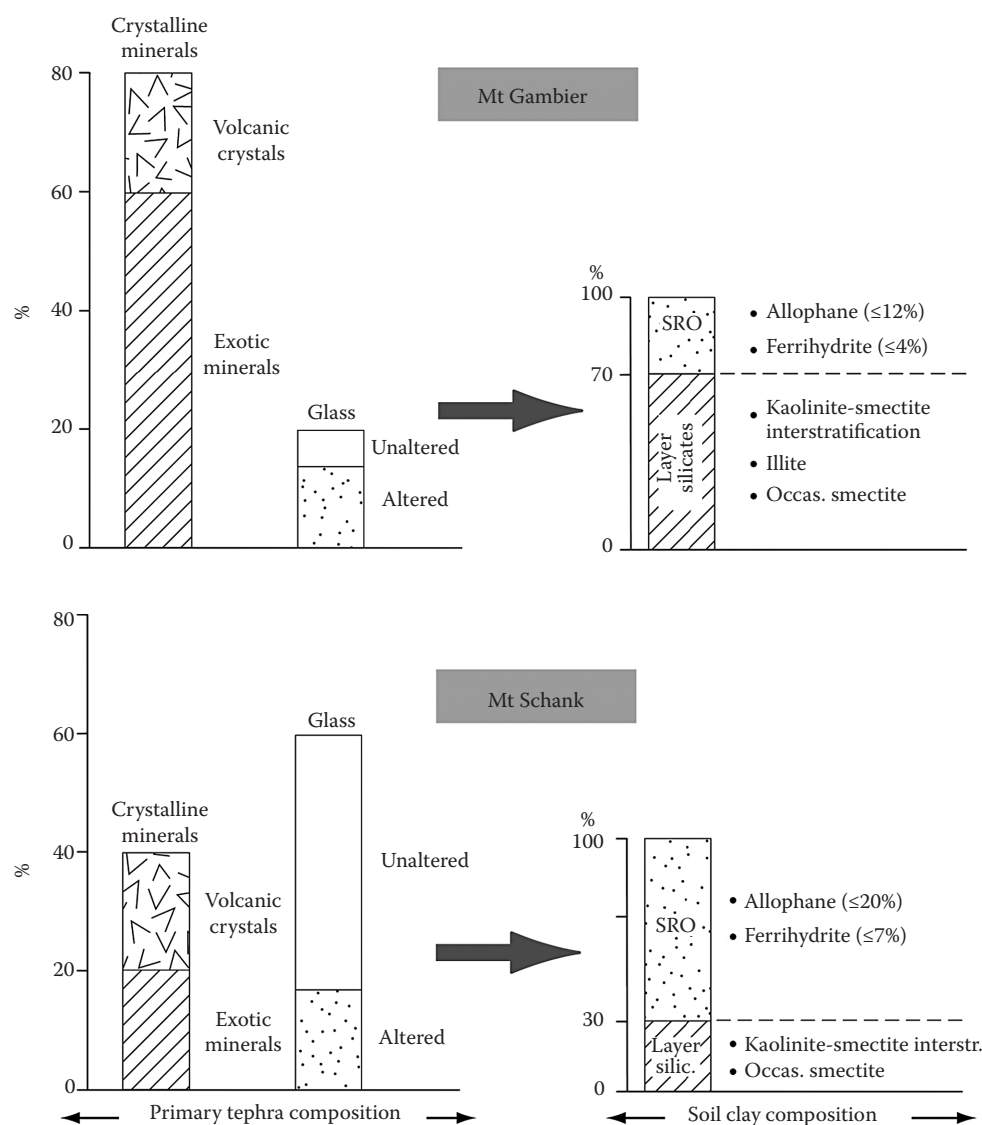
FIGURE 20.7 Contrasts in sand mineral assemblages of soils on mid-Holocene basaltic tephra erupted from Mt. Gambier and nearby Mt. Schank in South Australia have led to markedly different soil clay compositions under a xeric moisture regime. SRO, short-range order (i.e., nanocrystalline); interstr., interstratification; occas., occasional. (Modified after Lowe, D.J., and D.J. Palmer. 2005. Andisols of New Zealand and Australia. J. Integr. Field Sci. 2:39–65. With permission of Tohoku University.)

A variety of secondary minerals could be found in soils developed in the erupted material from the basaltic volcanoes at Mts. Gambier and Schank in southeast South Australia, which are only 10 km apart and both about 5000 years old (Lowe et al., 1996; Lowe and Palmer, 2005; Takesako et al., 2010). The main types of clay minerals formed in a xeric moisture regime with a relatively low average annual precipitation of 700 mm are shown for each location in Figure 20.7. They are compared with the categories of primary minerals in the parent materials, which comprise both volcanic and nonvolcanic "exotic" minerals incorporated into the eruptives from underlying calcareous sands and limestone. There is a contrast between the clays found at or near Mt. Gambier, on the one hand, and Mt. Schank, on the other. The

former are dominated by layer silicate minerals, including interstratifications of kaolinite and smectite, illite, and some discrete smectite. The secondary minerals in soils in the vicinity of Mt. Schank show less variety. They tend to contain more allophane and ferrihydrite than those from the Mt. Gambier area. For the soils from both sites, while both kaolinite–smectite and illite occur in many Australian soils (Norrish and Pickering, 1983; Churchman et al., 1994) and probably do not originate from the recent volcanic material, the discrete smectite found in poorly drained subsoils is probably formed by neogenesis from elements from the dissolution of the basaltic tephra (Lowe et al., 1996; Lowe and Palmer, 2005; Takesako et al., 2010). Notably, there was no halloysite formed in any of the soils at either Mt. Gambier

or Mt. Schank. Allophane showed a range of composition within profiles; its Al:Si ratio determined using acid oxalate-extractable Al (minus organically associated pyrophosphate-extractable Al) as a ratio to acid oxalate-extractable Si was Al:Si ~ 2 in upper parts of the soil profiles but dropped to Al:Si ~ 1 lower in the same profiles (Lowe et al., 1996; Lowe and Palmer, 2005). Lowe et al. (1996) suggested that there had been seasonal leaching during winter and early spring, leading to the formation of Al-rich allophane toward the surface, but weaker leaching (or even slight accumulation of Si) at depth, leaving Si-rich allophane rather than halloysite.

The volume of water draining through the upper part of the profile each year has been measured as about 280 mm, which just exceeds the threshold of about 250 mm for Al-rich allophane to form according to models derived from New Zealand data (Parfitt et al., 1984; Lowe, 1995). Jongmans et al. (1994) also found both Al-rich and Si-rich forms of allophane, respectively, in the B and C horizons, in the same profiles in soils in Guadeloupe. Alloway et al. (1992) similarly found differences in Al:Si ratios (and 15 bar water retention) in upbuilding Andisols in the Taranaki region of North Island, New Zealand, which they attributed to changing

climatic conditions from late glacial to postglacial periods; increasing rainfall and fewer prolonged dry periods resulted in higher Al:Si ratios because of increasing desilication.

Lowe (1986, 1995) and Lowe and Percival (1993) summarized data on the critical conditions for the formation of each of the main products of the weathering of volcanic ash. These are given in Figure 20.8, which also shows if Andisols are likely to have formed, and if they are of the allophanic or Al-humus (nonallophanic) type (McDaniel et al., Section 33.3).

A consideration of rates of chemical processes brings in the question of kinetics. Examples of kinetic studies dealing with the dissolution of glass in tephras and associated soils include those of Hodder et al. (1990, 1996) for rhyolitic tephras, Ruxton (1988), and Neall (1977) for dacitic and andesitic tephras, respectively, and Gislason and Oelkers (2003), Ziegler et al. (2003), Wolff-Boenisch et al. (2004), Shikazono et al. (2005), and Sigfusson et al. (2008) for basaltic tephras or lavas. Generally, basaltic and intermediate (andesitic, dacitic) tephras, with lower Si contents, tend to weather more readily than rhyolitic tephras, but in all cases glasses weather very quickly as noted earlier (Kirkman and McHardy, 1980;
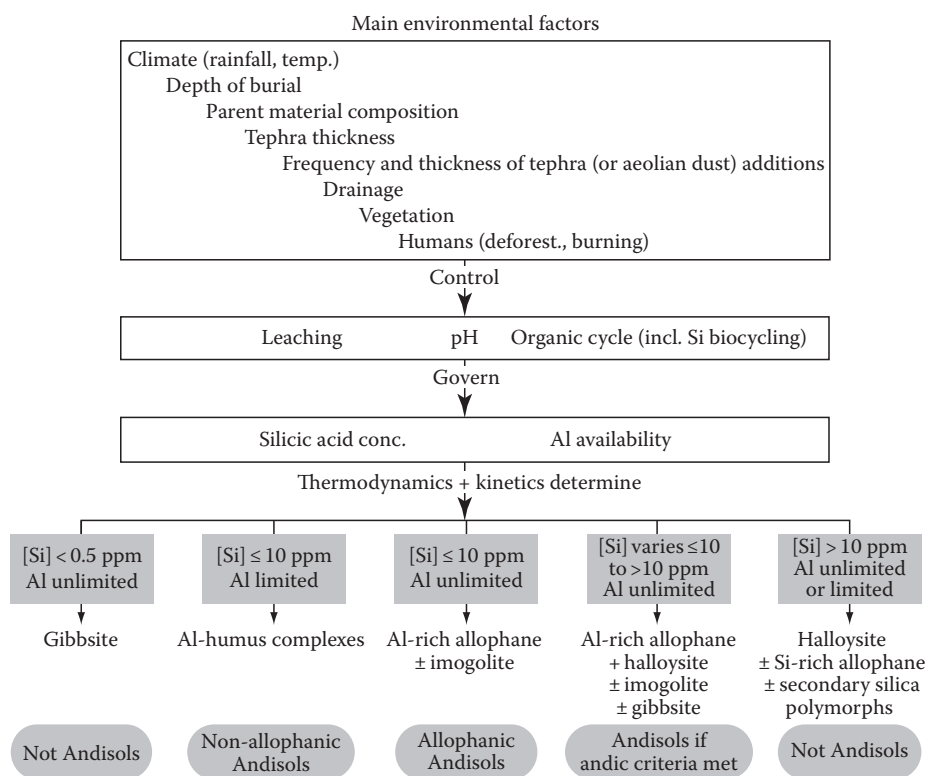
FIGURE 20.8 Environmental influences and controls that govern the critical conditions leading to the formation of different clays from the weathering of tephras, and the likely occurrence or not of Andisols as a result. ± indicates that the clay mentioned may also be present. (Modified after Lowe, D.J. 1986. Controls on the rates of weathering and clay mineral genesis in airfall tephras: A review and New Zealand case study, p. 265–330. *In* S.M. Colman and D.P. Dethier (eds.) Rates of chemical weathering of rocks and minerals. Academic Press, Orlando, FL; Lowe, D.J. 1995. Teaching clays: From ashes to allophane, p. 19–23. *In* G.J. Churchman, R.W. Fitzpatrick, and R.A. Eggleton (eds.) Clays: Controlling the environment. Proc. 10th Int. Clay Conf., 18–23 July 1993, Adelaide, Australia. CSIRO Publishing, Melbourne, Australia.)

Wolff-Boenisch et al., 2004). Compared with hard rock, the fragmental tephra components, especially vesicular glass and pumice, have a much greater surface area and high porosity and permeability, and so break down to constituent compounds very readily (Wolff-Boenisch et al., 2004).

Ziegler et al. (2003) suggested that kinetics controlled the formation of allophane and halloysite in arid soils on the basis of their work on a chronosequence of Hawaiian soils on basaltic lavas (with ash overlying lava in one case) in arid conditions (MAP 180–225 mm). Thermodynamics do not control the composition of soils unless soil solutions remain in contact with the mineral surfaces until equilibrium is reached. Where this does not occur, such as under arid conditions, kinetic factors control the soil system (Ziegler et al., 2003). Clay mineral synthesis in the Hawaiian arid-zone chronosequence was thus shown by Ziegler et al. (2003) to be controlled by the kinetics of soil drying, rather than thermodynamics, so that halloysite was the favored aluminosilicate end product with the formation of smectite inhibited by kinetic factors (and a lack of micas to "fuel" the dominant pathway for smectite formation; Chadwick et al., 2003). In Cameroon, the depth distributions of 0.7 and 1 nm halloysite in soils developed on weathered nephelinite lavas were able to be related to both kinetic and thermodynamic factors by Etame et al. (2009). That Si concentrations increased with depth in the profiles while Al remained relatively constant, and 1 nm halloysite occurred exclusively in lower horizons whereas both 0.7 and 1 nm halloysites were present in upper horizons, led Etame et al. (2009) to two suggestions. First, the 1 nm halloysite formation was controlled by direct precipitation from Si–Al-rich solutions (released from the weathering of primary minerals), and, therefore, the availability of Si was the only factor that controlled its formation at the base of the profile. Second, the presence of both forms of halloysite in the upper parts of the profiles implied that kinetics, hence time, controlled the evolution of halloysite through wet–dry seasonality, aided by the thermodynamically favorable factor of the availability of Si (Etame et al., 2009). In tropical Costa Rica, the predominance of 1.0 nm over 0.7 nm halloysite in the wetter, lower subsoil horizons of profiles studied by Kleber et al. (2007) was attributed to Si enrichment by percolating waters (i.e., mainly for kinetic reasons).

Returning to the question of whether it is time (Fieldes, 1955) or solution concentration/activity, particularly of Si (Parfitt et al., 1983, 1984), that governs whether or not allophane or halloysite persists in soils from volcanic parent materials (leaving aside Al-humus complexing), it may be said that while the evidence is very strongly in favor of solution conditions determining the nature of the minerals formed, the passage of time also appears to enhance halloysite formation at the expense of allophane via the so-called Ostwald ripening process (e.g., Chadwick and Chorover, 2001; Dahlgren et al., 2004; Rasmussen et al., 2007). This direction of change is consistent with the thermodynamic stabilities of the two mineral types (see, e.g., Figure 1.11 in Churchman, 2000). Structural considerations, however, mean

that an Al-rich allophane, with isolated tetrahedra on the inside of a spherule of coiled gibbsite-like sheets, cannot transform in the solid state into halloysite, where sheets of linked silica tetrahedra are generally coiled around alumina octahedra and hence are in the opposite conformation one to each other as in the allophane spherule. Therefore, dissolution and recrystallization must be involved in the change from allophane to halloysite. Nonetheless, it is likely that the passage of time effects other changes, which affect the environment in which mineral alteration and formation take place, so that conditions that favor the formation of allophane early in the weathering process change to those favoring halloysite formation. Most often in upbuilding terrains, this comes about through an increase in [Si], which may result from a buildup of overburden through ongoing tephra deposition, or the development of impediments to efficient drainage (e.g., Cronin et al., 1996). Another possible change may be toward a drier climate, so that there is less throughflow of water and hence a buildup of Si in solution. An alternative situation is that reported by Chadwick and Chorover (2001) for a Hawaiian chronosequence on basaltic lavas and ash deposits not subject to ongoing burial by later eruptives. There, the ongoing depletion of [Si] and increase in [H^+] under high rainfall after ca. 400,000 years eventually led to the formation of more stable kaolin (halloysite and kaolinite) and gibbsite. Previously, conditions had apparently favored the formation of allophane. Chadwick and Chorover (2001) noted that [Al] in solution also declined over time, a result that follows from increasing crystallinity and decreasing solubility of gibbsite and kaolin in the soil. Inconsistencies between observed soil mineral composition and stability-field plots of solution chemistry data are evidence of kinetic limitations to mineral transformations according to Chadwick and Chorover (2001), who also noted that such discrepancies can be resolved through consideration of mineral transformation rates and Ostwald ripening processes but that the requisite kinetic data are often lacking.

In considering longer timescales of glacial and interglacial cycles in temperate volcanic landscapes not directly glacierized, the marine oxygen isotope records show that cooler and drier conditions associated with glaciations persisted ~80%–90% of the time whereas warmer and wetter conditions associated with interglaciations occurred for ~10% to 20% of the time. In these landscapes, where long sequences are well preserved, therefore, it might be expected that soils developed from accumulations of weathered tephras of similar composition dating back several hundreds of thousands of years to one or two millions of years should be dominated by halloysite rather than allophane (e.g., as evident in the so-called Kauroa and Hamilton ash tephra sequences in northern New Zealand that date back to ca. 2.3 Ma; Lowe and Percival, 1993; Lowe et al., 2001).

Overall, the relationship between the minerals formed from volcanic parent materials (especially tephra) and solution [Si], as it is affected by the throughflow of water and given appropriate kinetic circumstances, can be expressed by Box 20.3:

Box 20.3   Relationship between Throughflow of Water, Si Concentration, and Types of Clays Formed from Volcanic Parent Materials

	Smectite	Halloysite/Si-rich Allophane	Al-rich Allophane	Gibbsite
Al:Si	0.5	1.0	2.0	∞

Increasing rate of throughflow of water →

Decreasing Si concentration/activity
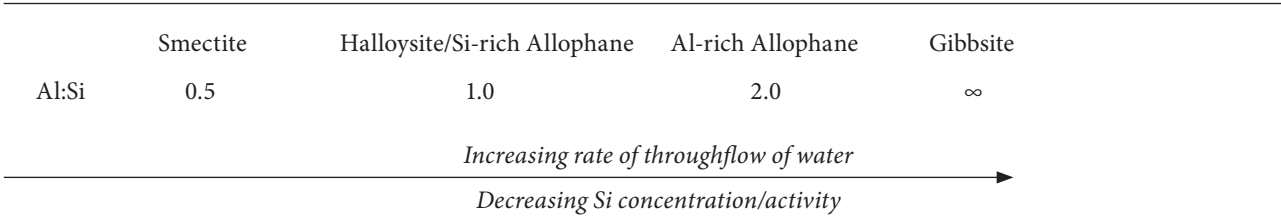
We note in addition that allophane can be formed in a range of nonvolcanic materials through strong leaching and acid conditions, as occur, for example, in loess in upland areas of southern New Zealand (e.g., Eger and Hewitt, 2008), and as a result of podzolization (see Section 20.2.5.3.2).

Parallel to the alteration of aluminosilicate minerals and the formation of allophane, halloysite, and perhaps also gibbsite or smectites, iron is also lost from parent Fe-bearing minerals and incorporated into secondary solid phases. Ferrihydrite forms early in the process of weathering of volcanic parent materials from the precipitation and oxidation to Fe(III) of soluble Fe(II) that had been released into solution from volcanic glass and also feldspars and mafic minerals (e.g., McDaniel et al., Section 33.3). It is the presence of other species in solution such as silicate and phosphate ions and soluble organic compounds that inhibit the formation of the more crystalline forms of iron, such as goethite and hematite (e.g., Childs, 1992; Bigham et al., 2002). However, the type of iron oxide that is formed is a function of the same environmental conditions that also affect which particular aluminosilicates are formed in volcanogenic soils. Hence, in young soils on andesitic mudflows, the buildup of ferrihydrite with time paralleled that of allophane, albeit that the amount of allophane formed was ~10 times or more than that of ferrihydrite (Lilienfein et al., 2003). In a chronosequence of soils in the humid tropics, Nieuwenhuyse et al. (2000) found that Fe occurred mainly as Fe-humus in the two younger soils, especially in their A horizons, but also as ferrihydrite, especially at depth, whereas it occurred mainly as goethite throughout the oldest soil in the sequence. Further, the iron oxides in soils in a xeric moisture regime that contain all of gibbsite, imogolite and/or Al-rich allophane, and halloysite as aluminosilicates were dominantly highly crystalline—they were mainly goethite, reflecting strong seasonal drying (Takahashi et al., 1993). Soils in a toposequence on andesitic lahar deposits covering a range of climatic zones generally showed a similar pattern to that for the aluminosilicates, with poorly crystalline forms giving a high oxalate-Fe analysis dominant in soils at higher altitudes and more crystalline forms giving a high dithionite-Fe analysis in soils at lower altitudes (Rasmussen et al., 2007). There was a further subtlety with Fe analyses, however, insofar as the proportion of oxalate-Fe to dithionite-Fe rose again at the lowest altitude, where the MAP was also lowest. In this drier zone, less leaching probably led to higher retention of Si in soil solution, which would inhibit Fe oxide crystallization (Rasmussen et al., 2007).

20.2.4.6.3 *The Special Case of "Palagonite"*

In the volcanic and geochemical literature especially, the term "palagonite" is widely used to describe the first stable product of basaltic volcanic glass alteration (e.g., Fisher and Schminke, 1984; Cas and Wright, 1987; Daux et al., 1994). Palagonite has also been described in the clay science literature but uncommonly (e.g., Singer and Banin, 1990; Drief and Schiffman, 2004) and it thus deserves some attention here. As reviewed by Stroncik and Schmincke (2002), palagonite (named after the location Palagonia in Sicily) is a heterogeneous substance, usually with highly variable optical and structural properties ranging from a clear, transparent, isotropic, smooth, and often concentrically banded material, commonly called "gel-palagonite," to a translucent, anisotropic, slightly to strongly birefringent material of fibrous, lath-like, or granular structure, commonly called "fibro-palagonite." It ranges in color from yellowish to shades of brown. Since it was first defined in 1845 as a new mineral, palagonite has been interpreted in numerous studies to be a heterogeneous material composed variously of different aluminosilicate clays, zeolites, and oxides, or mixtures of these. Based on XRD, HRTEM, atomic force microscopy (AFM), and electron microprobe analyses, palagonite is now interpreted by most to be composed of a variety of smectites and very minor amounts, if any, of zeolites and oxides (Stroncik and Schmincke, 2002). At more advanced stages of alteration, other secondary phases are known to form following Ostwald "ripening" processes.

Palagonite evidently comprises spherical structures 20–60 nm in diameter, which have been interpreted as "microcrystallite precursors of smectite" (Stroncik and Schmincke, 2002)—that is, palagonite would seem to be a nanocrystalline variety of smectite(s). Drief and Schiffman (2004) identified an Fe-rich montmorillonite-like composition in their study on Hawaiian basaltic deposits. The atomic structure of palagonite and, especially, its structural evolution are not yet fully understood, but both are important to glass-alteration rates and to the evolution of the whole water–rock–soil system (Stroncik and Schmincke, 2001). Because kinetic and thermodynamic modeling and mass–balance calculations necessitate exact differentiation of all the secondary products developing during alteration, the term "palagonite" should be used only for the hydrous nanocrystalline alteration product ("gel-palagonite"), not for the long-range-ordered crystalline material evolving from the palagonite itself (Stroncik and Schmincke, 2001, 2002).

The so-called palagonitization of basaltic volcanic glass is a continuous process of glass dissolution and palagonite formation and evolution, which, according to Stroncik and Schmincke (2001), can be subdivided into two different reaction stages with changing element mobilities. Stage 1 is the congruent hydration and dissolution of thermodynamically unstable glass and contemporaneous precipitation of "fresh," gel-like, amorphous (or nanocrystalline), optically isotropic, mainly yellowish palagonite. This stage, kinetically controlled and consistent with the models of Hodder et al. (1990, 1993), is accompanied by the loss of Si, Al, Mg, Ca, Na, and K, active enrichment (gain) of H_2O, and the passive enrichment ("default" accumulation relative to other elements) of Ti and Fe. Stage 2 is an aging process during which the thermodynamically unstable palagonite reacts with the surrounding fluid and crystallizes to smectite. This stage is accompanied by uptake of Si, Al, Mg, and K from solution and the loss of Ti and H_2O. Ca and Na are still showing losses, whereas Fe reacts less consistently, remaining either unchanged or showing losses (Stroncik and Schmincke, 2001, 2002).

20.2.5 Peculiarities of Processes and Products of Alteration by Weathering in Soils

Much—although not all—of the information on the processes and products of alteration of primary minerals by weathering that is summarized in the preceding Section 20.2.4 comes from observations of rocks or geological deposits, minerals associated with these materials, and saprolites deep within soil profiles. However, soil profiles can encompass a range of conditions, from fragmented, almost abiotic, rock through to organic-rich material containing a great variety of microbial, faunal, and also plant life. The different parts of soil profiles represent different regimes for water, from abundant and mobile to scarce and largely immobile, so studies of some nonsoil situations where each of these prevail are relevant to weathering in some parts of soils but not in others. Most soils have a significant biological input, at least in their surface layers, and, therefore, the biological factor in mineral alteration and formation tends to play a larger role in

the genesis of soil minerals than in those from rocks subject to largely abiotic weathering. Furthermore, some processes, notably podzolization, occur within permeable soils but not in impermeable rocks under certain environmental conditions and deserve particular attention. In addition, soils can be further developed even following the depletion of easily weatherable rock-derived minerals. Secondary minerals can themselves be altered and new phases formed. Each of these particular aspects of mineral development in soils will be discussed in the following sections.

20.2.5.1 The Effect of Position within Soil Profiles on Hard Rock

In a recent review, Wilson (2004) showed that position in the weathering profile on indurated rock has a crucial influence on the nature of the environment for weathering and hence on the processes involved and products formed, as well as the rates at which changes occur. Note that modified models are required for soils formed on unconsolidated materials or, especially, those developed through upbuilding pedogenesis. Following Delvigne (1998), a fully developed weathering profile comprises a soil, which is delineated (in a straightforward case) into A and B horizons (solum) and usually C or CR horizons comprising saprolite/alterite, which consists of completely disaggregated rock grading to the rock base, with fabric and texture changing with depth. At the base of the whole profile is the (partly) weathered rock (R horizon), which may be hard and coherent but also incorporates fractures and fissures. Among other features, an increase in surface area characterizes the upward change from the rock to the solum (Hochella and Banfield, 1995).

There is also a trend of decreasing porosity and permeability from soil down to rock, so that water can flow quite freely within the soil, but can be either stagnant or held in capillaries in the rock. Hence, there is a much greater chance of an equilibrium being established between the minerals and the largely immobile solution within the rock than between the minerals and the highly dynamic water in the soils zone.

Following Velde and Meunier (2008), we note that alteration within the rock proceeds in microsystems. Meunier and Velde (1979) followed the alteration of granites by weathering through developments occurring at the scale of microsites. The system is highly heterogeneous at this fine scale; grains of the primary minerals, muscovite, orthoclase, and biotite each give rise to multiphase assemblages upon alteration. Alteration of the whole mineral assemblage appears to take place in three stages. In the first stage, alkalis are lost but the Si:Al ratio is preserved as illitic mica, beidellite, and vermiculite are variously formed. The second stage involves a loss of both alkalis and silica with kaolinite and oxides resulting. The authors surmised that there is a third stage where quartz is dissolved, leaving only oxides, but not silica. This last stage is seen in tropical weathering. The processes of in situ alteration of ferromagnesian minerals to smectites and of feldspars to illite that were studied by Eggleton and coworkers and Banfield and coworkers, among others, and cited in Section 20.2.4 herein, also occur within microsites where the physical constraints as well as the chemical availability of reagents limit the reactions that

can take place. In a particularly close study of microsites within one mineral type (amphibole) on weathering, Proust et al. (2006) found that montmorillonite formed on the (001) amphibole face and saponite on the (110) face, while the sawtooth (001) face fracture surface hosted a kaolinite–smectite and, with time, also halloysite and montmorillonite. The establishment of local equilibria may occur at different sites, within both a single crystal, as in the amphibole studied by Proust et al. (2006), and also within different fractures and fissures in a rock (see the diagrammatic representation for granite in Figure 4.19 of Velde and Meunier, 2008). Within microsites where secondary products remain in close proximity to host primary minerals, reactions may even occur between product and host minerals (Velde and Meunier, 2008). As a result, vermiculite may form by recrystallization in a microsite from the dissolution products of the adjacent secondary minerals, saponite or Fe–beidellite and their primary precursor, amphibole for one example (Velde and Meunier, 2008). For another example (Velde and Meunier, 2008), product smectite and host K-feldspar may react together to give rise to illite and kaolinite.

The solid:solution ratio is very high in hard rocks that remain coherent yet contain small fractures and fissures. By great contrast, the solid:solution ratio in the solum or unconsolidated geological deposits within the weathering profile is likely to be both very low and quite variable on both a daily and seasonal basis, except in special circumstances, for example, in a desert or frozen soil. The dynamic nature of soils alone ensures that the possibility of predicting the course of weathering via stability diagrams for the component clay minerals (e.g., Kittrick, 1967) cannot be realized. Minerals awash in the dynamic water phase in soils almost certainly remain well out of equilibrium with those solutions (Chadwick and Chorover, 2001; Wilson, 2004). However, those formed in microsites within weathering rocks deeper in the weathering profile may not appear in bulk analyses of soils above the weathering rock. In particular, trioctahedral smectites formed during early stages of alteration are rarely reported in soils. Either they are unstable in the soil environment (Wilson, 2004) or they are present in only vanishingly small concentrations (Proust et al., 2006), or both.

20.2.5.2 Weathering in the Absence or Shortfalls of Water

Generally on Earth, weathering occurs because there is sufficient—and usually excess—water to effect alteration of primary minerals and either transformation to related secondary phases or neogenesis of new (secondary) minerals. However, mineral alteration and formation are also observed to occur where there is little, or no, water. Studies in the normally frozen Antarctic have shown clear evidence for chemical weathering occurring there. As well as salts such as gypsum and sodium sulfate formed by recrystallization from elements leached from rocks by occasional (liquid) water, abundant iron oxides also occur (Claridge, 1965). Furthermore, clay minerals occur—most commonly transformation products of micas such as vermiculites, but also montmorillonite. This last mineral was thought to form in the arid and highly alkaline environment by similar processes to

those seen in other terrestrial environments, but occurring at a much slower rate (Claridge, 1965). Some, but not all, clay minerals in Antarctic soils could be inherited from other land masses to the north (e.g., dust blown from southern South America; Delmonte et al., 2004). Boyer (1975) collected weathering products such as rinds on rocks from part of Antarctica and identified various 2:1 minerals, including chlorite–vermiculite and also identified montmorillonite. These products all indicated that chemical weathering had taken place. In the arid Sahel in west Africa, Ducloux et al. (2002) identified beidellite, kaolinite, and mica in soils. They concluded that these minerals formed by a process called "xerolysis," whereby desiccation in a hot climate leads to protons from dissociated water reacting with crystalline minerals to produce secondary products.

The new frontier for clay mineralogy is outer space, and especially other planets in our solar system, and there has been a particular focus on Mars from a clay mineralogical viewpoint quite recently. Earlier exploration of Mars involved fly-bys, with spectra being used to identify materials on the Martian surface by remote sensing. These data could have been interpreted to indicate the presence of some phyllosilicates (Hamilton et al., 2003), but the evidence was equivocal. The collection of samples by Exploration Rover from the planet's surface showed that the ferric sulfate, jarosite, which is commonly associated with acid sulfate soils on Earth (see Section 20.2.5.3.4), and also possible relicts of gypsum at the Meridian Planum landing site (Madden et al., 2004). Other work on samples from a number of different locations on the planet has confirmed the widespread occurrence of secondary sulfates and also iron oxides (Hurowitz and McLennan, 2007). Hurowitz and McLennan (2007) considered that they originated from altered olivines in Martian rocks. These various localized occurrences of sulfates and iron oxides were compared with the results of a general survey of most of the surface of Mars that was carried out by an image spectrometer during an aerial survey of the planet (Poulet et al., 2005). The survey had earlier identified sulfates in localities in addition to those reported by Madden et al. (2004) and Hurowitz and McLennan (2007), but Poulet et al. (2005) also identified phyllosilicates in several localities, albeit that their distribution was quite heterogeneous. They included Fe/Mg smectites in some areas and montmorillonite in others. These occurrences suggested that there had been water in some localities on Mars, in order to alter basaltic and similar rocks to give phyllosilicates, although the sulfates derive from a different, more acidic, environment. More recent work (e.g., Bishop et al., 2008; Mustard et al., 2008; Ehlmann et al., 2009) has identified phyllosilicates in a number of localities on the surface of Mars. The discovery of phyllosilicates has aided our understanding of the climatic history of Mars (Newsom, 2005; Poulet et al., 2005). On Earth, meteorites considered to have come from Mars have been studied for their mineralogy, and iddingsite, comprising a mixture of smectites, iron oxides, and silica, has been identified on the altered surfaces and in the veins of the meteorites (Wentworth et al., 2005). These authors drew an analogy between mineral alteration and formation that has taken place on Mars with that which has occurred—and is

occurring—in the Dry Valleys of Antarctica. Examined at the submicroscopic scale, the patterns of alteration of primary silicates, for example, amphiboles and pyroxenes, were similar in an Antarctic Dry Valley (Wright Valley) and in the Mars meteorites (Wentworth et al., 2005). In addition, soils in the Dry Valleys contain evaporated salts at their surfaces, as have been seen in soils on Mars. In both of these arid situations, alteration processes and their products are distributed heterogeneously, indicating an uneven distribution of water available for the alteration and subsequent formation of minerals, both in space (in the Antarctic Dry Valleys) and in time (in the Martian past).

20.2.5.3 Biota as Important Agents in Mineral Alteration and Formation

It has long been recognized, including by the pioneers in pedology, for example, V.V. Dokuchaev and Hans Jenny, that biota play an important, sometimes dominant, role as one of the agents in the development of soils from rocks and other geological materials. Plants, animals, and microbes share an equivalent status to parent materials, climate, topography, and time as soil-forming factors. Therefore, it comes as no surprise to find that biota in their different forms demonstrably play a major role in the alteration of primary minerals and the formation of secondary clay minerals.

Even so, it is arguable that the role of biota in the formation of clay minerals has been relatively neglected in comparison with those of other soil-forming factors, such as parent materials and climate. Churchman (2000) noted that studies of the alteration and formation of minerals had largely concentrated on processes involving inorganic agents, while observing that there had been an upsurge in studies of the role played by biological and biochemical agents in the previous decade. The effect of these agents on weathering changes is now one of the more active areas of research on mineral alteration and formation in relation to soils (e.g., Dong et al., 2009).

20.2.5.3.1 Plants as Weathering Agents

Plants have been shown to be important weathering agents. Mortland et al. (1956), Hinsinger et al. (1992, 1993), and Hinsinger and Jaillard (1993) showed that trioctahedral micas in soils lost potassium from their interlayers as a result of the growth of plants in the soils. In general, this process is that of vermiculitization (or the formation of beidellitic smectites) (Section 20.2.4.3). However, it generally occurs via mixed-layer randomly interstratified species. In soils, illite–smectites are the most common form of these (Velde and Meunier, 2008). Indeed, since this type of clay mineral is not found in buried sediments or rocks, it is peculiar to the soil environment, according to these authors. Furthermore, it is found particularly in the A and B horizons of soils (the solum), confirming its origin in surface alteration processes (ibid.). Although plants, in general, are the main agents of alteration of trioctahedral micas to produce illite–smectites, Velde and Meunier (2008) summarized observations in the literature to conclude that prairie soils (Mollisols) dominated by grasses contain illite–smectite together with varying amounts of illite under different climate conditions. Even

though vermiculites are rare in prairie soils, they can occur in significant amounts in young mountain soils according to Velde and Meunier (2008). Expanded (to vermiculite or smectite) and partially expanded (to mixed-layer illite–smectite) products occur, and these often acquire Al-hydroxy interlayers (see Section 20.2.4.3) under forest (Velde and Meunier, 2008). It was the development of a method of XRD peak decomposition by Lanson (1997) that enabled the detection of some of these mineral changes attributable to plants, which can be subtle.

Acidification can be an important effect of plant growth and consequently can play an important role in alteration by weathering. Protons are released into soil solutions to balance the overall charge when roots take up more cations than anions (e.g., Hinsinger et al., 2001; Courchesne, 2006; Calvaruso et al., 2009). For some tree species, it is the uptake of NH_4^+ in preference to NO_3^- by roots that ensures acidification, specifically in the rhizosphere (Calvaruso et al., 2009). Roots (and fungal hyphae tips as noted above) also exude organic acids into soils (Courchesne, 2006). More Ca, Na, Mg, Si, and especially Fe were released from leached basalt when plants were grown in the fresh rock than when there were no plants (Hinsinger et al., 2001). When soils within the rhizosphere of a Norway spruce and an oak were compared with the bulk soil, it was found that Fe and Al, and therefore their "amorphous" (i.e., nanocrystalline) mineral phases, as well as Al-hydroxy interlayers in 2:1 Si:Al aluminosilicates, were dissolved as a result of acidification by the plant roots (Calvaruso et al., 2009). Forest soils in Taiwan showed a lack of hydroxyl interlayering of 2:1 Si:Al aluminosilicates in the surface O, A, and E horizons but there was interlayering, which was shown to involve both Al and Fe, in deeper horizons (Pai et al., 2004). It was thought that a low pH and possible chelation of Al and Fe by organic material, which was abundant in and on the surfaces of the soils, led to the clean interlayers in the expandable aluminosilicates near the surface, whereas the higher pH and lower organic contents in lower horizons were conducive to their formation in the clay minerals there (Pai et al., 2004). In a later study, Pai et al. (2007) found that acidification at the surface of forest soils elsewhere in Taiwan decreased the charge of the K-depleted aluminosilicate layers, which they characterized as vermiculite, occurring either as the discrete phase or in mixed-layers with vermiculite. In the North Island of New Zealand, Jongkind and Buurman (2006) found that weathering under kauri (*Agathis australis*) trees, which produce especially low pHs (4.0 ± 0.2) left vermiculites depleted of hydroxyl interlayering and even brought about some conversion of vermiculite to smectite, as Churchman (1980) had found to occur in soils under another native (beech) tree species (*Nothofagus* sp.) in the South Island of New Zealand. By matching the rare-earth compositions of the silica particles found within the roots of ferns with those of aluminosilicate minerals in the soil surrounding the roots, Fu et al. (2002) found strong evidence for the dissolution of clay minerals to leave deposits of silica following their incorporation into plant roots.

Plants can bring about very rapid changes in soil minerals (Turpault et al., 2008). A seasonal study by these authors showed that in the rhizosphere close to tree roots, mineral dissolution,

with loss of citrate-extractable Al and Si, occurred in surface layers in spring because of enhanced biological activity. That some of the Al came from interlayers was shown in XRD by an increased ease of collapse of these upon K-saturation and also an increase in cation exchange capacity (CEC). There was a corresponding migration of Al and Si to lower layers. Perceptible changes in the XRD patterns occurred within only 3 months.

20.2.5.3.2 *The Special Case of Podzolization*

Organic compounds from the activities and decomposition of biota have long been thought responsible for extracting and transporting Fe and Al ions and incidentally producing new minerals from them in the process of podzolization (e.g., Russell, 1973), even if equally valid inorganic mechanisms have also been put forward for this process (e.g., Farmer et al., 1980; Wang et al., 1986), as will be discussed further.

Podzolization usually involves the intense leaching of parent materials and soils with acidic solutions. In podzolization, Al, Fe, Mg, and Si are mobilized (e.g., Farmer, 1982; Farmer and Fraser, 1982; Farmer et al., 1983; Taylor, 1988; Churchman, 2000; Giesler et al., 2000). Podzolization can be contrasted with the processes, sometimes collectively referred to as "andosolization," involved in forming allophane and other andic soil materials mainly in tephra-derived soils (Section 20.2.4.6). Normally, allophane and other nanocrystalline clays are formed in situ during andosolization (even though the process typically requires desilication via leaching) whereas podzolization always involves the *translocation* of mobile constituents as organic or inorganic complexes, or both, from very acid horizons in the upper profile and subsequent immobilization in lower horizons through an increase in pH, via microbial action, or by adsorption, precipitation, etc. That such movement takes place has been confirmed by micromorphological studies (e.g., Farmer et al., 1980, 1983, 1985; Farmer and Lumsdon, 2001). Broadly speaking, the processes involved in podzolization are as follows: (1) enhanced mineral weathering in upper A and E ("eluvial") horizons; (2) transport of Al, Fe, and Si and probably also organic matter from upper horizons to lower B horizons—either in metal–organic complexes (chelates) or as independent inorganic complexes including "proto-imogolite sols," and as independent organic complexes; and (3) deposition of the Al and Fe as oxides, together usually with allophane, from (according to Farmer et al., 1985) a mixed Al_2O_3–Fe_2O_3–SiO_2–H_2O sol, in so-called illuvial Bs horizons.

There have been a number of theories put forward to explain the changes that take place in soils as a result of podzolization (see Churchman, 2000, for proponents and arguments involved in earlier discussions, and elsewhere in this handbook). Briefly, these can be grouped into two main hypotheses, both of which may apply to some degree depending on site-specific soil environments (e.g., Wang et al., 1986; Lundström et al., 2000): (1) the organic or "fulvate" hypothesis and (2) the inorganic or "proto-imogolite" hypothesis. A third is the so-called rock-eating fungi hypothesis, which is associated with mycorrhizal fungi found in many soils under pines and heaths commonly in parts of Europe and Scandinavia and elsewhere (Jongmans et al., 1997; van Breemen et al., 2000a).

The fungal hyphal tips, which produce organic acids (e.g., citric or oxalic), "drill" directly into mineral grains including feldspars, thereby bypassing the external acid soil solution, and hence the E horizon is regarded as the fungal "eaten" part of the soil.

An eccentric view is that of Do Nascimento et al. (2004, p. 536) who have described podzolization as a deferralitization process whereby the chemical elements that had accumulated during ferralitization are "progressively exported towards the drainage network under the combined effects of waterlogging, organic matter accumulation, and lateral subsurface and surface water flows." In a somewhat related study, Lucas (2001) compared the dynamics operating in podzols with those in ferralsols (Oxisols) regarding the rates of Si recycling (including storage as phytoliths) by plants and the rates of loss, by leaching, of Si and organometallic complexes involving Al, Fe, and organic matter. This author concluded that the main constituents of the "weathering mantle" ultimately depended on the balance between (1) the stability of the clay minerals, sustained by the plant Si cycling, and (2) the leaching of plant-induced organo-Al compounds. These findings were evident in equatorial areas where old soils are markedly affected by biological activity, and they were also seen in temperate areas (Lucas, 2001).

After reviewing the literature, Lundström et al. (2000) concluded that organic complexes were important in both the weathering and deposition (or immobilization) processes and Jansen et al. (2005) also decided, on experimental grounds, that organic matter played an important role in the transport of Al and Fe down profiles to create illuvial horizons, at least in podzols in the Netherlands and similar climatic zones, while conceding that podzols may form by slightly different mechanisms in both cooler and warmer climatic zones. Contrary views are evident in, for example, Farmer and Lumsdon (2001) who argued that "fulvic acid plays no active role in podzolization, but only recycles Al and Fe, that has been transferred by biological processes to the O horizon, back to the Bh horizon" (p. 177). Buurman and Jongmans (2005) proposed that podzolization in boreal zones, on nutrient-rich parent materials at high latitudes and high altitudes, hence cold climates, occurred with organic matter (OM) dynamics that led to little accumulation of OM (so-called "fast" OM dynamics), and, hence, illuvial B horizons were dominantly inorganic. These authors proposed an alternative route for podzolization in hydromorphic situations on nutrient-poor parent materials, including in warmer climates, where OM dynamics were "slow," by their nomenclature and would lead to organic-rich illuvial horizons.

Regardless of mechanism, podzolization leads to essentially two compartments of the soils according to their weathering status and the nature of the secondary phases (Ugolini et al., 1991). These are (1) the upper horizons, and especially the distinct eluvial (or "fungal eaten") E horizon, and (2) the taxonomically diagnostic lower illuvial horizons, and especially the equally distinct Bh and Bs horizons. (Both or only one of the Bh or Bs may be present in podzol soils.) Considerable recent effort has been expended into studying the early stages of podzolization, largely in soils in the boreal zone, namely, under forests, in northern and/or alpine Europe, from a mineralogical viewpoint

(Melkerud et al., 2000; Mossin et al., 2002; Mokma et al., 2004). Mokma et al. (2004), tracing the earliest stages of podzolization in sandy soils in Finland, found visual evidence for translocation of C, Al, and Fe after 230 years of soil formation while there was evidence of allophane in a Bs horizon after 900 years. There were few differences between the crystalline minerals in the soils of different ages (up to 11,300 years). Buried podzol profiles in northern New Zealand exhibit thin E, Bh, and Bs horizons (about 20, 2, and 5 cm in thickness, respectively) formed on a pumiceous rhyolitic tephra layer deposited in ca. AD 232. The Bs horizons contain allophane. The developing podzols were subsequently buried in ca. AD 1314 by deposition of another rhyolitic tephra (Lowe, 2008). The profiles thus reflect about 1082 years of podzolization under ~1500 mm annual rainfall in a mesic temperature regime under broadleaf–podocarp forest.

In three soils on Quaternary deposits in Fennoscandia, Melkerud et al. (2000) found that the most easily weatherable primary minerals, biotite, chlorite, and hornblende, were depleted from the eluvial horizon, which had the highest concentration of quartz. Vermiculite dominated the clay mineralogy of the E horizons and was also present, but with incorporated Al-hydroxy interlayers, in the B horizon, where poorly ordered or nanocrystalline imogolite-type materials (i.e., allophane) also occurred. Mossin et al. (2002) specifically searched for imogolite in three Danish podzols. Some proto-imogolite allophane rather than imogolite per se was found, but in smaller amounts than had been found in podzols further north in Scandinavia and not at all under spruce, which was responsible for the lowest pH for the soils. They ascribed the generally relatively lower contents of allophane to a lower pH, an increase in organic matter, and a parent material with a low content of easily weatherable minerals. The pH conditions and other factors such as aluminum extractability and solubility are important in explaining the differences in composition in different parts of the podzol profile (e.g., Zysset et al., 1999; Lundström et al., 2000; Farmer and Lumsdon, 2001).

Recently, several other authors have particularly examined the effect that the podzolization, namely, acid leaching, process has had upon micaceous clay minerals. The results of some earlier studies of this kind are discussed in Section 20.2.4.3 and also earlier in this section herein and in Churchman (2000). While the general consensus from earlier studies was that beidellitic smectites are often identified as a stable phase, if not an end product, of the alterations of micaceous minerals that take place in podzol E horizons as a result of eluviation, Righi et al. (1999), Gillot et al. (2001), Mirabella and Egli (2003), and also Egli et al. (2004) have all found that the smectites that form in this way can be quite complex. They constitute a mixture of several populations with various layer charges (Gillot et al., 2001). A chronosequence of soils in Finland enabled these last authors to find that the products of mica transformation included both a relatively high-charge ferromagnesian smectite and also a vermiculite with higher charge in younger soils. They thought that the Fe–Mg smectite arose from the transformation of the ferromagnesian primary mineral biotite and from Fe–Mg-bearing chlorite present in the rocks, whereas the vermiculite, which was aluminous, was a transformation

product of dioctahedral micas (muscovite and phengite). In older soils, smectites with a lower charge resulted from both the dissolution of the higher charged trioctahedral smectites and also a progressive decrease in the charge of the dioctahedral expanded phases (Gillot et al., 2001). Righi et al. (1999) found the nature of the aluminosilicate products of podzolization in the eluvial A and E horizons differed from those in the B horizons within a different chronosequence of soils. The A and E (eluvial) horizons contained a mixed-layer mica–smectite, and, given time, also discrete smectite from the transformation of dioctahedral mica. The B (illuvial) horizons, by contrast, contained mixed-layered mica–vermiculite and smectite. These last clays were the products of the alteration by transformation of trioctahedral minerals, notably biotite and chlorite (Righi et al., 1999). These authors explained the differences between the transformational products by a difference in the nature of the acids involved as weathering agents in the two different parts (or compartments) of the soil profile. In the upper, eluvial compartment, organic acids are responsible for the mineralogical changes and these are more aggressive than the carbonic acid in the lower compartment, which includes the illuvial horizons (Righi et al., 1999). It is only in the upper compartment that dioctahedral minerals become altered, and the organic acids responsible for the alteration there dissolve the products of weathering of the trioctahedral minerals. Their dissolution leads to phases that are translocated to the illuvial Bh and Bs horizons where they accumulate as "amorphous" (i.e., probably nanocrystalline) Fe and Al oxides and as recrystallized gibbsite in these particular soils (Righi et al., 1999). Studying an altitudinal sequence of five podzolized soils in Italian alpine regions, Mirabella and Egli (2003) also found that the smectites formed in either eluvial or illuvial horizons were highly heterogeneous, mostly comprising mixtures of montmorillonite with interstratified beidellite–montmorillonite, but never consisting of beidellite alone. Some were interlayered with hydroxy-Al species, which needed to be removed by citrate treatment before the underlying phases could be distinguished, but their removal showed that the specific expandable micaceous phases with their variety of layer charges had been formed prior to interlayering (Mirabella and Egli, 2003). The most intense weathering, leading to the lowest charged smectites, occurred in podzols below the tree line but closest to it. Clearly, in many of these studies, certain trees are a major factor in the intensification of the podzolization process, if not in its occurrence, in some environments. Churchman (1980) also found that, although podzolization could occur under grasslands provided the rainfall was sufficiently high, the mineralogical changes to micaceous phases were more advanced in soils that were just a few meters away, but which were below the tree line. In the podzols under grassland on both mica–chlorite schist and graywacke lithologies, regularly interstratified mica–beidellite dominated the clay fractions and there was no discrete beidellite present, but discrete beidellite occurred in substantial amounts in podzols under native *Nothofagus* beech forest in New Zealand (Churchman, 1980).

Although beidellitic smectites appear to constitute a stable phase in podzols formed in boreal zones, there can be a sequence

of products formed at different stages during the podzolization process, as discussed by, for example, Churchman (2000), and also by Righi et al. (1999). Studying seven podzol profiles in the Tatra Mountains in Poland, Skiba (2007) has concentrated on the alteration processes of dioctahedral micas within all parts of these soils. This author found a variety of transformational products of dioctahedral micas that varied in the regularity of ordering of their interstratifications with (at first) vermiculites, in the expandability of the vermiculites formed, both as interstratified layers and also as discrete phases, and in the degree of interlayering of vermiculites with smectite, leading ultimately to a discrete smectite phase. Many of these also exhibited interlayering with hydroxy-Al species when pH was ≥4.4. Skiba (2007) also claimed that kaolinite formed neogenetically in these podzols from Al and Si released by weathering.

Three podzols forming in tropical and subtropical climatic zones have been studied in Taiwan by Lin et al. (2002), among others. There was a very similar mineralogical pattern to those seen in podzols in the cooler boreal zone. Illite was altered to vermiculite, or sometimes to interstratified illite–vermiculite in eluvial horizons. There was no interlayering with hydroxy-Al species in the eluvial horizons, which exhibited very low pHs under forest but some occurred in B horizons at higher pHs, and Lin et al. (2002) also considered that translocated Fe, Al, and organic acids coated the Al-interlayered vermiculitic phases with organometallic compounds and thus stabilized them against breakdown. As in many other podzols, the system was characterized by the eluviation, translocation, and deposition of mainly Al and Fe by organic acids.

The general mineralogical pattern in podzols is summarized in Box 20.4. It is given for podzols on micaceous parent materials, which are common for many podzolized soils, although podzols can form on many parent materials or lithologies.

20.2.5.3.3 *Apparent Counteraction of Weathering by Plants*

Some work has indicated that plants, as well as acting as potential weathering agents, may also act to counter the course of mineral development that might otherwise be expected from the operation of the other main weathering agents such as climate. Thus, in a chronosequence on basalt in the tropics, both the early formation of an illite and also a lower content of gibbsite in A horizons than in C horizons across much of the sequence were attributed by He et al. (2008) to the transport by plants of nutrients (K, and also Si, as siliceous phytoliths) to the surface of soils. Together the increases in these nutrients have led to the formation and retention of illite and 2:1 aluminosilicates more generally, at least until they are finally weathered out in the oldest soils. Both April and Keller (1990) and also Calvaruso et al. (2009) found that there was a buildup of the discrete mica phase (respectively muscovite and illite in the two reports) relative to its degraded counterpart in the rhizospheres of trees in comparison with the bulk soils. Barré et al. (2009) have concluded from a review of the literature that this effect of plants retarding weathering changes, which is known as "nutrient uplift," "element translocation," or "biological pumping," is widespread, and that "the clay mineralogy observed in surface soils is probably largely plant mediated." The uptake and biocycling of Si by plants has also been studied by Drees et al. (1989), Lucas et al. (1993), Derry et al. (2005), Farmer (2005), and Henriet et al. (2008). The concentration of Si by "nutrient uplift" has probably contributed, at least, to the preservation of kaolinite and halloysite, respectively, in the surfaces of forest soils in Brazil (Furian et al., 2002) and Costa Rica (Kleber et al., 2007), while strong tropical weathering has led to gibbsite dominating the mineralogy of the saprolite in the Brazilian case and the subsoil in the Costa Rican one. An examination of published data on Si concentrations in Bs horizons in several European forest soils, as well as experimental data for some of these, led Farmer et al. (2005) to postulate that plant phytoliths provide a sink, hence short-term nutrient uplift, for Si from soil solutions during the growing season that is released back to solutions during winter and spring. Such a viewpoint is supported by data derived from studies of plant-related movement in soils of Al and Fe as well as that of Si in both temperate and tropical soils that were reviewed by Lucas (2001), although differences are evident for different environments (see Figure 6 in Lucas, 2001). Nonetheless, Lucas (2001) concluded

---

Box 20.4   Mineralogical Pattern in Podzols Developed on Micaceous Parent Materials

	Horizons	Type	Alteration	Phyllosilicates	Nanocrystalline (SRO) Minerals
Incr. Al, Fe, pH ↓	A, E	Eluvial	Accelerated loss of cations including Al, Fe, K, Si	Beidellitic or smectite or vermiculite; interstratified with illite, or discrete	—
	Bs, Bh	Illuvial	Normal extent of alteration, for example, for climate	Vermiculitic, *not* smectitic, i/strat. or discrete, poss. Al-interlayered	Allophane and/or imogolite, Fe oxides, especially ferrihydrite

that in most places (tropical as well as temperate), the weathering mantle can be regarded as being in a "dynamic equilibrium sustained by plants." Derry et al. (2005) found that most of the silica released to Hawaiian stream water had passed through the biogenic silica pool rather than originating directly from mineral–water reactions in soils in this basaltic environment.

20.2.5.3.4 Humans and Other Fauna as Weathering Agents

The human being is one biological agent with a potential to affect soil mineralogy, which cannot always be ignored. Apart from such effects on soil mineralogy as those from enhanced erosion—leading to, for example, truncation of profiles and transportation of soil minerals over distances from a few meters to the diameter of an entire hemisphere (e.g., Syers et al., 1969) by water or wind—people's activities can also either enhance or retard the influence of other organisms on mineral weathering, whether through the effects of agriculture, pollution, construction, and engineering activities and/or effects on climate and hydrology at either local or global scales (e.g., see Yaron et al., 2010). In the particular case of acid sulfate soils, which are increasing in area globally because of human activities, and which are of increasing concern as a form of both soil and environmental degradation (Dent and Pons, 1995; Ritsema et al., 2000), large-scale changes wrought by humans have led to mineralogical changes, which contribute to the properties of these soils, which have been labeled "the nastiest soils in the world" by Dent and Pons (1995). Classed often as Sulfaquents, acid sulfate soils, which are most commonly coastal, but may also occur in nontidal seepage and marsh areas affected by dryland salinity (Fitzpatrick et al., 1996), originate from the oxidation of sulfides, most commonly of Fe and most often pyrite FeS_2, but sometimes also Fe monosulfides such as mackinawite (tetragonal FeS), pyrrhotite (hexagonal FeS), and greigite (Fe_3S_4) (Burton et al., 2009). These Fe sulfides are themselves products of earlier bacteria-mediated reduction of sulfate ions (Dent and Pons, 1995; Fitzpatrick et al., 1996). Sulfuric acid results from the oxidation of the sulfides (such as occurs after draining the soils), which occurs only slowly in an abiotic environment, but which is greatly accelerated in the presence of Fe-oxidizing bacteria (Ritsema et al., 2000). The sulfuric acid thus produced reacts with soil minerals generally to give a range of compounds. Of these, jarosite ($KFe_3(OH)_6(SO_4)$), also known as "cat clay" because of its yellow, fecal-like color, is most widely recognized as a component of acid sulfate soils (Dent and Pons, 1995; Fitzpatrick et al., 1996), and schwertmannite ($Fe_8O_8(OH)_{8-2x}(SO_4)_x$), has been recognized in some under very acid conditions (Fitzpatrick et al., 1996; Burton et al., 2006, 2008). The Fe oxidic phases, ferrihydrite $Fe_5HO_8 \cdot 4H_2O$ and goethite, can also form (op. cit.), but many other minerals have been found to form as a result of the processes contributing to the evolution of these soils. These minerals include hydrated sulfates of Na and Fe (sideronatrile and natrojarosite), Na and Al (tamarugite), Mg and Fe (copiapite), and Pb and Fe (plumbojarosite), as well as gypsum, halite, the Fe oxyhydroxide, akaganéite, and a poorly ordered Al oxyhydroxide (R.W. Fitzpatrick and M.S. Skwarnecki, unpublished results), as well as elemental sulfur (Burton et al., 2009). In the case of acid

sulfate soils, the effects of human intervention in certain environments have been to promote mineralogical changes that are mediated by microbes.

Agriculture brings about disequilibrium in soils (Velde and Meunier, 2008). The effect of cropping on clay mineralogy was assessed through analyses of samples taken from the long-term Morrow experimental plots at the University of Illinois for >80 years cropping of different types and with different management regimes (Velde and Peck, 2002). Continuous cropping with corn has led to a marked loss of illite in favor of illite–smectite phases. In contrast, rotation cropping over the same period has had little effect on clay mineralogy. However, 80 years of growing rice in flood-irrigated soils in China showed strong effects on clay minerals in the soils (Li et al., 2003). Micaceous phases, including interstratified phases, have largely been lost, as have Fe oxides, but a Si-poor ferromagnesian chlorite has formed. Pastures can also influence clay mineralogical development. Poldered sediments in France with up to ~850 years of development under prairie enabled Velde et al. (2003) to show that the dominantly illite–smectite clay minerals in topsoils, which were influenced the most by grass growth, hardly changed with time from those in the sediment. On the other hand, the same type of clay minerals in the subsoils became more smectitic over the same time period. The subsoils had a higher pH than the topsoils, which were more-or-less neutral in pH. The grasses appeared to stabilize the particular mixed-layer composition of the original clay minerals.

The common agricultural practice of adding fertilizers to soils can also have effects on the mineralogy of soils. Thus, Simonsson et al. (2009) found that soils subjected to the long-term use of K fertilizers had contents of illite, both as a discrete phase and also within mixed layers, which reflected the rate of addition of the fertilizers—illite was more concentrated where fertilizer additions had been higher. K release from the soil minerals appeared to be reversible except where hydroxyaluminum interlayers occurred, presumably blocking K^+ from entering or leaving the interlayers (Simonsson et al., 2009). Velde and Peck (2002) also found that additions of NPK fertilizer after >40 years of continuous cropping with corn restored the clay minerals to their original state whereas, as already noted, illite was sufficiently depleted of K over >80 years of this treatment without fertilizer additions to register a significant change to illite–smectite (I–S). These authors suggested that I–S plays the role of a K buffer that releases K when plants cannot access other sources of this essential nutrient and stores it when there is an excess of K. Pernes-Debuyser et al. (2003) found that the content of nonexpandable illite increased at the expense of expandable I–S in long-term K-addition treatments of soil without plants with either KCl or manure at Versailles in France. Fertilization with ammonium-based N-fertilizers has also been shown to affect clay mineralogy through an increase in the contents of nanocrystalline aluminosilicates and also hydroxyl-Al interlayers in 2:1 minerals (McGahan et al., 2003). On the contrary, Tye et al. (2009) noticed that the greatest mineralogical changes in the long-term plots at Rothamsted in the United Kingdom occurred when fertilization with ammonium sulfate led to pHs falling below 3.7, when

hydroxyl-Al interlayers in 2:1 minerals became solubilized, allowing entry of the interlayers to K^+ and NH_4^+.

Fauna can also aid weathering processes by moving soil materials, especially by bringing less-weathered or unweathered subsurface materials to the soil and land surface as occurs by termites for example in building mounds, and worm ingestion and excretion (e.g., Paton et al., 1995; Johnson, 2002; Meysman et al., 2006). In maritime Antarctica, which receives more liquid water than the Antarctic continent, ornithogenic soils formed on basaltic and andesite rocks in abandoned penguin rookeries and which are organic-rich due to penguin guano show considerable development of secondary minerals, mainly as noncrystalline or nanocrystalline phases, including 1:1 (Al:Si) allophane (Simas et al., 2006).

20.2.5.3.5 Microorganisms as Weathering Agents

It is well recognized that microbes act as reducing agents and thereby contribute to the reassignment of Fe and Mn from primary minerals or secondary oxides, oxyhydroxides, and hydroxides into solution and ultimately into new oxidic phases upon drying, sometimes with the solubilization and release of trace metals such as Co and Ni (e.g., Quantin et al., 2001). Microbes also act commonly as oxidizing agents (such as in the formation of ferrihydrite in seepages). Their activity, when simulated by a series of oxidation-reduction cycles, transformed the short-range-ordered iron oxide "nanogoethite" to microcrystalline goethite and microcrystalline hematite (Thompson et al., 2006). They may also bring about mineral alteration through root exudates, as in the tunneling action of fungal hyphae on feldspars and other mineral grains (van Breemen et al., 2000b; Smits et al., 2005, see Section 20.2.4.2). Müller (2009) has shown that the weathering effect of bacteria can vary greatly between different genetic derivatives of a single strain. Organic acids resulting from the life and death of organisms have also been recognized as important agents in mineral alteration (e.g., Huang and Schnitzer, 1986; Robert and Chenu, 1992). In particular, dicarboxylic acids have been identified in soils (Hue et al., 1986) where they greatly enhance rates of dissolution (Amrhein and Suarez, 1988) through ring compounds as ligands for sequestering metals on mineral surfaces (Casey, 1995). One of the main effects of organic acids on weathering may come through their supply of protons (Ugolini and Sletten, 1991; Nesbitt, 1997). It has also been shown in nature and in experiments that both aluminosilicate minerals and oxides of Fe and Mn can form by nucleation on bacteria (Urrutia and Beveridge, 1994; Ferris, 1997; Tazaki, 1997, 2005; Chan et al., 2009). Magnetite formed partly by bacteria in soils has been described, for example, by Maher and Taylor (1988) and Geiss and Zanner (2006). Observations made by Ueshima and Tazaki (2001) in both deep-sea sediments and laboratory experiments showed that aluminosilicates (nontronite in this case) could be formed on extracellular polysaccharides from microbes. The role of microorganisms in the weathering of glass or volcanic materials has been described, amongst others, by Thorseth et al. (1995), Kawano and Tomita (2001, 2002), and Stroncik and Schmincke (2002). According to the last authors,

bacteria and microorganisms create a local microenvironment as a result of the fluids of their metabolic products. These fluids are either acidic or basic in pH, depending on the type of bacteria: an acidic pH essentially results in incongruent glass dissolution, whereas a basic pH results in congruent glass dissolution, leaving large pits on the glass surface (Stroncik and Schmincke, 2002). Microbial activity thus enhances the rate of dissolution of volcanic glass and microbial alteration also results in the formation of authigenic phases and is accompanied by redistribution of elements.

Hence, microorganisms may play a number of roles in mineral weathering. As summarized by Barker et al. (1997), these include the following: (1) physical disintegration of rocks; (2) production of acids and chelating ligands to accelerate chemical alteration; (3) stabilization of soils to increase their time of exposure to chemical agents; (4) production of extracellular polymers to moderate water potential, maintain diffusion channels, and themselves act as ligands for chelation and also as nuclei for mineral neoformation; and (5) absorption of nutrients to lower solution concentrations and increase the chemical potential for weathering. All these roles may also be performed by plants, including grasses, crops, and trees (e.g., Courchesne, 2006). In addition, they may be carried out by other biological agents, including lichens (Adamo and Violante, 2000; Chen et al., 2000), fungus-growing termites (Jouquet et al., 2002), and mycorrhizal fungae (Jongmans et al., 1997; van Breemen et al., 2000a, 2000b; Certini et al., 2003). With regard to mycorrhizal fungae, the cited authors concluded that the fungal hyphae dissolved and extracted nutrients directly from mineral grains, such as feldspar and hornblende, through their intrusion into fractures and fissures in the geological deposits and by "drilling" directly into minerals, thereby bypassing the external solution phase. However, in studying the microbiological colonization of primary mineral weathering in volcanic soils in Italy, Wilson et al. (2008) found in this case that microbial activity, particularly fungal activity, was not an effective agent of mineral weathering, that the association with clay minerals was indirect, and that fungal weathering of primary minerals was probably less important than previously claimed as a source of plant nutrients. These authors concluded that preformed cleavages and microporous features in primary minerals enable the penetration of fungi for them to bring about mineral alterations and noted that these features were generally absent in the soils they studied.

20.2.5.3.6 The Ecology of Biota and Soil Minerals

Bennett et al. (2001), in noting that microbes bring about the alteration of minerals through their selective extraction of essential nutrients from the minerals, posed the question of whether the suite of microorganisms is partly, at least, controlled by the mineralogy and by the ability of particular microorganisms to use the nutrients contained within the minerals present. Soil minerals, both primary and secondary, may thus be viewed in the ecological context that is characterized by Young and Crawford (2004) as "the soil-microbe complex." Building on the concept of plants as ecosystem engineers, and on published information on

the effects of particular plant species on soils, van Breemen and Finzi (1998) reviewed the evidence that such effects can provide a positive feedback to such plants. They concluded that there was strong evidence for plant-soil feedbacks in a variety of ecosystems, and they argued that these feedbacks could have played a role in the evolution of the plant species in question. Earlier, van Breeman (1992) had asked the question: "have soils merely been influenced by biota or have biota created soils as natural bodies with properties favorable for terrestrial life?" Coleman and Crossley (2003) argued for the latter by reference to the storage in soils of the life-limiting element, phosphorus. They contended that (1) the interaction of biota with primary minerals produces soluble Fe (and other cations) by weathering; (2) Fe in particular is brought into the solid phase as oxides (including oxyhydroxides) by oxidation aided by (iron oxidizing) bacteria; and (3) the Fe oxides are kept in a largely amorphous or nanocrystalline form, from which P can be extracted by plants more efficiently than from crystalline oxides, through their interaction with (biota-derived) humic substances. Recent experimental

work has shown both that (1) bacteria can effect the production of polymeric molecules within the structures of (smectitic) clay minerals, hence probably improving the environment for the survival of the bacteria (Alimova et al., 2009), and (2) the structure of bacterial communities in soil is influenced by the (primary) minerals in their microhabitat (Carson et al., 2009). Chickens and eggs come to mind in assessing whether biota or minerals are the limiting factors in the ecology of the soil, but it is clear that, while both are important, it is their interactions that are of overriding and vital importance to the sustainable functioning of soil and the life it supports.

20.2.6 Processes of Mineral Alteration and Formation by Weathering: Summary

Secondary minerals are formed as a result of the alteration of primary minerals or mineraloids by either (1) transformation in the solid state or (2) hydrolysis, dissolution, and recrystallization out of solution, namely, neogenesis. Figure 20.9 summarizes these two possible mechanisms.

20.3 Occurrence of Clay Minerals in Soils

This section concentrates on the particular forms in which clay minerals occur within soils, rather than the general groups of soils in which they occur, or even their distribution in particular environments or globally. These aspects are dealt with elsewhere in this handbook. Instead, our focus is on the sizes and shapes of mineral particles and peculiarities of their chemical compositions and crystal structures, largely in relation to those of "type" occurrences of minerals with the same names in more or less pure "geological" deposits, including those often used as standard specimens of each type. In this section, we also explore the associations of minerals within soils in relation both to other types of minerals and also to organic materials. In particular, we examine the possibility that there is a relationship between the mode of formation (and alteration) of clay minerals, on the one hand, and on the other, the forms they adopt and also their associations within soils.

20.3.1 Occurrence of Kaolinite in Soils

Kaolinite is the most ubiquitous phyllosilicate in soils (White and Dixon, 2002). Nonetheless, kaolinites are most abundant in highly weathered soils, which are formed in warm humid climates, although a high rainfall can result in their formation within soils in temperate climates (Weaver, 1989; White and Dixon, 2002). They can also be found in soils in colder climates that were nevertheless formed under an earlier, warmer climate, for example, by preglacial weathering in Scotland (Wilson et al., 1984). Generally, kaolinites in soils are highly disordered and many of the peaks in their XRD patterns are poorly resolved. As a result, the common index for assessing the degree of crystalline order by XRD, namely, the Hinckley index, is impossible
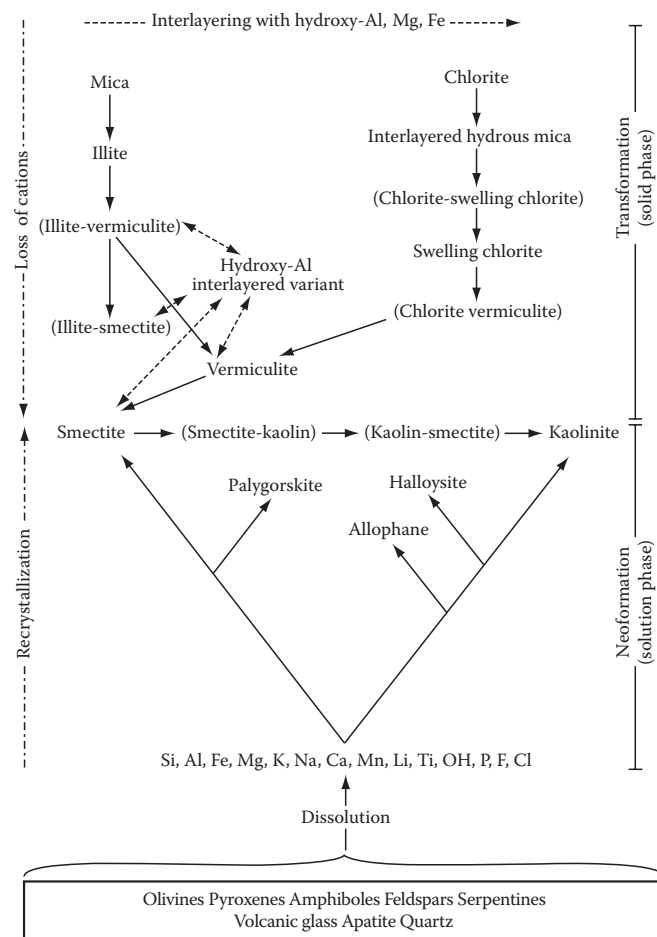
FIGURE 20.9 Main pathways for mineral alteration and formation in soils via either neoformation (lower part of figure) or transformation (upper part). (After Churchman, G.J. 1978. Studies on a climosequence of soils in tussock grasslands. 21. Mineralogy. N.Z. J. Sci. 21:467–480; Wilson, M.J. 2004. Weathering of the primary rock-forming minerals: Processes, products and rates. Clay Miner. 39:233–266.)

to measure and an alternative, empirical index was devised by Hughes and Brown (1979), which remains in use for soil kaolins (including halloysite). In developing this index, Hughes and Brown (1979) noted that kaolinites in the set of 26 soils they mainly studied, which were from Nigeria, invariably occurred along with other minerals in their clay fractions, as well as containing weatherable minerals in the coarser fractions. These authors suggested that the weatherable minerals present hindered the crystallization of well-ordered kaolins. Millot (1970) had come to a similar conclusion, suggesting more specifically that alkaline-earth cations had caused the inhibition of crystallization. Iron is a common impurity in the kaolinite structure, and there appears to be a relationship between structural Fe and crystalline order (Herbillon et al., 1976; Mestdagh et al., 1980). As Fe increases in amount, displacing the smaller Al atoms from the structure, crystal strain may increase, thereby limiting crystal growth, so that particle size as well as structural order is likely to be diminished by a high Fe content. Lim et al. (1980) had found that nonsoil kaolinites could also be poorly crystallized. Compared with their well-crystallized counterparts, seven examples of these poorly ordered kaolinites, from Georgia in the United States of America, showed not only much higher values for specific surface area particularly (between 78 and 114 m^2 g^{-1}, cf. 10–15 m^2 g^{-1} for the well-crystallized kaolinites), but also some relatively high values for CEC. All were shown to comprise some montmorillonite and vermiculite as impurities, while six of them also contained mica. Kaolinites often occur in vermiform books, suggesting a micaceous precursor. White and Dixon (2002) suggested, rather, that micas, or, more commonly, their expanded transformation products with the same crystal

form, for example, vermiculite or beidellitic smectite, provide nucleation sites on their 001 surfaces for authigenic kaolinite formation, as shown in Figure 20.4.

A number of papers (Singh and Gilkes, 1992a; Melo et al., 2001; Hart et al., 2002; Kanket et al., 2005) have compiled data describing crystal form, measures of crystal order, and related properties such as Fe contents for kaolinites extracted from highly weathered soils in (Western) Australia, Brazil, Thailand, and Indonesia, although the last also included some halloysite or "tubular kaolins" (Churchman and Gilkes, 1989), so the total set is better described as kaolins. Together with information from Hughes and Brown (1979) for Nigerian soils, these properties are able to be compared in Table 20.4 (from Kanket et al., 2005, with data for the reference kaolinites from Hart et al., 2002). The various kaolins had many properties in common. Although there was a range for each value within each set of soil kaolins, their specific surface areas (overall range 34–88 m^2 g^{-1}) all exceeded those of any of the reference samples (all ≤25 m^2 g^{-1}) and their particle sizes, given by coherent scattering domains (CSDs), were generally less than half those of the reference kaolinites. Surface areas and particle sizes are expected to be inversely related to each other. Their Hughes and Brown (HB) index values were always less than those of the reference kaolinites, and, although some of the soil kaolins had quite high CECs, some were comparable to those of several, at least, of the reference kaolinites. Regardless of their other characteristics, however, all of the soil kaolins on which Fe was measured (those from Thailand, Western Australia, and Indonesia) had much higher contents of Fe than the reference kaolinites. With the most clear-cut differences between soils and reference kaolinites being in their surface areas, the HB measure

TABLE 20.4   Properties of Kaolins in Groups of Soils (with Standard Deviations, Where Given, in Brackets) and of Standard Kaolinites

Location	No. of Samples	d_{001} (Å)	CSD[a] (nm)	HB[b] Index	% Fe_2O_3 (g kg^{-1})	CEC (cmol$_+$ kg^{-1})	SSA[c] (g kg^{-1})
Soils							
Thailand (1)[d]	20	7.208 (0.042)	19.3 (8.9)	6.7 (1.6)	19.6 (1.8)	8.5 (2.9)	47.4 (8.8)
Thailand (2)[e]	18	7.19 (0.03)	15.4 (5.5)	6 (1)	25.3	15.2 (4.7)	44.9 (12.1)
Western Australia	7	7.213	22.9	5.6	25.7	5.0	50.8
Indonesia	6	7.212	10.8	5.6	25.4	9.4	72.8
Nigeria	26	—	—	8.0	—	—	—
Brazil	21	7.19	—	12.7	19.6	—	48.5
Standards							
Georgia #1460	1	7.148	26.6	18.9	8.9	3.2	25
Georgia #1261	1	7.160	43.4	32.6	2.5	3.9	10
Georgia #MP5	1	7.132	47.6	34.2	4.2	0.4	5
New Mexico (Wards)	1	7.148	43.1	60.5	4.3	4.6	13
Cornwall (ECC)	1	7.160	48.8	26.3	4.2	0.5	11

Source: Data for the soil kaolins from Kanket, W., A. Suddhiprakern, I. Kheoruenromne, and R.J. Gilkes. 2005. Chemical and crystallographic properties of kaolin from Ultisols in Thailand. Clays Clay Miner. 53:478–489, with data for the reference kaolinites from Hart, R.D., R.J. Gilkes, S. Siradz, and B. Singh. 2002. The nature of soil kaolins from Indonesia and Western Australia. Clays Clay Miner. 50:198–207.

[a] CSD, 001 peak.

[b] Hughes and Brown (1979).

[c] Specific surface area from N_2-BET.

[d] From Hart et al. (2003).

[e] From Kanket et al. (2005);

– Not determined.
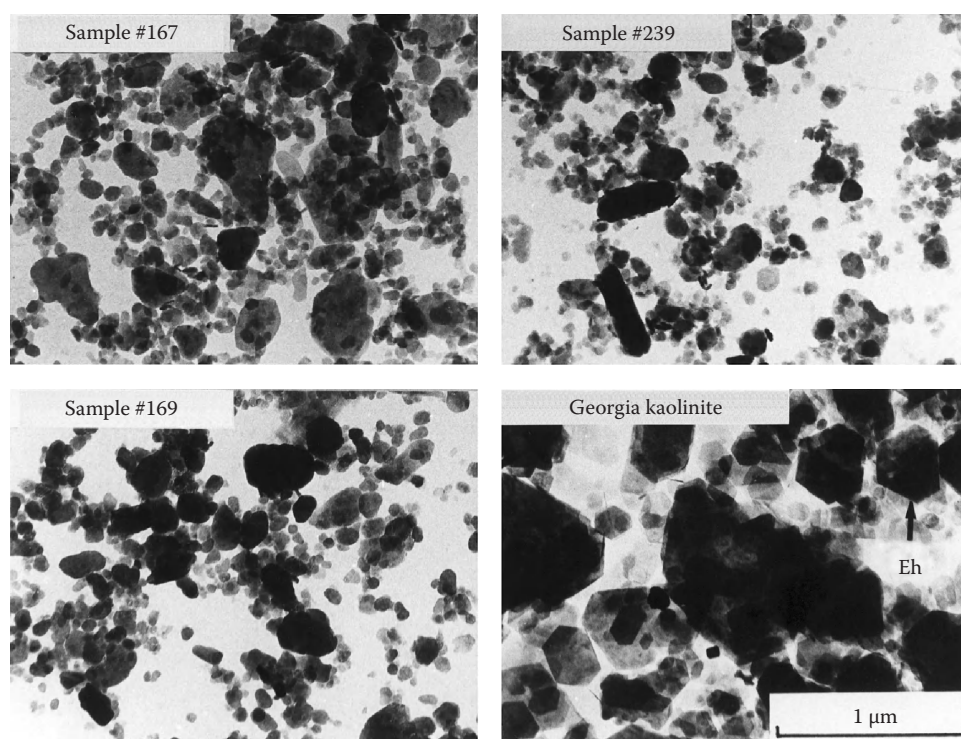
FIGURE 20.10 Transmission electron micrographs showing comparisons between some soil kaolinites from Western Australia and a reference kaolinite (from Georgia). Eh indicates a euhedral hexagonal shape for particles in the reference sample. (Micrographs copied from Singh, B., and R.J. Gilkes. 1992a. Properties of soil kaolinites from south-western Australia. J. Soil Sci. 43:645–667. Reproduced with permission of Wiley-Blackwell.)

of structural order, and their Fe contents, these results bear out the proposition from Herbillon et al. (1976) and Mestdagh et al. (1980) that a high Fe content leads to a low degree of structural order and small particles. Their small size, particularly, is illustrated by transmission electron micrographs showing comparisons between some soil kaolinites from Western Australia from Singh and Gilkes (1992a) and a reference kaolinite (from Georgia) (Figure 20.10). Furthermore, Singh and Gilkes (1992a) found that the kaolinites contained both K and Mg, which are largely absent from pure, well-crystallized kaolinites. Brazilian soil kaolinites were also found to contain considerable K and Mg (Melo et al., 2001). These authors suggested that there are residual micaceous layers incorporated in the kaolinites. This finding is consistent with the results of Lim et al. (1980) and also with the common observation that kaolinites appear in vermicular books by crystallization on the surfaces of micas and their transformation products (Figure 20.4). The crystal growth of the kaolinites may have been constrained by interleaved micaceous layers as well as by Fe incorporated into the kaolinite layers.

Kaolinite (and also, we shall see, halloysite) quite often occur in interstratifications with a smectite. Kaolinite–smectite interstratified phases have been found to occur worldwide, including in Scotland, Canada, throughout Africa, Australia, Central America, India, Japan, China, and the United States (see Churchman, 2000, for citations), and a more recent report (Vingiani et al., 2004) described their occurrence in Italy (Sardinia). Norrish and Pickering (1983) speculated that

kaolin–smectite interstratified phases (or "kaolin–smectites") are likely to be a very common component of Australian soils. In Australia, and elsewhere, they have probably been often overlooked in identifications using XRD patterns. This "omission" is because the main effect of the mixed layering is to broaden and weaken XRD peaks, so that basal peaks may become invisible even though nonbasal peaks appear (Norrish and Pickering, 1983). In part, this effect is a consequence of the inherent variability of clay minerals in soils, so that the properties of the two (and sometimes more) components and their degree of order can vary from particle to particle (ibid.). Most kaolinite–smectites described in the literature are more kaolinitic than smectitic (Churchman et al., 1994), although Bühmann and Grubb (1991), Churchman et al. (1994), and also Ryan and Huertas (2009) each found examples where the proportion of smectite layers was >60%. In the same way that many interstratified specimens with high concentrations of kaolinite have been identified incorrectly as kaolinites (Norrish and Pickering, 1983; Churchman et al., 1994), so also some with high contents of smectite have been misidentified as smectites (Cuadros et al., 1994). Some with a high kaolinite content were probably misidentified as halloysites that have a platy morphology (Carson and Kunze, 1970; Wada et al., 1987), but each of these was later interpreted as a kaolin–smectite by Sakharov and Drits (1973) and Parfitt and Churchman (1988), respectively. The identification of phases as kaolin–smectites rather than as either of their components is important in order to fully describe the effect of clay mineralogy on soil properties

such as CEC (e.g., Seybold et al., 2005) because they give rise to higher values than can be expected from kaolinites and lower values than from smectites.

Kaolinite–smectites have been found in intermediate zones in toposequences spanning smectitic black soils at their base and kaolinitic red soils near their summits and on their steep slopes (Herbillon et al., 1981; Bühmann and Grubb, 1991; Delvaux and Herbillon, 1995; Vingiani et al., 2004). They have also been recognized within single soil profiles on basalt (Churchman et al., 1994; Vingiani et al., 2004). From the base of the profile to near its surface, their composition changed from highly smectitic to highly kaolinitic, but nonetheless with both phases being interstratified kaolinite–smectites (Churchman et al., 1994), or else more subtly, from ~35% to 40% to ~40% to 45% kaolinitic within the interstratifications (Vingiani et al., 2004). Ryan and Huertas (2009) have also located kaolinite–smectites in a set of soils within a chronosequence on mostly basaltic and andesitic parent materials. As it was within the residual soil profiles on rock (Churchman et al., 1994; Vingiani et al., 2004), the progression from youngest (found at the base of the profiles) to the oldest (at the surface in the case of the profiles) was from more smectitic (or pure smectite, in the chronosequence) to more kaolinitic. Kaolinite also occurred as a discrete phase alongside its interstratifications with smectite in the most developed (i.e., oldest) soil in the chronosequence of Ryan and Huertas (2009) and also in the soil profiles of Churchman et al. (1994) and Vingiani et al. (2004). A variety of techniques, including XRD, together with modeling of the effect of interstratifications and sometimes also peak decomposition, FTIR, HRTEM, differential thermal analysis (DTA), and permanent and variable charge analysis, have enabled various authors to seek an understanding of the origin of the process of interstratification of the 2:1 and 1:1 layers. Ryan and Huertas (2009) concluded that a layer-by-layer transformation occurs as smectite is progressively altered to kaolinite via an interstratification of the two endmember phases with each other. The transformation proceeds by localized dissolution of the 2:1 layers with initial loss of Fe and Mg from octahedral sheets and of Al from tetrahedral sheets followed by a loss of Si from tetrahedral sheets ("subtraction") and/or a deposition of hydroxyl-Al species in interlayers, leaving 1:1 layers ("addition") (Ryan and Huertas, 2009). These two alternative types of mechanism for the alteration had been proposed earlier by Wada and Kakuto (1983, 1989), who favored a subtraction mechanism, and by Altschuler et al. (1963) and Poncelot and Brindley (1967), who favored an addition mechanism. Albeit through studying a clay deposit rather than a soil, Watanabe et al. (1992) found support for a subtraction mechanism as a result of their discovery that opal C–T, or microcrystalline opal comprising clusters of stacking of cristobalite and tridymite over very short length scales, forms during the early stages of formation of kaolinite from smectite. Vingiani et al. (2004) discovered that the smectite in the sequence they studied consisted of a mixture of a more ferric and a more aluminous type. In the course of mineralogical evolution, there was a decrease in the proportion of the ferrous smectite layers relative to their aluminous counterparts. These

authors proposed that the (subtraction) mechanism operated through a preferential loss, by either transformation or dissolution of the ferrous smectite layers, to give kaolinite layers, with the excess Fe forming either Fe oxides (or oxyhydroxides) or becoming incorporated in kaolinite layers, or both.

Kaolinites can also form interstratifications with other types of layers besides smectites, for example, vermiculite (Jaynes et al., 1989), and can also participate in interstratified phases involving several different types of layers. The elevated CECs and also high K_2O contents of some of the soil kaolins studied by Singh and Gilkes (1992a), Melo et al. (2001), and Kanket et al. (2005), in particular, are at least partly ascribed to their interstratifications by micas and/or vermiculites. Ma and Eggleton (1999) found that some soil kaolinites in Queensland, Australia, had quite high CECs (16–34 $cmol_+\ kg^{-1}$) and HRTEM had shown that their crystals commonly comprised stacks of many kaolinite layers, which terminated in smectite layers, possibly as single layers. These stack-end smectite layers affected CEC but were usually not detectable by XRD (Ma and Eggleton, 1999).

20.3.2 Occurrence of Halloysite in Soils

As already noted (Section 20.2.4.5), halloysite is a common component of soils formed from volcanic parent materials, particularly tephra including finer components referred to as volcanic ash (e.g., Kirkman, 1981; Lowe, 1986). It has also been identified as a weathering product of a wide range of rock types including nephelinite, granite, gneissic granite and granitic gneiss, gneiss, dolorite, schist, graywacke, greenstone, granodiorite, gabbro, shale, and amphibolites (see Churchman, 2000, and also Joussein et al., 2005, for citations). Halloysite can occur in a variety of particle morphologies including tubes and microtubules, spheroids and microspheres, crumpled lamellae, crinkly films, needle-like or fiber-like forms, and prismatic forms (ibid.; see also Hong and Mi, 2006; Etame et al., 2009). Its morphology appears to relate to the content of impurity iron with substitution of the larger Fe(III) for Al in the octahedral sheet, which, without substitutions, is smaller than the tetrahedral sheet. This situation leads to partial or full correction of the misfit between the octahedral and tetrahedral sheets that, therefore, causes curling (as in tubes) in halloysites with little or no Fe. The impurity Fe also constrains crystal growth and leads to smaller particles being formed. Recently, Singer et al. (2004) have found halloysites with high contents (up to 6.5%) of Ti in their aluminosilicate structure that were plate-like, probably from the same cause: larger Ti atoms replace some Al in the octahedral sheets, hence also correcting at least some of the mismatch with the tetrahedral sheet. Even so, whereas a peculiar shape was once regarded as the distinguishing feature of halloysite in comparison with kaolinite, it is rather the occurrence, or else evidence for prior occurrence, of interlayer water that demarcates halloysites from kaolinites (see Section 20.2.4.6.2 earlier herein). Interlayer water is lost irreversibly from halloysites on drying.

However, some shapes, particularly spheroidal, appear to reflect a particular mode of formation for halloysite (Churchman,

2000). Spheroidal forms of halloysite appear often in weathering products of volcanic glass, which has a fast dissolution rate, and recrystallization from the resulting supersaturated solution appears to favor this particular shape (review by Bailey, 1990; Adamo et al., 2001; Singer et al., 2004), possibly when physically constrained, for example, within pumice cavities (Adamo et al., 2001). Nevertheless, both Churchman and Theng (1984) and Singer et al. (2004) found that spheroidal particles had higher Fe contents than even short tubes, suggesting a structural control on shape for spheroids as well as for tubes. By contrast, Johnson et al. (1990) found a spheroidal halloysite with virtually no Fe whereas Noro (1986) and also Adamo et al. (2001) found few, or only minor, differences in composition between spheroidal and tubular forms of the mineral. In any case, Soma et al. (1992), using XPS, found that Fe content varied between different layers in a halloysite, whether spheroidal or tubular. Furthermore, spheroidal- and tubular-shaped particles occur together in halloysites (Saigusa et al., 1978; Churchman and Theng, 1984), and one may alter to another in situ (Sudo and Yatsumoto, 1977), suggesting there may be no clear-cut distinction between the conditions that lead to the different particle shapes for halloysite. On the other hand, a very detailed microtextural and microanalytical study of samples taken at various depths in a weathering profile on gneiss in Greece led Papoulis et al. (2004) to deduce a sequence of mineral alteration that encompasses different morphologies for halloysite and includes two forms of kaolinite while simultaneously indicating a correlation between structural Fe and the various products in the alteration sequence. Papoulis et al. (2004) divided the alteration sequence into four successive stages. In the earliest stage, halloysite formed in very small spheres (~0.1 to 0.3 μm in diameter) that were located on the surfaces of plagioclase in this case. These small spheres then apparently coalesced to form tubes while new tubes also formed in the second stage. These tubes were generally quite short (<1 μm in length), although they could develop with time to form longer tubes with tapered ends, which, together with a relatively loose packing and subparallel orientation, suggest crystallization from solution (a conclusion supported by studies on halloysite with tapered ends by Hong and Mi, 2006). In the third stage, platy halloysite formed apparently by interconnection of the preformed tubes. This form even occurred in crude booklets, which are commonly regarded as more typical of kaolinites, albeit interleaved with a low proportion of kaolinite plates, according to Papoulis et al. (2004). These authors considered that the platy halloysite was unstable and led to the formation of plates of a poorly ordered kaolinite, with dimensions of ~1 μm in length and ~0.1 to 0.2 μm in thickness. The change from halloysite plates to kaolinite was considered to occur within the single (third) stage by Papoulis et al. (2004). The fourth stage was one of kaolinization, with a disordered kaolinite forming first but with a well-formed book-type kaolinite as the ultimate product. Concomitant with these changes in morphology, increasing Fe in EDAX/EDS analyses of the products through the first three stages led these authors to allocate increasing proportions of Fe in octahedral sheets at the expense of Al, so that spheroidal halloysite had less substitution

of Al by Fe than tubular halloysite, whereas the platy form had the most structural Fe among the halloysites. The poorly ordered kaolinite that developed from halloysite in the third stage had a comparable degree of substitution of octahedral Al by Fe as its supposed platy halloysite precursor, and the disordered kaolinite formed first in the fourth stage had even more substitution of Fe for Al. However, the book-type kaolinite had a vanishingly small content of structural Fe. It was almost pure aluminosilicate. Overall, Papoulis et al. (2004) attributed the changes in the sequence from small spheroidal particles of halloysite through tubular and play halloysite to poorly ordered platy, then well-ordered book-type kaolinite, largely to a changing chemistry of the ambient solutions, with a successively decreasing availability of Fe for incorporation in the aluminosilicate layers, although they conceded some role to microenvironmental conditions (e.g., time, and space available) in affecting particle morphology.

Bailey (1990) concluded that the structural difference between halloysite and kaolinite is substantial enough to make it impossible for one to give rise to the other without dissolution and recrystallization. Nonetheless, there have been many studies showing an apparent genetic relationship between halloysite and kaolinite in weathering profiles including soils and deeper regolith. There is a general trend from halloysite at depth toward kaolinite at the surface of profiles on residual rock materials (e.g., Eswaran and Wong, 1978; Calvert et al., 1980; Churchman and Gilkes, 1989; Churchman, 1990; Takahashi et al., 2001, Singer et al., 2004; Jongkind and Buurman, 2006). These studies together have covered a range of parent rock types and also climates, including tropical (Eswaran and Wong, 1978, and part of Churchman, 1990), temperate (Calvert et al., 1980; part of Churchman, 1990; Jongkind and Buurman, 2006), and Mediterranean, or xeric (Churchman and Gilkes, 1989; Takahashi et al., 2001; Singer et al., 2004). According to Papoulis et al. (2004), these trends with depth were explained in the context of their morphological and compositional development sequence of halloysite and kaolinites. They pointed out that a depth sequence in a residual weathering profile on rock is inter alia also a time sequence, with the earliest stages concentrated at depth and the latest (oldest) stages to be found near the surface (unlike we note the situation in upbuilding profiles where the opposite is true). Like Churchman and Gilkes (1989), these authors considered that the products of the different stages could also exist simultaneously with one another. Singer et al.'s (2004) examination of the weathering of basaltic scoria at a fine scale showed that, whereas 1.0 nm halloysite was found in vesicles in the scoria, where it was protected from dehydration, 0.7 nm halloysite and also kaolinite were found in the less scoriaceous basalt, where minerals had been exposed to seasonal drying. 1.0 nm halloysite was also found in a buried paleosol in the same region. Its burial in the paleosol had protected it from drying. In New Zealand, Lowe (1986) found that 0.7 nm halloysite predominated in the surface horizons of a soil developed on weathered tephra subject to reasonably frequent seasonal drought, whereas 1.0 nm halloysite occurred in deeper horizons that were less affected by drying. An intermediate zone of variably dehydrated

halloysite with XRD peaks between 1.0 and 0.7 nm occurred between upper and lower subsoil horizons.

These findings were corroborated by Churchman et al. (2010), who found that alteration of granite and volcanic tuff in Hong Kong had led to halloysite forming only in veins within the rocks where there was no evidence of drying, namely, no Fe or Mn oxides were present. By the same token, this study showed that kaolinite formed in these veins, either alone or in association with halloysite, where the appearance of Fe and Mn oxides indicated that drying had occurred. A weathering sequence on trachy-basaltic parent materials in Sicily was found to produce imogolite-like allophane (i.e., Al-rich allophane) in the early stages of alteration of volcanic glass and kaolinite, but no halloysite in later stages (Egli et al., 2008). This absence is probably because of the xeric moisture regime in the area. Similarly, basaltic parent materials in a xeric moisture regime in South Australia were found to produce allophane, but not halloysite (Lowe et al., 1996; Lowe and Palmer, 2005; Takesako et al., 2010; also Section 20.2.4.6). Kaolinite per se was not formed in this case, and, although kaolinite–smectite was often present in the soils, it could have been derived ultimately from limestones and calcareous dune sands underlying the basalt that were incorporated into the soil parent material during the basalt emplacement, with 1:1 Al:Si allophane the likely result of further development of initially formed 2:1 Al:Si allophane by resilication. Albeit that a drier regime tends to favor halloysite over allophane in soils from volcanic ash (see Section 20.2.4.6.2), these various observations suggest that the presence of water is a necessary condition for the formation and preservation of halloysite rather than kaolinite under similar conditions of, for example, parent material, temperature, and time.

Even so, a trend from halloysite at depth toward kaolinite at the surface in residual soil profiles does not necessarily imply that halloysite transforms to kaolinite as weathering intensifies with time. Churchman and Gilkes (1989) found rather that hydrated, tubular halloysite, which dominated the clay fraction of the saprolite deep in a lateritic profile in dolerite in Western Australia, became progressively more dehydrated and progressively more difficult to intercalate with the polar liquids—formamide, hydrazine, and a concentrated potassium acetate solution—with closer approach to the surface. As weathering intensified up the profile, kaolinite began to appear as hexagonal particles but by a different pathway from that for halloysite. As has been found elsewhere in the old, strongly leached landscapes of much of Australia (Janik and Keeling, 1993), Churchman and Gilkes (1989) observed that tubular particles with a kaolin layer structure and composition appeared even in the uppermost duricrust of the weathering profile, but they were dehydrated and completely resistant to intercalation by polar liquids.

In seeming contradiction to this trend and also to Bailey's (1990) contention that halloysite and kaolinite cannot transform from one to the other in the solid phase, both Robertson and Eggleton (1991) and also Singh and Gilkes (1992b) showed electron micrographs from within weathering profiles in Australia in which tubular particles, suggesting halloysite, are seen to apparently form on the surface of kaolinite platy particles. Dissolution of kaolinite and recrystallization of halloysite could have occurred in both situations.

Like kaolinites, halloysites have been found to form interstratified phases with smectites. However, halloysite–smectites form under conditions different from those that result in kaolinite–smectites (Delvaux and Herbillon, 1995). Delvaux and Herbillon (1995, and earlier papers cited therein) have generally found interstratified halloysite–smectites in soils formed from volcanic ash. They form under similar (usually alkaline) conditions, which give rise to smectites but where the drainage is less restricted (i.e., more leaching occurs) than for smectite formation (Glassmann and Simonson, 1985; Smith et al., 1987).

There do not appear to be any studies of the form of halloysites and their associations in soils that are similarly detailed to those of soil kaolinites by Singh, Gilkes, and coauthors (Section 20.3.1). However, many electron micrographs, including those of the highly halloysitic Naike and Hamilton soils in New Zealand (formed from strongly weathered tephra) that are shown in Fieldes (1968) and Churchman and Theng (1984), respectively, display mainly tubular particles that are practically all <0.2 μm. Lowe (1986) also found tubes, some split, between ~0.07 and ~0.3 μm in length in similar materials. Furthermore, high values for surface areas of halloysites in soils or halloysite-rich soils from around the world have been recorded by several workers (Quantin et al., 1984; Theng, 1995; Takahashi et al., 2001; Singer et al., 2004). Specific surface areas >100 m^2 g^{-1} were measured for soil halloysites or halloysitic soils in all of the studies cited. These area measurements point to the likelihood that halloysites, like kaolinites, commonly occur as quite small particles in soils. The fineness of their particles very likely reflects compositional or structural features that have limited crystal growth. In the case of the high surface area halloysites studied by Singer et al. (2004), it is likely that the high degree of substitution of Al by Fe and/or Ti introduced strains into the layer structure that limited the size of particles. The clay fractions of two halloysitic soils studied by Takahashi et al. (2001) were found to have high CECs and a high selectivity for K$^+$ in addition to high surface areas, and ^{27}Al-NMR showed that their constituent halloysites had ~2% tetrahedral Al and probably also some Al vacancies and impurity Fe in their octahedral sheets. They contained some Fe as oxides, which could have been present as surface coatings, and some 2:1 Si:Al minerals could also have been present in the clay fraction, even if not detected there, because of their obvious appearance in coarser fractions. While these authors could not reach unequivocal conclusions about the extent to which halloysite per se contributed to the unusual properties of these soils, the dilemma they encountered was the result of the complex mineral assemblages that could involve halloysites in soils and emphasizes how halloysites, like kaolinites, can be bound into multiphase associations in soils.

Halloysites often occur in close associations with iron oxides and oxyhydroxides in soils (Churchman and Theng, 1984; Bakker et al., 1996; Takahashi et al., 2001; Pochet et al., 2007; Etame et al., 2009). Pochet et al. (2007) showed that they can

form stable microaggregates with Fe oxides, and especially ferrihydrite that probably preserve the halloysites in spite of nearby soils being at an advanced state of weathering, which would normally drive the change toward kaolinite in a humid tropical environment.

In a study of soils formed from nephelinite (alkali- and rare-earth-rich lava) in tropical Cameroon, Etame et al. (2009) demonstrated a chemical dependence of the halloysites upon parent minerals: Ce-rich halloysite characterized alteromorphs formed from the mineral phillipsite, Fe-rich halloysite characterized alteromorphs on clinopyroxene, Ca-rich halloysite characterized alteromorphs on hauyne, and K-rich halloysite characterized alteromorphs on leucite. A Ce-rich halloysite (Ce_2O 0.1%) was also identified in saprolite derived from hydrothermally altered nephelinite (Etame et al., 2009).

20.3.3 Occurrence of Illite in Soils

Illites can form in soils by weathering as a result of transformation, in the solid phase, from micas (Section 20.2.4.3, and also, via biota, Section 20.2.5.3) and, like most of the other secondary minerals, they can also precipitate by neogenesis out of solution. There is little doubt that they most commonly form by transformation from primary micas. The degree of transformation they have undergone differs greatly, so that, in younger and less-weathered soils, especially those in arid environments (Section 20.2.5.2), the clay-sized micaceous phases may be mainly a product of comminution of primary micas by physical weathering processes (e.g., Fanning et al., 1989). In this case, the secondary phase may be better described as clay-sized mica rather than as illite. However, the term illite is used quite widely to describe mica-like clay-size minerals in soils, as already noted (Section 20.2.4.3). In all but the youngest or little-weathered soils, the micaceous phase in their clay fraction, hereafter referred to as illite, or illitic, contains some expanded layers; hence, some degree of interstratification of the K-rich illite layers with vermiculite or smectite layers that have become depleted in K through its displacement by hydrated cations. Hence, they are expandable.

Micas, especially muscovite and biotite, are particularly abundant in mica schists and some other metamorphic rocks (Fanning et al., 1989). Mica schists are especially abundant because they originate from the metamorphosis of argillaceous sediments such as shales and slates (ibid.). According to Fairbridge and Bourgeois (1978), shales comprise ~40% to 60% of all sedimentary rocks by volume and ~80% of crustal weathering products. As relatively soft minerals, micas in the schists are readily broken down during transport and sedimentation. Thence, they become important components of shales and slates, the types of sedimentary rocks from which they originated. Hence, illite may originate by inheritance from micas in underlying or nearby sedimentary shales or slates that may ultimately be recycled by metamorphosis into new mica schists (e.g., Garrels and Mackenzie, 1971). Illite in soils formed from these widespread sedimentary rocks is most abundant in the coarse clay (0.2–2 μm) size fraction (Fanning et al., 1989).

On the other hand, illite that has undergone considerable transformation to include expansible 2:1 layers of vermiculite and/or smectite is more likely to occur in the fine clay (<0.2 μm) size fraction (Fanning et al., 1989; Laird et al., 1991; Robert et al., 1991). Studying soil clays derived from the major sedimentary rocks, Robert et al. (1991) found that illite–smectite interstratified phases (I–S) are "the most widespread and representative soil clays," a conclusion that was echoed by Velde and Meunier (2008; see also Section 20.2.5.3). Generally, they are randomly interstratified (Thompson and Ukrainczyk, 2002). Robert et al. (1991) distinguished two types of I–S. One, which they termed "micromicas," may occur as large crystals and have layers with 1 nm periodicity that are separated by "swelling spaces" (Robert et al., 1991). It is a "structural" type of interstratification. The other type consists of random superpositions of elementary particles, "like monolayers," to quote Robert et al. (1991). These authors characterized it as a "textural" type of interstratification, which is the same type as that identified by Nadeau et al. (1984) in the context of diagenetic clay minerals in sandstones.

One of the main conclusions to be drawn from Robert et al.'s (1991) work is that I–S, as common soil clays, differ significantly from reference clays. Among their defining characteristics, they are always multiphase, their smectitic phase is dominantly dioctahedral, and Al-rich, hence beidellitic (Robert et al., 1991), their particles have a very small lateral extension (often 30–50 nm) and a very small number of layers. Furthermore, their smectite component includes partially collapsed layers, and the clays have a high external surface area, particularly relative to their internal surface area (Robert et al., 1991). Laird et al. (1991) studied an I–S mineral that fully occupied fine clay fractions of a soil and found that the smectitic layers comprised a high-charge Fe-rich montmorillonite. This finding is at odds with several reports of soil smectites being more beidellitic, albeit Fe-rich (see Section 20.3.5). Laird et al. (1991) also found that the illitic layers had a low, largely tetrahedral charge, and were dioctahedral. Their layer charge was more typical of a beidellite than an illite. These authors speculated that the illitic layers in the I–S could themselves constitute a breakdown product of illite in the coarse clay fraction or else they could have formed by the collapse of pre-formed beidellite layers, but the evidence available to Laird et al. (1991) favored neither mechanism. Examination of a fine clay (0.02–0.06 μm) fraction by HRTEM in a later study (Laird and Nater, 1993) revealed that this was composed of (1) two-layer elementary illite particles, which were most abundant, (2) one-layer elementary smectite particles, and (3) rare discrete multilayer illite particles. The elementary illite particles could be separated from the smectitic component and were shown to comprise a rectorite-like phase with alternate swelling and nonswelling layers except that, in this case, the swelling and nonswelling layers each had a similar low charge of ~0.47 per formula unit in contrast to rectorite, with ~0.35 for the swelling and ~0.85 for the nonswelling layers (Laird and Nater, 1993). Like Robert et al. (1991), Laird and coauthors identified this I–S mineral phase as being characteristic of soils. Whereas it was identified as a component of an interstratification with smectite by Laird

et al. (1991), the ease of its separation, together with the lack of a 1 nm (10 Å) peak for illite in XRD analysis, indicated that the I–S phase in this fine clay fraction is a "textural" type of interstratification, similar to those identified by Nadeau et al. (1984), and, in soils, by Robert et al. (1991).

Illite can also form by neogenesis. Such illites include some that have formed in close proximity of feldspars, most probably out of solutions bathing the feldspars (Section 20.2.4.2). In many soils in Australia (Norrish and Pickering, 1983), and also in Iran (Mahjoory, 1975) and in some semiarid soils in North America (Nettleton et al., 1970), there is a concentration of illite toward the surface of soils, contrary to the trend in soils where micas are subject to transformation, which leads to increasing vermiculite and smectite at the expense of illite toward soil surfaces. The trend toward more illite in surface soils than at depth suggests that illite may have formed by neogenesis in these soils. Norrish and Pickering (1983) characterized a Fe-rich illite from a deposit ("Muloorina") near a seasonally wet and dry central lake in Australia (Lake Eyre) as a product of neogenesis in the lacustrine environment (although it may not be pedogenic) (Wilson, 1999). Its neogenetic origin was suggested by the very uniform shape and size of its particles, which are almost all ~0.07 μm in diameter (see Churchman, 2000, Figure 1.8a). Similarly sized and shaped particles of illite are also found in a number of soils in various parts of Australia (e.g., Willalooka—see Norrish and Pickering, 1983). The size, shape, and uniformity of these illites are in complete contrast to those originating from micas in rocks. They are more similar in their size and shape to the lacustrine illite from Muloorina (Norrish and Pickering, 1983) than to clay-size breakdown products of rock-derived micas. Berkgaut et al. (1994) found a mineral alongside halloysite in soils formed from pyroclastic layers in a semiarid climate in Israel that comprised very thin crystallites of a randomly interstratified I–S with 70% illite layers and which they considered had formed authigenically. An authigenic origin for illite/mica in soils in Hawaii that was proposed by Swindale and Uehara (1966) was later discredited when oxygen isotope ratios for the fine quartz found alongside the mica in the soils indicated its origin as tropospheric continental dust (Rex et al., 1969) and it was concluded by Syers et al. (1969) that the mica had arrived in the soils in aerosolic dust along with the quartz. Similar findings have been made in other countries such as Japan and New Zealand (e.g., Mizota and Takahashi, 1982; Stewart et al., 1984; Inoue and Sase, 1996; Nagashima et al., 2007; Marx et al., 2009).

20.3.4 Occurrence of Vermiculite in Soils

Most vermiculites in soils have formed by the transformation of micas by weathering (Wilson, 1999), although some form by the analogous transformation of chlorites (Wilson, 1999; Velde and Meunier, 2008) (see also Sections 20.2.4.3 and 20.2.4.4). Dioctahedral vermiculite is much more common in soil clays than its trioctahedral counterpart (Wilson, 1999). This is because, although the first transformation products of trioctahedral biotite (and phlogopite) and also chlorite on weathering

may well be trioctahedral, for example, the regularly interstratified phases, hydrobiotite and chlorite–vermiculite, a variety of structural and composition changes occur within the aluminosilicate layers alongside the loss of K^+ during vermiculitization (see Section 20.2.4.3). The net result of these changes is the loss of Al from the tetrahedral sheet and its gain in the octahedral sheet at the expense of Fe^{2+} and Mg. As a consequence, the layer structure becomes more dioctahedral (Wilson, 2004). Nonetheless, in the North Island of New Zealand, the occurrence of trioctahedral vermiculite in a soil derived from a composite of weathered tephras was attributed to the presence of biotite, which was known to occur in some of the parent tephras because such tephras had been preserved unweathered in layers within sediments in lakes adjacent to the soil (Lowe, 1986).

Although dioctahedral micas are much more stable in the weathering environment than trioctahedral micas, their alteration can also lead eventually to vermiculites, sometimes via regularly interstratified mica–vermiculites, as found in soils formed on mica schists in the South Island of New Zealand (Churchman, 1980). However, where pHs are moderately acid, that is, between about 4.6 and 5.8, organic matter contents are relatively low and there is frequent wetting and drying—all conditions that are conducive to the vermiculitization of micas—Al-hydroxy species are deposited, usually as polymeric cations, in the interlayers of vermiculites (and also some smectites). As a result, vermiculites in soils most often occur with incorporated hydroxyl-Al interlayers (see Churchman, 2000; Wilson, 2004).

Only few reports have considered the possibility of a neogenetic origin for vermiculite. These include those of Smith (1962), who observed that vermiculite replaced feldspars in some Scottish soils of igneous rock origin, and also Barshad and Kishk (1969), who found dioctahedral micas in soils derived from rocks containing no mica. Ildefonse et al. (1979) found that a vermiculite formed as large (up to 150 μm) crystals by crystallization from solution, hence by neogenesis, within a weathered metagabbro-containing plagioclase and amphiboles, hence no micaceous precursor. However, this trioctahedral vermiculite proved to be unstable within the soil above the weathered rock, where it was replaced by a Fe-rich smectite.

20.3.5 Occurrence of Smectite in Soils

Smectites occur in soils either because they are inherited from parent materials, have been formed by neogenesis by crystallization from solutions of the constituents of rocks, or else they are the products of the strong, or long-term, transformation of micas and chlorites. They share the possibility of an inherited origin with most other types of clay minerals. Either (1) preformation in, or prior to, the development of parent materials, for example, sedimentary rocks; (2) a detrital origin from other eroded soils or soft sediments; and arguably also (3) preformation in the soil profile but under previous conditions that differ from those currently prevailing (Buol, 1965), have been responsible for the occurrence of at least a large part of the smectites found in skeletal soils and soils in arid zones and also in some

soils found in strongly leaching climatic regimes (Wilson, 1999). Hence, inheritance may be the main or a major source of smectites in Entisols, Aridisols, and Ultisols and could also have contributed to soils classified into almost all the other orders of *Soil Taxonomy* (Wilson, 1999).

The possible origin of smectites by the transformation of micas and chlorites by acid leaching, and particularly, podzolization, processes was discussed in Sections 20.2.4.3, 20.2.4.4, and 20.2.5.3 (and particularly Section 20.2.5.3.2). The smectites formed in this way, and hence those found in the eluvial horizons of podzols, that is, Spodosols, are beidellitic (Ross and Mortland, 1966; Churchman, 1980; Carnicelli et al., 1997; Wilson, 1999; April et al., 2004), consistent with the effect of acid leaching in selectively removing structural Mg, especially, from 2:1 Si:Al aluminosilicates, which thereby become more aluminous. As the products of changes in the solid state they may also be expected to occur, like most illites (Section 20.3.3), mainly within the coarse clay fractions of soils.

Most typically, however, smectites have been formed in soils by neogenesis from alkaline solutions that have developed under poor drainage. Their formation requires high concentrations of Si and basic cations and is therefore favored typically from basic parent materials. Concentrations of smectites comprising bentonite deposits form under similar conditions, most commonly from large bodies of water, for example, lakes or marine lagoons or shallow-marine embayments into which tephra has fallen or been deposited fluvially (e.g., Millot, 1970; Chamley, 1989; Naish et al., 1993; Velde, 1995; Gardam et al., 2008). Work in New Zealand on detrital glass deposits in shallow-marine Holocene muds showed that the early diagenetic transformation of glass to smectite occurred as a two-stage process initially with diffusion-controlled hydration and dissolution of glass followed by first-order precipitation of smectite (Hodder et al., 1993). Interstitial marine waters of slightly reduced salinity (lower Ca^{2+}, Na^+) and higher concentrations of dissolved silica relative to those in ocean water played a key role in the smectite formation. Rate constants for both the dissolution of glass and the precipitation of smectite were two orders of magnitude greater than those determined for the formation of clay minerals on similar tephra-derived glassy materials in soils (Hodder et al., 1993).

Some of the knowledge of the conditions for the formation of smectites by neogenesis has come from studies of the conditions for their formation in the laboratory (e.g., Harder, 1972, 1977; Siffert, 1978; Farmer, 1997). The requirements for a successful synthesis of a smectite included pH > 7.5 and some Mg and/or Fe in solution. In nature, smectites are often found in calcareous environments, which provide a high pH, and often also where leaching is infrequent or minimal, or where there is an impediment to drainage (e.g., Fieldes and Swindale, 1954; Lowe et al., 1996). However, smectites can form from many types of rocks or geological deposits and minerals, provided that they can provide a sustained supply of the necessary cations and silica (Churchman, 2000).

The origin of smectites occurring in soils is not always clear-cut. While smectites formed by transformation may occur in relatively large particles, as already noted, smectites most often are especially fine-grained materials, a feature more consistent with a neogenetic origin. Somewhat confusingly, Boettinger and Southard (1995) found an aluminous dioctahedral smectite occurring in each of the sand, silt, and clay fractions of an arid zone soil. Although the soils also contained biotite, a trioctahedral vermiculite, and a hydroxyl-interlayered 2:1 Si:Al aluminosilicate, Boettinger and Southard (1995) considered that the smectite in all size fractions was neogenetic, with the coarser particles comprising aggregates of smectite cemented by opaline silica. A geomorphic and salinity gradient transect in another arid environment contained smectite in both its beidellitic and montmorillonitic varieties (Reid et al., 1996). These authors considered that there were multiple origins for the smectite in the soils. The beidellites probably originated from shales by transformation processes, and, while some montmorillonite was inherited, the groundwater chemistry was suitable for some having been neoformed. Studying a toposequence of soils surrounding an alkaline–saline lake in Brazil, Furquim et al. (2008) found smectites in both an upper zone that is largely higher than lake level and also a lower zone that is submerged seasonally by the lake. However, the smectites in the two zones were different structurally and compositionally, pointing to different origins for each of them. In the upper zone, a dioctahedral Fe- and Al-rich ferribeidellite had formed as a result of the transformation of micas whereas in the lower zone, trioctahedral Mg-rich smectites that were characterized as saponitic or stevensitic had formed, apparently by chemical precipitation from the lake, that is, by neogenesis (Furquim et al., 2008).

Most smectites in soils are dioctahedral, including the products of the alteration of Mg-rich serpentinite, where saponite might be expected to form (Wildman et al., 1968). Saponite often does form as a weathering product—for example, within the crystals of altered rocks containing ferromagnesian minerals (Section 20.2.4.1)—but it can give way to dioctahedral smectites in the upper parts of soil profiles (Section 20.2.5.1). Of these, the beidellitic and montmorillonitic varieties of smectites are most common in soils, although nontronites have been found there (Sawhney and Jackson, 1958). Even so, nontronite is thought to be particularly susceptible to attack by complexing agents from roots (Farmer, 1997). Generally, soil smectites have more tetrahedral Al and more octahedral Fe than the type of montmorillonite associated with most bentonites (Wilson, 1987; Weaver, 1989). There appears to be a tendency toward a ferribeidellite composition for smectites in soils, albeit that a given soil smectite is most probably heterogeneous in composition. Petit et al. (1992) found that the smectite in a lateritic profile in Brazil could be described as a true solid solution between nontronite and beidellite endmembers.

Smectites themselves can be subject to alteration and change within soils. As already discussed (Section 20.3.1), smectites can incorporate kaolinite layers to become interstratified kaolinite–smectites (K–S), most probably following the prior adsorption of hydroxyl-Al species into their interlayers (e.g., Ryan and Huertas, 2009). Kaolinite appears to be the ultimate product of this process (Section 20.3.1). Nevertheless, Fisher and Ryan (2006), studying

a chronosequence of soils on fluvial fill terraces in tropical Costa Rica, found that there was a transition with age from smectite-rich to kaolinite-rich soils without any K–S intermediates being formed along the way. The smectite was a beidellite and the kaolinite was disordered. These authors generalized their results and those of others who had studied this transition by suggesting that K–S forms as an intermediate phase in the transition from S to K where the MAP is between ~500 and ~2000 mm, whereas the transition occurs without a K–S intermediate when MAP is more than ~3000 mm. The changes may be more subtle, however, resulting in a change within the composition and structure of the smectite. Thus, a low-charge beidellite inherited from the parent pelitic sediment was converted to a high-change beidellite during the formation of a Vertisol (Righi et al., 1995).

20.3.6 Occurrence of Palygorskite and Sepiolite in Soils

The fibrous 2:1 Si:Mg clay minerals, palygorskite and sepiolite—the so-called hormite minerals—undoubtedly occur in the sedimentary parent materials of many soils containing these minerals; and, until relatively recently, their inheritance from these parent materials was considered to be their source in soils (Singer, 2002). However, both have since been found to form in soils in many countries, including Israel, Syria, Egypt, Morocco, Iraq, Iran, Turkey, Spain, Portugal, China, the United States, Mexico, Australia, South Africa, and Argentina (see summaries in Zelazny and Calhoun, 1977; Singer, 1989, 2002; Churchman, 2000; and reports by Owliaie et al., 2006; Bouza et al., 2007; Kadir and Eren, 2008). Even so, they almost invariably form in calcareous environments (Singer, 2002) and often in calcretes, in particular. Verrecchia and Le Coustumer (1996) have listed many reports of their occurrence in calcretes and there are also later similar reports (e.g., Kadir and Eren, 2008). The main conditions under which they form have been summarized as (1) a fluctuating saline or alkaline groundwater, (2) strong and continuous evaporation, and (3) a sharp textural transition (Singer and Norrish, 1974; Singer, 1989, 2002). A strong textural transition means that water is likely to accumulate at the transitional boundary, hence maintaining a saturated solution from which the hormite minerals can precipitate, under even arid conditions. Calcretes (or caliches) provide such a transition and, hence, these minerals are often found in or below these secondary carbonate concretions. Albeit that palygorskite is most commonly associated with calcareous soils, gypsiferous soils appeared to contain more palygorskite than calcareous soils within the same region in Iran (Khormali and Abtahi, 2003; Owliaie et al., 2006). It appears that an increase in the concentration of Mg that occurred upon the precipitation of gypsum was particularly conducive to the neogenesis of palygorskite (Owliaie et al., 2006).

Chemically, the neogenetic formation of both palygorskite and sepiolite requires considerable magnesium and silicon in a high pH solution. Palygorskite would also require some Al and Fe (Singer, 1989). Although there had been no reproducible synthesis of palygorskite, at least up until Singer (2002), sepiolite has been synthesized many times from alkaline solutions with high activities of Mg, and Si (see, e.g., Churchman, 2000), confirming the conditions for their formation in nature. In nature, the required Mg may originate in dolomite or Mg-containing calcite. The conversion of high-Mg calcite into low-Mg calcite has been observed to accompany palygorskite formation (Singer, 1989, 2002). The required Si may derive from the dissolution of any silicate mineral, including illite and chlorite (Galán et al., 1975; Galán and Castillo, 1984; Dias et al., 1997; Owliaie et al., 2006), smectites (see Singer, 1989; Bouza et al., 2007; Kadir and Eren, 2008), and even quartz (Singer, 1989, 2002). It is not surprising that smectites may give rise to palygorskite and/or sepiolite because the chemical conditions of high [Mg], high [Si], and a high pH are shared among these three mineral types. Bouza et al. (2007) suggested that, as Mg concentration increased with time in calcic and calcic–gypsic horizons in a weathering profile in Argentina, smectite gave way to palygorskite. Nor, therefore, is it surprising that palygorskites and sepiolites often occur together with smectites in soils and may disappear in their favor, as observed by Paquet and Millot (1972), Khormali and Abtahi (2003), Owliaie et al. (2006), and Bouza et al. (2007), for example. Paquet and Millot (1972) suggested that this dissolution and reformation occurs when rainfall rises to a MAP of merely 300 mm. Khormali and Abtahi (2003) found an inverse relationship between the occurrences of palygorskite and smectite with regard to soil available water, measured by the ratio of precipitation to plant transpiration. Smectite was favored over palygorskite at higher values of soil available water. On the other hand, in soils in an irrigated cropping area in India, smectite predominated in the rain-fed soils, but irrigation, which exacerbates salinity and sodicity (and attendant waterlogging), had led to the formation of palygorskite in just 40–50 years (Hillier and Pharande, 2008). It appears that evaporative concentration of the irrigation waters has led to solution conditions more favorable to the formation of palygorskite than smectite. It has also been observed that the highly siliceous environment in which palygorskite and sepiolite form may give rise to an excess of Si in solution, which results in deposits of secondary silica upon their formation (Singer, 1989; Verrecchia and Le Coustumer, 1996). Experiments with nonsoil palygorskite and sepiolite showed that they were unstable, giving rise to kaolinite, within the rhizospheres of common agricultural crops (Khademi and Arocena, 2008). In this situation, the high acidity of the rhizospheres and the extraction of Mg from the minerals by the crops had destabilized them (see also Section 20.2.5.3).

Palygorskite is much more common than sepiolite in soils (Singer 1989, 2002). However, Bouza et al. (2007) found that sepiolite formed chronologically after palygorskite within a mature calcrete.

20.3.7 Occurrence of Iron Oxides in Soils

Iron oxides and oxyhydroxides (generically known as "iron oxides") occur in almost all soils (Allen and Hajek, 1989). They are usually the product of the oxidation of Fe(II) occurring

either within the structure of primary minerals or upon their breakdown by weathering, whereupon the newly formed Fe(III) hydrolyzes (e.g., Churchman, 2000). The resulting iron oxides are essentially insoluble at the pH values encountered in soils and are mainly remobilized by reduction. As we have seen for soils from volcanic materials particularly (Section 20.2.4.6), the specific Fe oxides that occur in a soil depend largely on the environmental conditions under which they were formed.

Following Schwertmann and Taylor (1989), Cornell and Schwertmann (1996), Churchman (2000), Bigham et al. (2002), and Schwertmann (2008), in the main, we can make some generalizations regarding the conditions under which goethite or hematite are formed. Goethite, which is yellow to yellowish brown in color depending partly on particle size, is found in soils almost everywhere. But it is most common in cool and temperate climates and where there are high contents of organic matter. Hematite, by contrast, is most common in aerated soils in generally warmer temperatures in the tropics and subtropics and also in arid and semiarid regions, as well as in xeric (or Mediterranean) climate zones. Hematite is characteristically red in color. High organic matter content favors the formation of goethite over hematite, and it is not uncommon to find soils with yellowish goethitic topsoil over red hematitic subsoil. Goethite also tends to be favored at low pH, hematite at higher pHs. Laboratory syntheses have shown that from ferrihydrite as a starting material, the most goethite was formed at a pH near 4, while the maximum yield of hematite occurred at pHs between 7 and 8. Synthetic work has also shown that hematite is favored over goethite by a high rate of release of Fe into solution. This finding is consistent with the observed occurrence of hematite in a soil on basalt but not in a nearby soil on shale (Kämpf and Schwertmann, 1982). Hematite is the preferred phase where the activity of water is low, as in small pores, and is also the most stable phase thermodynamically in smaller particles (Langmuir and Whittemore, 1971). These latter authors predicted that goethite would be the stable phase in particles >76 nm and hematite in smaller particles. While Al substitution for Fe is common in all Fe oxides, synthetic work has shown that goethite can accommodate more Al in its structure than hematite. Fritsch et al. (2005) explained yellowing toward the surface of a weathering profile on sediments in Brazil by upward changes in both the relative amounts of goethite and hematite and also in the extent of Al substitution of goethite. Hematite and Al-poor goethite that are remnants of prior weathering of the sediments in a different climate have dissolved, with Al-rich goethite forming in their place by recrystallization, according to Fritsch et al. (2005). In some tropical soils, the formation of hematite is at the expense of goethite when temperatures are high, the aeration good, the pH near neutrality, and organic matter is rapidly transformed (Etame et al., 2009).

Poorly crystallized or nanocrystalline ferrihydrite is also widespread in soils, particularly in those formed from tephra, and also in the illuvial or placic B horizons of podzols (see Churchman, 2000, and McDaniel et al., Section 33.3). Its formation is favored by rapid oxidation of Fe and when silicates and organic matter, particularly, but also other ions, are present

in soil solutions to inhibit crystallization of goethite and other Fe oxides, including lepidocrocite (Churchman, 2000; Bigham et al., 2002), and soil ferrihydrites can contain significant amounts of Si (Childs, 1992). Nonetheless, ferrihydrite often also occurs in association with goethite and lepidocrocite (Bigham et al., 2002). In the laboratory, it has been found that ferrihydrite is easily converted to hematite (Schwertmann and Taylor, 1989). Nonetheless, ferrihydrite and hematite are rarely found together in soils, indicating that the conversion is also rapid in nature. Even so, Adamo et al. (1997) found ferrihydrite together with both hematite and goethite in the products of weathering of a volcanic rock by lichens. They appear to occupy different microsites in the altering rock.

The formation of lepidocrocite is favored by seasonally anaerobic conditions in noncalcareous soils. It forms when there is an accumulation of Fe^{2+} together with a slow rate of oxidation (Alekseev et al., 2005). Lepidocrocite is found in ocherous mottles, root mats, and crusts (Cornell and Schwertmann, 1996; Churchman, 2000; Bigham et al., 2002). It is often found in association with goethite. Lepidocrocite was found in all three soils in a hydromorphic sequence in Ohio (Smeck et al., 2002), but in different amounts according to ease of drainage. These authors concluded that the formation of lepidocrocite was favored in horizons that were saturated for between 5% and 50% of the part of the year when the soil temperatures were >5°C. Alekseev et al. (2005) found that lepidocrocite was concentrated in the bottom of the "active" layer immediately above permafrost in Russia. The existence of the permafrost below concentrated solutes in this layer and the low temperature slow oxidation.

Maghemite occurs in soils in the tropics and subtropics, often in association with hematite, and most often in surface soils (Schwertmann and Taylor, 1989; Churchman, 2000; Bigham et al., 2002). A common mode of formation of maghemite appears to be through the heating of other Fe oxides in the presence of organic matter, as occurs in bush and forest fires. Schwertmann and Taylor (1989) suggested that a temperature of between 300°C and 425°C is conducive to the formation of maghemite. It may also form by the oxidation of precursor magnetite in parent materials, according to the authors cited herein. However, a completely different, alternative origin for this magnetic mineral was suggested by Rivers et al. (2004) to explain its occurrence in association with hematite in soils on alluvium that has in-filled low-lying areas in Tennessee. These authors suggested that there had been reduction of Fe(III), either microbial or abiotic, to form magnetite in the hydromorphic soil. Magnetite occurs in many soils but is usually found in the sand and silt fractions rather than the clay fraction and is generally considered to have been inherited from the parent rock. Nonetheless, a partially oxidized form of magnetite was found to occur along with hematite in nodules in an Oxisol in Brazil that has developed from magnetite-free parent materials, so a pedogenic origin has been invoked for the mineral in this soil (Viana et al., 2006). These authors considered that the magnetite in this soil formed under seasonal burning and also that it is becoming converted to hematite by atmospheric oxidation in air. Some

authors (Maher and Taylor, 1988; Geiss and Zanner, 2006) have considered that some magnetite in soils may have had a bacterial origin (see Section 20.2.5.3.5).

Other Fe oxides formed in soils include schwertmannite, from acid sulfate soils (see Section 20.2.5.3.4), and "green rusts," found in a highly hydromorphic soil by Trolard et al. (1997).

Iron oxides can be extremely labile in the soil environment. They are quickly reduced or oxidized when redox conditions change in soils, as they can do intermittently over even quite short time periods (hours to days or weeks) and also seasonally. The activity of microbes can play a central role in the formation and dissolution of iron oxides. A wide variety of microorganisms are involved in these processes and much remains to be understood about their operation (Fortin and Langley, 2005). Suffice it to say here that Fe oxides are often biogenic (see also Section 20.2.5.3). Furthermore, there is widespread recognition that Fe is commonly mobilized in soils by reduction occurring through the activity of microbes in anoxic conditions (Robert and Berthelin, 1986; Schwertmann and Taylor, 1989; Robert and Chenu, 1992). The ensuing reduction occurs along with the oxidation of decomposing biomass.

It is a feature of iron oxides in soils that they are most often closely associated with the other components of soils, including particularly aluminosilicate clay minerals, which as small-charged particles attract Fe oxides and other reactive species to their surfaces and pores. Following the well-accepted convention that (citrate-) dithionite-extractable Fe represents free, or well-crystallized oxides, (acid ammonium) oxalate-extractable Fe represents poorly crystallized oxides, and pyrophosphate-Fe stands

for the metal associated with organic matter, Table 20.5 shows the relative amounts of these forms of Fe and their distribution in relation to aluminosilicate clays in a range of soils, mainly taken from papers in the most recent literature, which have presented data for both different forms of Fe and also estimates of the relative abundance of the phyllosilicates present. The data for selected horizons within one soil of each type are given in Table 20.5. The soils in Table 20.5 represent a wide range of soil orders and lithologies, and soil horizons are also given for each soil. The table shows that there were substantial proportions of iron oxides and Fe-organic complexes in their various forms in all horizons of all the soils examined. They were associated with virtually all of the types of phyllosilicates that occur in soils.

20.3.8 Occurrence of Manganese Oxides in Soils

Manganese oxides are a widespread but usually minor component of soils (McKenzie, 1989; Dixon and White, 2002). Generally, they occur in soils in small particles, leading to poor peaks in XRD. They are more abundant in soils from mafic, that is, basic, rocks than from siliceous materials (Dixon and White, 2002). Because reduced Mn (Mn(II)) is very soluble as a divalent cation and is mobile in soils, they commonly form as coatings on rocks, on ped surfaces, and on other soil particles, as well as in concretions, segregations, pans, and nodules, often with Fe oxides (McKenzie, 1989; Churchman, 2000; Dixon and White, 2002). They are characteristically black in color but give a brown streak.

TABLE 20.5 Occurrence of Iron in Oxides and Oxyhydroxides and in Organic Complexes in a Range of Soils with Different Phyllosilicate Clays

Soil Subgroup	Main Lithology	Horizon	pH	Fe^a ($g\ kg^{-1}$)				Phyllosilicates in Clay Fraction[b]	References
				Fe_d	Fe_o	Fe_p	Fe_t		
Gypsic Haplustept	Gypsiferous	Ay	7.6	7.8	n.d.	n.d.	16.0	Paly ~ smec ~ chl > ill	Owliaie et al. (2006)
		2By2	7.7	2.1	n.d.	n.d.	9.9	Paly > chl ~ ill > smec	
Humic Umbrisol	Granite, orthogneiss	Ah1	4.0[c]	22.0	14.0	15.2	51.2	Verm ~ HIS ~ HIV > ill = kaol > chl	Zanelli et al. (2006)
		BC	5.0	6.6	2.2	0.5	9.3	Ill ~ HIV ~ (ill-HIV) ~ allo > verm ~ chl	
Mollic Eutrocryept	Calcschists, serpentinites	A	6.5	16.6	8.8	3.4	28.2	Chl > ill > serp > talc	Scalenghe et al. (2002)
		C	8.1	12.3	4.5	0.8	17.8	Chl > ill > serp > talc	
Typic Haplorthod	Shale, sandstone	A	3.7	1.4	0.8	1.1	3.3	Ill ~ kaol ~ (ill-verm) > verm	Pai et al. (2004)
		Bs2	4.8	48.4	16.9	10.1	75.4	Ill ~ kaol ~ HIV > (ill-verm) >verm	
Ultic Haplorthod	Marine sediments	E	3.7	37.6	8.4	5.5	51.4	Verm ~ (i-v) ~ ill ~ qtz > kaol ~ gibb	Lin et al. (2002)
		Bhs	4.2	59.2	17.7	19.9	86.8	Verm > HIV > ill ~ kaol ~ qtz ~ gibb	
Typic Haploxerert	Olivine basalt	Ap	6.7	26	n.d.	n.d.	53	(Kaol-smec 40:60) > (k-s 95:5) >ill	Vingiani et al. (2004)
		C	8.1	21	n.d.	n.d.	96	(K-s 40:60) > kaol ~ (k-s 95:5) >ill	
Andic Haploxeralf	Rhyolitic ash on andesite	A1	6.8	3.2	0.5	0.1	3.8	Hall ~ HIV > allo	Takahashi et al. (1993)
		2Bt3	6.2	10.1	0.5	0.1	10.7	Hall > gibb ~ allo	
Acrudoxic Fulvudand	Andesitic lava	Ah1	3.9	2.8	1.7	1.5	6.0	Gibb > kaol > goe > ferr	Nieuwenhuyse et al. (2000)
		BC	5.2	4.1	0.8	0.0	4.9	Allo > goe > ferr	
Orthoxic Tropohumult	Olivine basalt	Ap	5.9	9.0	n.d.	n.d.	n.d.	Kaol > goe	Gallez et al. (1976)
		B21t	5.8	10.3	n.d.	n.d.	n.d.	Kaol > goe	

[a] Values of Fe extractable by dithionite (Fe_d), oxalate (Fe_o), and pyrophosphate (Fe_p), also total (Fe_t).

[b] Minerals usually given by first three or four letters of their name, except for quartz (qtz) and HIV; interstratified minerals in brackets: their components often by first letters only; relative amounts of minerals as given in the appropriate reference.

[c] In $CaCl_2$, other pH values in H_2O.
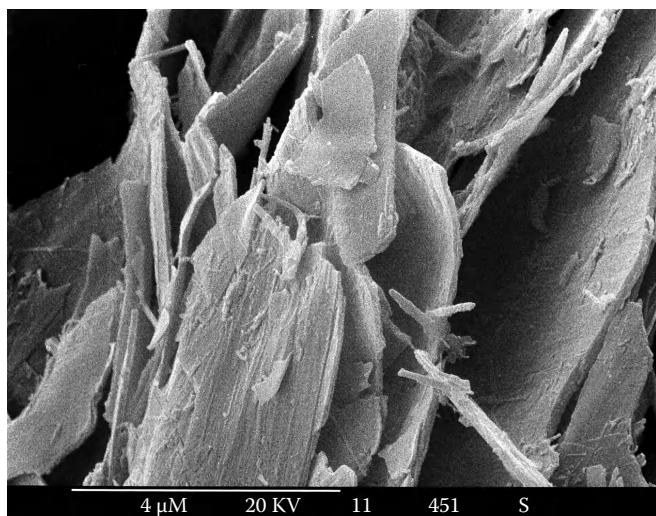
FIGURE 20.11 Scanning electron micrograph of crystals of todorokite, with some kaolin minerals, in infill formed as a result of the weathering of granitic saprolite in Hong Kong. (Photo by Stuart McClure. Reproduced with the permission of the Geotechnical Engineering Office, Civil Engineering Office, Hong Kong.)

Their distinctive appearance, especially in, for example, concretions and nodules, gives a strong indication of prior oxidation, including by seasonal drying as noted earlier in this chapter (e.g., Birkeland, 1999; Vepraskas et al., 2004). Churchman et al. (2010) used their occurrence to trace a history of dehydration, leading to the conditions for the formation of kaolinite rather than halloysite in some saprolites in Hong Kong. Manganese oxides form when conditions favor their oxidation; hence, they are often found in subsoils with a higher pH than the upper layers and are common in calcareous soils. Manganese also occurs as a Mn carbonate (rhodochrosite) in calcareous soils (Dixon and White, 2002).

Among the polymorphs of MnO_2, birnessite, especially, and lithiophorite have most often been identified in soils. However, birnessite occurs in soils as poorly developed crystals and the crystals of birnessite produced by biological oxidation of Mn(II) in the laboratory are similarly poorly developed, as are those precipitated from solutions with a high Si content (Dixon and White, 2002), suggesting that their formation in soils occurs via microbiological oxidation or at an early stage of soil formation, or both (McKenzie, 1989; Dixon and White, 2002). Todorokite has been identified in Texas in nodules from a Vertisol and on the surface of weathered siderite boulders (Dixon and White, 2002) and appeared in large crystals in saprolite on both granite and volcanic tuff in Hong Kong (Churchman et al., 2010), as seen in Figure 20.11, alongside both halloysite tubes and platy kaolinite particles.

20.3.9 Occurrence of Aluminum Hydroxides, Oxyhydroxides, and Oxides in Soils

Gibbsite is the most common of the various hydroxides, oxyhydroxides, and oxides of aluminum in soils (e.g., see Churchman, 2000). It occurs in soils of any age whenever Si is

in low supply. Hence, it occurs under strong leaching. Where there is a gradation in the extent of leaching, often because of an altitudinal change, gibbsite may occur in the most heavily leached soils whereas kaolinite occurs when there is a lesser degree of leaching. This tendency appeared to be the case for temperate mountain soils in South Carolina (Norfleet et al., 1993), as well as for soils in the humid tropics, where gibbsite was dominant in the highlands and kaolinite in the lowlands (Herbillon et al., 1981). In some Oxisols, gibbsite may be most abundant in upper horizons but decrease in concentration with depth (Huang et al., 2002). These authors also pointed out that, in the tropics, particularly, landscapes may be extremely old and the occurrence of gibbsite not necessarily related to the present soil environment. This polygenesis was illustrated in a study by Bhattacharyya et al. (2000), who found gibbsite occurring alongside mica–vermiculite phases containing hydroxy-Al interlayers in an acidic tropical soil in India. However, it has long been considered (Jackson, 1963) that the presence of these latter aluminous interlayers precludes the formation of gibbsite by the so-called antigibbsite effect. Bhattacharyya et al. (2000) explained the presence of gibbsite and these interlayers together by the prior formation of gibbsite, possibly mainly from sillimanite, in an earlier neutral to alkaline pH environment. The aluminous interlayers, by contrast, are products of weathering in the current acidic environment. An apparent contradiction of the "antigibbsite" effect was also seen by Jolicoeur et al. (2000) when they observed each of hydroxyl(-Al) interlayered vermiculite (HIV), kaolinite and/or halloysite, and gibbsite occurring together as pseudomorphs after biotite in scanning electron micrographs of saprolites and soils in central Virginia. They were considered to coexist because they each occupy their own particular microsites and microenvironments of weathering, especially in the vicinity of the altering rock material (Jolicoeur et al., 2000). Gibbsite in soils may also be inherited from saprolites (e.g., Huang et al., 2002; Simas et al., 2005). In these cases, its content decreases toward the surface within the soil profile. In another situation, the co-occurrence of both gibbsite and halloysite, together with kaolinite, goethite, hematite, maghemite, quartz, and cristobalite in old soils formed on strongly weathered andesite or associated alluvium in tropical Costa Rica, was studied by Kleber et al. (2007). They attributed the clay mineral depth-distribution patterns first to intense weathering, with gibbsite increasing with depth as the end product of prolonged tropical weathering, as noted earlier. However, the seemingly paradoxical enrichment of kaolin group minerals and quartz near the soil surface was attributed to ongoing dissolution of Si (quartz being unstable at the silica concentrations necessary for gibbsite formation, namely <0.5 ppm) and its vertical redistribution by a plant-based "biological resilication mechanism" (Kleber et al., 2007, and see also Section 20.2.5.3.3).

Gibbsite may also occur as coatings or infillings of voids in weathered upper soil horizons (e.g., Huang et al., 2002). Gibbsite is also favored if parent materials are rich in Al (e.g., Lowe,

1986), as illustrated by the example of sillimanite within gneiss in Bhattacharyya et al.'s (2000) study.

Among Al oxyhydroxides, boehmite has been identified, along with some other poorly ordered, nanocrystalline, and/or amorphous phases in lateritic profiles, particularly at their surfaces, and also in tropical soils (e.g., Churchman, 2000; Huang et al., 2002). Some may have resulted from the dehydration of gibbsite. Other Al hydroxides, bayerite, nordstrandite, and doyleite, have been identified only rarely in soils (Huang et al., 2002). The Al oxyhydroxide, boehmite, and the Al oxide, corundum, have been reported at the surfaces of ferralitic (lateritic) profiles, whereas corundum has often been seen in the same soils that contain maghemite, for example, in ferralitic duricrusts in Western Australia (Anand and Gilkes, 1987), and these authors proposed that it formed from gibbsite and/or boehmite by dehydration in bushfires. A bushfire origin has also been suggested for maghemite (Section 20.3.7 herein).

20.3.10 Occurrence of Phosphate, Sulfide, and Sulfate Minerals in Soils

Phosphate minerals in soils appear to derive mainly by inheritance or by alteration through weathering of forms of apatite, namely, fluorapatite and hydroxyapatite, from rocks or other geological deposits, or else from reactions between phosphorus fertilizers and soil minerals (Lindsay et al., 1989; Churchman, 2000; Harris, 2002). The products are various forms of phosphates in which the Ca in apatite is substituted, to a greater or lesser degree, by other cations. Hence, there are forms of variscite, where the major cation is Al, strengite (Fe), also barrandite (Al, Fe), vivianite [Fe(II)], crandallite (Ca, Al), plumbogummite (Pb, Al, and other elements), gorceixite (Ba, Al), leucophosphite (K and Fe), along with many other variations (Norrish, 1968; Lindsay et al., 1989; Churchman, 2000; Harris, 2002). A hydrated Al hydroxyl phosphate, wavellite, is the most common Al phosphate in soils derived from phosphoritic marine sediments, which are the most extensive phosphate-rich geological materials at the Earth's surface (Harris, 2002). Nevertheless, phosphates generally occur very sparsely in soils and have to be either concentrated or examined by electron optical methods to be identified (e.g., Norrish, 1968).

Sulfide minerals in soils originate ultimately from the bacterial reduction of sulfate in seawater (Doner and Lynn, 1989; Fanning et al., 2002). They occur in the unoxidized parent materials of soils and are very readily oxidized on exposure to air and drying. Pyrite is the most common sulfide mineral in the unoxidized parent materials. Sulfates, including sulfuric acid, are the products of their oxidation. The reaction of the sulfuric acid thus produced with silicates leads to hydroxyl Fe sulfate minerals, especially jarosite and schwertmannite. These, and their variants, occur typically in acid sulfate soils with their attendant environmental problems (see also Section 20.2.5.3.4). Gypsum occurs in soils as a result of its inheritance from gypsiferous parent materials (or from agricultural application), while barite, of unknown origin, has also been found in some soils, often as a white powder (Fanning et al., 2002).

20.3.11 Occurrence of Pyrophyllite, Talc, and Zeolites in Soils

Both pyrophyllite and talc are rare in soils (Zelazny and White, 1989; Churchman, 2000; Zelazny et al., 2002). As minerals originating at high temperatures and pressures, they are seldom inherited in soils. Pyrophyllite, particularly, is broken down into finer fractions by physical processes (Zelazny et al., 2002). As with other types of minerals, it is the dioctahedral species, pyrophyllite, that is more stable in the soil environment than its trioctahedral counterpart, talc (Zelazny et al., 2002). Generally, talc weathers quickly to nontronite or to Fe oxides. Both pyrophyllite and talc have been protected from breakdown by their encapsulation within Fe oxides in some soils (Zelazny et al., 2002).

Occurrences of zeolites have rarely been reported in soils, but some may have been overlooked because they are generally only present in trace amounts (Boettinger and Ming, 2002). Apart from their inheritance from zeolitic parent materials, such as through hydrothermal alteration (e.g., Kirkman, 1976), some zeolites have formed pedogenically (Ming and Mumpton, 1989; Churchman, 2000; Boettinger and Ming, 2002). The most common pedogenic zeolite is analcime. Along with chabazite, mordenite, natrolite, and phillipsite, analcime has been found to form in salt-affected, alkaline soils, which contain sodium carbonate and have a high pH (Churchman, 2000; Boettinger and Ming, 2002). These soils were both volcanic and nonvolcanic in origin. A number of zeolites have been found in soils in Antarctica that appear to have a secondary origin. These include stilbite, from the weathering of dolerite (Claridge and Campbell, 1968), phillipsite, from tephra dissolved in lakes (Claridge and Campbell, 2008), and chabazite, which has been identified by many workers and which Claridge and Campbell (2008) suggested is quite widespread as a product of thin saline films in intergranular spaces in soils. Other zeolites that have been inherited in soils from parent materials include clinoptilolite, heulandite, gismondine, mordenite, stilbite, and laumontite (Boettinger and Ming, 2002).

20.3.12 Occurrence of Neogenetic Silica in Soils

Because leaching in soil formation results in desilication, silica is often mobilized and then (if not combined with Al for example to form an aluminosilicate clay) can be reprecipitated deeper in the soil environment. Chemically precipitated overgrowths of quartz, particularly on carbonates, are relatively common in soils (Drees et al., 1989; Churchman, 2000). Quartz apparently formed by crystallization out of the dissolution products of microcline (Estoule-Choux et al., 1995; Section 20.2.4.2). In soils across two chronosequences on fluvial deposits in arkosic sands in southern California, Kendrick and Graham (2004) found that the amount of pedogenic, opaline silica (measured by extraction with tiron) increased with duration of soil development from 1.2% to 4.6% of the soil. The loss of Si from primary minerals approximately equaled the gain of secondary (opaline)

silica (Kendrick and Graham, 2004). A number of studies of fine quartz from soils and also shales found a systematic increase in oxygen isotopic ratios with decreasing size, suggesting that the smallest particles of quartz were authigenic (Sridhar et al., 1975; Churchman et al., 1976; Clayton et al., 1978). However, much neogenetic silica in soils occurs in other forms besides quartz (Monger and Kelly, 2002).

Silica cements are common in hardpans or duripans (Chadwick et al., 1987) and also fragipans (Harlan et al., 1977; Marsan and Torrent, 1989) in soils and have been seen to coat surfaces of primary minerals, for example, micas, and to fill gaps in altered minerals including feldspars and micas (Chadwick et al., 1987; Singh and Gilkes, 1993). In the extreme, silica forms massive silcretes in acidic environments (Thiry and Simon-Coinçon, 1996). Secondary silica is sometimes identified as a form of opal, either opal-CT, which is usually inherited from rocks (Munk and Southard, 1993), or opal-A, which is biogenic, and includes plant opals. These are common in some Andisols (Drees et al., 1989) and may be recycled through the death and regrowth of plants (Alexandre et al., 1997). Precipitation of secondary silica may be aided by plant-based processes related to Si uptake and recycling (Henriet et al., 2008) and by evaporation or freezing of soil water. Inorganic silica polymorphs are distinguishable from plant-derived forms of silica, phytoliths, because the latter have more complex shapes inherited from biological cells (Nanzyo, 2007). Farmer et al. (2005) concluded from their study on the concentration and flux of Si in European forest soils (both podzols and acid brown soils) that phytoliths must be the principal immediate sink of silica in soil solution (although weathering is the ultimate source). During the growing season, forest vegetation takes up most of the Si released through weathering of soil minerals and phytolith dissolution and converts it into (new) phytoliths, the latter then becoming the main source of Si leached from the soil during winter rains and spring snowmelt (Farmer et al., 2005). Opals may be transformed to quartz in duripans in soils and also by diagenesis in fossilized wood and sediments (Drees et al., 1989).

Elsass et al. (2000) examined hard plates in laminar horizons and gray mottles at depth in indurated volcanic soils ("tepetates") in Mexico and found strong evidence for the transformation of halloysite to cristobalite via an amorphous opal-A stage. In this case, it appears that the secondary silica phase, first opal-A and then eventually cristobalite, had formed as a result of subtraction of Al from halloysite into solution, rather than through addition of Si from solution as in other cases of the occurrence of secondary Si in soils. Other studies, however, have demonstrated that cristobalite in soils is usually of primary volcanic origin (e.g., Mizota et al., 1987; Wallace, 1991; Mizota and Itoh, 1993). In northern New Zealand, cristobalite isolated from three weathered, halloysitic volcanic soils was identified as alpha cristobalite (opal-C) and was invariably found to be accompanied by tridymite (Wallace, 1991). SEM and oxygen isotope data, along with the highly ordered crystal structure and subhedral grain morphologies, were interpreted to indicate that the cristobalite and tridymite had formed at high temperature from primary volcanic sources (cf. Lowe, 1986; Wallace, 1991).

20.3.13 Occurrence of Titanium and Zirconium Minerals in Soils

Although the major forms of titanium oxide, namely, anatase and rutile, and also their polymorph, brookite, as well as the Fe–Ti minerals, ilmenite, pseudorutile, titanomaghemite, and others, can all form by neogenesis in intensely weathered soils, only anatase and pseudorutile commonly form this way (Milnes and Fitzpatrick, 1989; Churchman, 2000; Fitzpatrick and Chittleborough, 2002). Otherwise, the titanium minerals in soils are inherited from parent materials. Fitzpatrick and Chittleborough (2002) leave open the possibility that zircon, the most common Zr-containing mineral, could also form authigenically in intensely weathered soils. Nevertheless, it is generally inherited from parent materials.

Pseudorutile, which occurs widely in soils, is primarily found there as an alteration product of ilmenite (Fitzpatrick and Chittleborough, 2002). In turn, pseudorutile can dissolve to yield either rutile (Grey and Reid, 1975) or anatase (Anand and Gilkes, 1984). The authigenic formation of titanomaghemite and also ferrian ilmenite may occur by heating Fe–Ti oxides to high temperatures as occurs in bushfires (Fitzpatrick and Chittleborough, 2002). Some Ti^{4+} can become incorporated in other minerals, notably the Fe oxides goethite and hematite, by isomorphous substitution, thereby potentially leading to a reversal of the charge of the Fe oxides (Fitzpatrick and Chittleborough, 2002). Singer et al. (2004) have also found Ti as an isomorphous substitute for Al in halloysite (Section 20.3.2).

20.3.14 Occurrence of Highly Soluble Minerals in Soils

The most common minerals that are laid down following their dissolution from the preformed solid state, whether primary or secondary in origin, and then reprecipitation after evaporation, include calcium carbonate, mainly as calcite, but also as aragonite and Mg calcite (Doner and Grossl, 2002). Magnesite and also dolomite may be formed in that fashion and gypsum probably has a similar origin in arid and semiarid environments (Kohut et al., 1995; Churchman, 2000). Bassanite (or hemihydrite) is usually found in surface soils, indicating its likely evaporative origin (Doner and Grossl, 2002). The highly soluble salts, halides, sulfates, and some carbonates often occur in saline soils and also in arid environments such as Antarctica (Section 20.2.5.2), which is where Bockheim (1997) observed niter ($NaNO_3$) in soils. It is an indication of its extreme aridity that nitrate, borate, chromate, and perchlorate salts have been found in the Atacama desert in Chile (Erickson, 1983; Doner and Grossl, 2002).

20.3.15 Occurrence of Secondary Minerals in Soils: Summary

The main properties of the common types of secondary minerals, and characteristics of their occurrence in soils, are summarized in Table 20.6.

TABLE 20.6 Properties of the Common Types of Secondary Minerals and Characteristics of Their Occurrence in Soils

Group	Common Soil Mineral	Chemical Formula	Related Phases and Names	Specific Surface (m^2g^{-1})	CEC $(cmol_+ kg^{-1})$	Characteristics in Soils	Soils of Common Occurrence
Silicates							
Kaolin	Kaolinite	$Al_2Si_2O_5(OH)_4$	Dickite, nacrite	6–40	0–8	Very small euhedral particles, associated with much Fe	Widespread, high in well-weathered soils
	Halloysite	$Al_2Si_2O_5(OH)_4 \cdot 2H_2O$	Endellite[a], metahalloysite[a] 1.0 nm (10 Å) halloysite (= hydrated phase), 0.7 nm (7 Å) halloysite (= dehydrated phase)	20–60	5–10	Generally small tubular or spheroidal particles, associated with Fe	Where humid, especially from tephra (including volcanic ash)
Interstratified kaolin	Kaolinite–smectite	Variable, depending on proportions of K:S	Halloysite–smectite	Unknown	30–70		Moderately drained
Mica	Illite	$K_{0.6}(Ca,Na)_{0.1}Si_{3.4}Al_2 Fe^{III}Mg_{0.2}O_{10}(OH)_2$		55–195	10–40		Widespread, especially weakly weathered soils
Interstratified mica	Illite–vermiculite	Variable, depending on proportions of I, V	Mica–vermiculite	Unknown	Unknown	Usually regular, mica–vermiculite	Eluvial horizons of podzols
	Illite–smectite	Variable, depending on proportions of I, S	Mica–smectite	Unknown	Unknown	Either regular, mica–beidellite, or random, poss. with single layers	Regular: eluvial horizons of podzols. Random: very widespread, often in agricultural soils
Interstratified chlorite	Chlorite–vermiculite	Variable, depending on proportions of C, V	Corrensite	Unknown	Unknown		At very early stages of weathering
	Chlorite–smectite	Variable, depending on proportions of C, S	Chlorite-swelling chlorite	Unknown	Unknown		At very early stages of weathering
Vermiculite	(dioctahedral)	$K_{0.2}Ca_{0.1}Si_{3.2}Al_{0.8}$ $(Al_{1.6}Fe_{0.2}Mg_{0.2})$ $(Al_{1.5}[OH]_4)O_{10}(OH)_2$	Pedogenic chlorite, HIV, 2:1–2:2 intergrade, chloritized vermiculite	Unknown	pH-variable		Leached, mildly acid soils
	(trioctahedral)	$M_x^{II}(Mg,Fe)_3$ (Al_xSi_{4-x}) $O_{10}(OH)_2 \cdot 4H_2O$		50–150	100–210		Early stages of weathering, especially below soil zone
Interlayered	Pedogenic chlorite	Variable, depending on whether layers are vermiculite or smectite and interlayered species	HIV, 2:1–2:2 intergrade, chloritized vermiculite HIS, chloritized smectite	Unknown	Unknown		Intermediate pHs, ~4.6 to 5.8, with wetting and drying, low organic matter
Smectite	Beidellite	$M_{0.25}^{II}Si_{3.5}Al_{2.5}$ $O_{10}(OH)_2$	Nontronite, hisingerite[a], saponite	15–160 Interlayer ~800 (all smectites)	45–160 (all smectites)	Often occur as ferribeidellites in soils, but heterogeneous	More beidellitic in acid-leached horizons; otherwise where drainage retarded and pH high
	Montmorillonite	$M_{0.25}^{II}Si_4Al_{1.5}Mg_{0.5}O_{10}(OH)_2$	Stevensite, hectorite				
	(Palagonite)	Prob. variable		Unknown	Unknown	Nanocrystalline precursors of smectites	Early weathering products of basaltic volcanic glass

Hormite	Palygorskite	$Si_8Mg_5O_{20}(OH)_2 (OH_2)_4 \cdot 4H_2O$	Attapulgite[a]	140–190	3–30	Fibrous	In dry, usually calcareous regions, near textural transition in profile (both)
	Sepiolite	$Si_{12}Mg_8O_{30}(OH)_2 (OH_2)_4 \cdot 8H_2O$		Generally 260–330	20–45	Fibrous	
Al–Si	Imogolite	$SiO_2 \cdot Al_2O_3 \cdot 2H_2O$; Si tetrahedra within Al octahedral in tube		1500	pH-variable	Nanocrystalline: bundles of very thin hollow threads or tubules (nanotubes)	Limited, mainly from pumice, also podzol illuvial B horizons and related soil horizons
	Allophanes	Al:Si 2:1—1:1; variable between imogolite and allophane		700–1500	pH-variable	Nanocrystalline: very small spherules (nanospheres or nanoballs); closely associated with organic matter	From tephra (including volcanic ash) and in podzol illuvial B horizons, and in some other materials that are strongly leached
Nonsilicates							
Al hydroxide, oxyhydroxide, oxide	Gibbsite	$Al(OH)_3$	Bayerite, nordstrandite, doyleite, α, or γ alumina trihydrate	Unknown	pH-variable	Hexagonal crystals	Where Si low, especially in strongly leached soils
	Boehmite	$\alpha AlOOH$	Diaspore, α, or γ alumina monohydrate	Unknown, prob. high	pH-variable		Strongly weathered soil materials or soils, ferricrete (laterite), bauxite
Fe oxyhydroxide, oxide	Goethite	$\alpha FeOOH$	Limonite[a]	14–77	pH-variable	Yellow to yellow-brown color	Widespread—most common iron "oxide"
	Lepidocrocite	$\gamma FeOOH$	Akagenéite	Unknown, prob. high	pH-variable	Orange color	Noncalcareous soils that are seasonally anaerobic
	Hematite	αFe_2O_3		35–45	pH-variable	Bright red color	Soils of warmer climates
	Maghemite	γFe_2O_3	Titanomagnetite	Unknown, prob. high	pH-variable	Ferrimagnetic	Tropical and subtropical soils, from fires, poss. from bacteria
	Ferrihydrite	$Fe_5HO_8 \cdot 4H_2O$	Feroxyhite	200–500	pH-variable	Nanocrystalline: spherical nanoparticles, pinkish to yellowish-red color	Widespread, where Fe(II) is oxidized rapidly
Mn oxide	Birnessite	$Na_{0.7}Ca_{0.3}Mn_7O_{14} \cdot 2.8H_2O$	Todorokite, hollandite, lithiophorite, pyrolusite	Unknown, prob. high	pH-variable	(Blue-) black color	Widespread but usually minor component; in concretions, segregations, pans and nodules in subsoils
Sulfide	Pyrite	FeS_2	Mackinawite, greigite, pyrrhotite	Unknown	Unknown		Coastal regions and from some sediments when unoxidized
Sulfate	Gypsum	$CaSO_4$	Bassanite, anhydrite, barite ($BaSO_4$)	Unknown	Unknown		Often in desert soils
	Jarosite	$KFe_3(OH)_6(SO_4)_2$	Natrojaroite, schwertmannite	Unknown	Unknown		Acid sulfate soils or strongly acid seepages
Phosphate	Plumbogummite	$PbAl_3(PO_4)_2(OH)_5 \cdot 2H_2O$	Variscite (Al), strengite (Fe), crandallite (Ca), gorceixite (Ba)	Unknown	Unknown		Rare, from breakdown of rock phosphate
Ti oxide	Anatase	TiO_2		Unknown	Unknown		Widespread, in small amounts
Chloride	Halite	$NaCl$		Unknown	Unknown		Seasonally dry saline soils

[a] Discredited name.

20.4 Influence of Mode of Formation upon Predictions of Properties of Soils from Their Clay Mineralogy

20.4.1 Introduction: The Role of Mineralogy in Soil Science

Broadly speaking, soils are studied and interpreted either as products of the natural environment, which reflect their environmental history, or else as useful materials in which to grow plants or for removing and cleaning wastes and pollutants or providing resources or "environmental services" such as regulating water flow. Minerals, and particularly secondary (clay) minerals, have likewise been regarded and assessed either as indicators of the origin of soils and of changes, which have occurred in them through their development, or else as the most reactive inorganic components of soils. For the latter role, it has often been considered that knowledge of the properties of clay minerals would be useful in predicting the agronomic and adsorptive capacities of soils. However, several indicators have suggested that soil mineralogy has come to be regarded as less important to practitioners of the discipline of soil science in recent years than its other main subdisciplines such as soil chemistry, soil physics, soil biology, pedology, spatial analysis, and pedometrics, let alone the major applied aspects of soil fertility and soil pollution, than once was the case. For example, an analysis of papers that were published in the first 100 volumes of *Geoderma*, spanning from 1967 to 2001, showed a sharp decline in those on soil mineralogy, but either a rise or no significant decline in those from the other main subdivisions of soil science (Hartemink et al., 2001). It appears that mineralogical studies have been able to make fewer worthwhile contributions in comparison with those of other aspects of soil science than in earlier times. Nonetheless, the survey of the literature for this chapter suggests that a great deal of research continues to be undertaken into the origin of secondary minerals in soils and also, to some extent, into their mode of occurrence within soils. This new work includes the recognition that the clay fractions of many soils and associated materials contain nanominerals (including allophanes and ferrihydrite as discussed earlier), which confer unique properties and reactivities (Theng and Yuan, 2008; Waychunas and Zhang, 2008). Nanoscale minerals are defined as having at least one dimension in the nanorange, that is, 1–100 nm (Hochella, 2008). Nanoscience is now regarded as being a critically important new offshoot of more traditional colloid or clay mineralogy, and the growth rate of articles being published on nanoparticle science (relevant to the geosciences) is around 10% per year, about triple the average growth rates of all scientific disciplines over a 5-year period (Hochella, 2008).

20.4.2 Contributions of Classical Clay Mineralogy toward Explanations of Soil Properties

Clay mineralogy is a relatively young area of study. Up until 1929, with the publication of a pioneering bulletin by Hendricks and Fry on the X-ray examination of soil clays that was quickly followed by a journal paper (Hendricks and Fry, 1930), most had regarded clays, which were largely indistinct in optical microscopes, as being amorphous (Cady and Flach, 1997). XRD was very successful in determining that much of the clay-sized inorganic material in soils was composed of regular crystals like those that had been identified in minerals in rocks or other geological deposits. Although many other instrumental techniques have been applied to their identification and characterization over the intervening years, the search for and refinement of crystal structures has remained a major pursuit of clay mineralogists and XRD has continued to be their major workhorse. Its use, whether for structural determination or as the major instrument for the identification of clay-sized minerals in soils, has defined classical clay mineralogy.

Among the most useful roles promised for classical clay mineralogy in soil studies has been that of explaining the abundance of certain plant nutrients by soils. Probably chief among these has been potassium, with K-micas apparently providing a labile source of K^+ (e.g., Norrish, 1973; Loveland, 1984). Loveland et al. (1999), in writing a history of clay mineralogy at the Rothamsted Experimental Station from 1934 to 1988, attempted to identify the reasons why considerable effort was put into this area of soil science at one of the world's leading centers of soil research. At its outset in 1934, clay mineralogy was supposed to provide an understanding of the physicochemical behavior of the soil clay fraction mainly with respect to "the sorption and desorption of water and nutrients, with their practical consequence for soil workability and plant nutrient supply" (Loveland et al., 1999, p. 165). Its most important contributions are judged by these authors to have been in the identification of the phenomenon of interstratified minerals and the behavior of soil K. Even so, in spite of a large number of British studies, at Rothamsted and elsewhere in the United Kingdom, on the influence of clay mineralogy on the potassium-supplying power of soils, Loveland (1984, p. 700) stated that "none were successful in using clay mineralogy as a reasonably exact predictive tool for this property." Loveland (1984) concluded that (classical) clay mineralogy related better to geotechnical properties of soils than to their nutrient-supplying power. In apparent contrast, the K-supplying power of soils formed from quartzo-feldpathic loess or alluvium and colluvium from sedimentary rocks in New Zealand showed a close relationship to the mica contents of their clay fractions (Surapaneni et al., 2002). Barré et al. (2008), studying two French soils and one from the United States, also found that micaceous minerals in soil clay-size fractions serve as reservoirs for potassium. However, they found that the ability of soils to supply K to plants, and also to extract K when it was in excess, through fertilizer additions, could only be fully understood when the micaceous minerals were delineated into five different types. These were a well-crystallized illite, a poorly crystallized illite, a highly smectitic interstratification of illite and smectite and its highly illitic counterpart, and a soil vermiculite. These could be separately determined quantitatively, relative to one another, by decomposition of the XRD peak profile, following Lanson's (1997) procedure. Barré et al. (2008) claimed that the approach

could be used to predict the K release potential of soils generally and Velde and Barré (2010) provide many examples of the utility of this approach. It may offer a wider applicability than the relationships developed by Surapaneni et al. (2002) for soils, which have quite similar origins in just one country.

20.4.3 Potential of Classical Clay Mineralogy for Explaining Soil Properties

A different mineralogical approach using X-ray fluorescence (XRF) analyses of whole soils and also analyses of acid digests of soils by inductively coupled plasma mass spectroscopy (ICP-MS) was used to determine the normative mineralogies of forest soils in Finland at the early stages of weathering in order to explore the availability of reserves of the nutrients Ca and Mg (Starr and Lindroos, 2006). At the earliest stages of weathering, it is the extent of alteration of primary minerals rather than the nature of any secondary products that is the mineralogical characteristic that best relates to agronomic properties. Clay mineralogy per se is of little predictive use when clays have hardly been formed. For soils in most of the world's agricultural areas, however, the overwhelming contribution to the surface area of the soils, as well as virtually all of their charge, arises from material in their clay fraction (e.g., Gilkes, 1990; Churchman, 2006). The highly reactive clay-size material includes both organic and inorganic components, and it is the task of soil clay mineralogists to understand the source of the contribution made by the inorganic components. Gilkes (1990) noted that Norrish and coworkers recognized some time ago that secondary minerals in soils could hold and, when required, supply other important nutrients besides K for plants. These included manganese oxides, for cobalt (Adams et al., 1969), and iron oxides, for phosphorus (Norrish and Rosser, 1983). Classical clay mineralogy, using XRD, may be capable of detecting the presence of these oxides in soils provided that they are reasonably well crystallized, and that their XRD peaks are not masked by stronger peaks for other minerals. Even so, it is not simply their presence, but rather their detailed structural, chemical, and physical characteristics, as well as their associations with other soil entities that govern the reactivities—both chemical and physical—of these and other oxides, hydroxides, and oxyhydroxides, as well as of aluminosilicate phases in soils. Furthermore, much evidence points to the idea that metal oxides (used generally hereafter to include both oxyhydroxides and hydroxides) play a role in determining many of the useful properties of many soils that is out of proportion to their content in the soils. This status includes the indication from several earlier studies, van Raij and Peech (1972) and also Gallez et al. (1976), among them, that the surface charge–pH curves of many tropical soils show the characteristic features of metallic oxides, which have a pH-dependent variable charge. Similarly, nanocrystalline aluminosilicates such as allophane also make an extremely disproportionate contribution to both the surface area and reactivity of soils, including through a pH-dependent variable charge, even where only small amounts are present (e.g., Lowe, 1995; McDaniel et al., Section 33.3). In an effort to try to

reconcile studies that have found that Fe oxides play a positive role in aggregation in soils (e.g., Colombo and Torrent, 1991; Oades and Waters, 1991) with others in which Fe oxides do not appear to affect aggregate stabilities (e.g., Deshpande et al., 1968; Borggaard, 1983), Duiker et al. (2003) found that their effect depended upon their crystallinity. The most poorly crystalline Fe oxides played an important role along with organic matter in aggregate stabilization and, where organic matter contents were low, as in B horizons, their contributions to indices of aggregation were found to equal or exceed that of organic matter (Duiker et al., 2003), confirming similar trends seen by Churchman and Tate (1987) for the particular case of soils formed on volcanic ash. However, Al oxides may be more important than Fe oxides in governing both aggregate stability and also P adsorption in some soils. This relationship appeared to hold for Latosols in Brazil (Schaefer et al., 2004). In one soil, at least, an amorphous Al-hydroxy phase appeared to play an important role in both aggregation and P adsorption.

One property of soils that particularly demands explanation and, if possible, prediction, at this time is that of the capacity of soils to adsorb and retain organic carbon. The momentum for work to be done in this area comes from growing community concerns about the increased levels of carbon dioxide in the atmosphere, its links to global warming, and the role that soils may be able to play in sequestering excess carbon and isolating it from the atmosphere (Amundson, 2001; Kahle et al., 2002a). As Kahle et al. (2002a) pointed out, many workers have discovered relationships between clay content and carbon content, although it is noteworthy that not all have found such a link, for reasons we will discuss herein. Nevertheless, it has long been known (e.g., Greenland, 1965) that clays and organic matter can be strongly associated in soils. Even so, it has long been appreciated (e.g., Hamblin and Greenland, 1977) that the links between clays and organic matter can occur through Fe and Al oxides. Mikutta et al. (2005) and also Kleber et al. (2005) reported that for acid subsoils under forest and representing a wide range of parent rock types and rainfall, poorly crystalline oxides of both Fe and Al provided the strongest binding agents for organic matter. According to Mikutta et al. (2005), poorly crystalline to nanocrystalline Fe oxides such as ferrihydrite were particularly effective for stabilizing organic C in these soils. It is also significant that these authors found that none of the factors of clay content, namely, the specific surface area of the minerals as determined by nitrogen gas adsorption using the BET equation, total iron oxide content, or the nature of the phyllosilicate clays, was a good predictor of the ability of the soils to stabilize C. Their results confirmed those of Percival et al. (2000) who correlated the organic C contents of 167 New Zealand soils with several soil properties and found they showed significant linear relationships with oxalate-extractable Si and Al, which together represent allophane content, and also with pyrophosphate-extractable Al, which represents Al associated with organic matter, that is, "Al-humus complexes," but found no relationship with total clay content. Rasmussen et al. (2005), also studying forest soils, came to similar conclusions as those of Percival et al. (2000). In a different,

tropical environment, but again studying forest soils on volcanic ash, Basile-Doelsch et al. (2005) found that organic C in buried horizons was strongly bound in abundant amounts in soils containing allophane, whereas buried horizons with large amounts of crystalline minerals, including feldspars and gibbsite, had little capacity for C storage. The studies by Mikutta et al. (2005) and also Kleber et al. (2005), which were on samples devoid of allophane, showed that it is not the particular occurrence of allophane or even of ferrihydrite that is necessary to provide strong links to organic C. Rather it is the capacity of poorly crystalline or nanocrystalline minerals of both Fe and Al to provide singly coordinated hydroxyl groups that are able to participate in ligand exchange with organic functional groups, which ensures strong associations with organic C, according to Kleber and coworkers in the two related 2005 reports. This generalization also bears out the results of experiments by Kahle et al. (2004) in which dissolved organic carbon (DOC) was added to soil clays that had been depleted of organic C and also to "type" phyllosilicates from nonsoil sources. This experiment showed that soil clays adsorbed much more DOC than the type clay minerals and that the difference appeared to be due to pedogenic oxides, mainly of Fe, which provide reactive hydroxyl groups for DOC sorption. In the case of some tropical soils, Barthès et al. (2008) found that organic C content and also aggregate stability were promoted when sesquioxides, especially goethite and hematite, in the soils included substitutions by Al, which increased their specific surface areas.

One of the common characteristics of Andisols is accumulation of relatively large quantities of organic matter, both in the allophanic (moderate pH) and in the nonallophanic Andisols (low pH). Allophanic Andisols may contain ~10% or more C, whereas nonallophanic andisols (dominated by Al- and/or Fe-humus

complexes) may contain as much as ~30% of C. However, the residence time of C in soils containing allophane, as measured by ^{14}C, is much greater than that of other soils (Figure 20.12; also see Parfitt, 2009).

Albeit that the studies by Kleber and coworkers on acid subsoils showed no relationship between stabilized organic C and the surface area of minerals in the soils, extent of surface intuitively appears likely to be able to explain the capacity of soils for organic C, and several workers have proposed such a relationship (e.g., Saggar et al., 1996; Kaiser and Guggenberger, 2000; Eusterhues et al., 2005). Indeed, to illustrate the point that different soils lead to different generalizations about the relationship between C content and surface area, and, incidentally, forms of Fe in soils, Kahle et al. (2002b) found that a combination of specific surface area and total (i.e., dithionite-extractable) Fe oxide content predicted C content almost completely, with an R^2 of 0.96, in the topsoils of a set of illitic soils with more or less neutral pHs that had been farmed since Medieval times. The surface areas were determined using the sorption of nonreactive nitrogen gas and hence gave a good measure of external surface (e.g., Churchman and Burke, 1991). Using isotopically labeled plant material, Saggar et al. (1996) found that its residence time during its initial decomposition in soils correlated well with the surface area of the soil as measured by the adsorption of p-nitrophenol. It is noteworthy that it was this particular (p-nitrophenol adsorption) method that showed a good relationship with organic C retained in this case. Like Churchman and Burke (1991), these authors also tried to explain soil properties (organic C uptake in this instance) by the sum of the surface areas from the literature for the mineral components of the soils and also largely failed. By contrast, consideration of the specific surface areas (SSAs) from N_2 sorption of different soil types in relation to those calculated
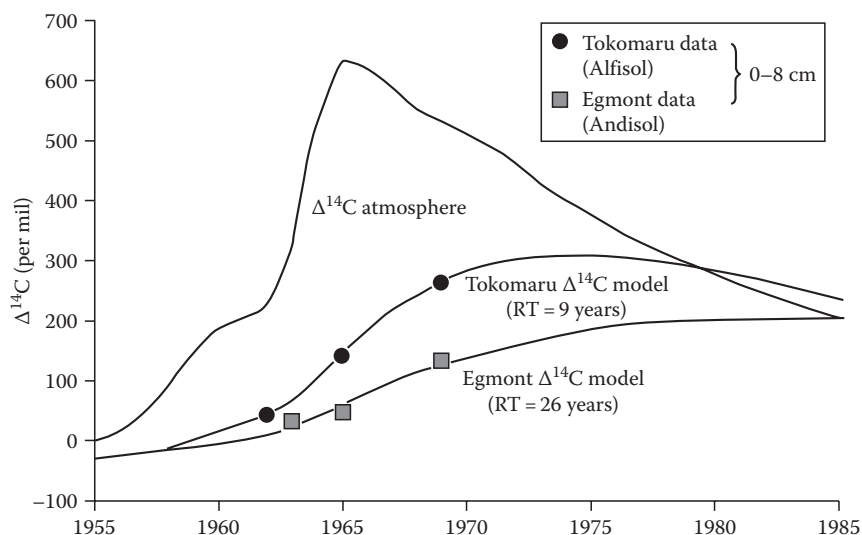
FIGURE 20.12 Model showing slower rates of incorporation and turnover of organic C using bomb-derived ^{14}C in an allophanic soil (Egmont series) compared with those of a soil without appreciable allophane (Tokomaru series) under similar climate and land use in New Zealand. (Based on Baisden, W.T., R.L. Parfitt, and C.W. Ross. 2010. Radiocarbon evidence for contrasting soil carbon dynamics in an Andisol and non-Andisol pasture soil comparison. J. Integr. Field Sci. 7:59–64.) RT, residence time of carbon. (Modified diagram from Parfitt, R.L. 2009. Allophane and imogolite: Role in soil biogeochemical processes. Clay Miner. 44:135–155. With permission of The Mineralogical Society of Great Britain and Ireland.)

for their constituent Fe oxides and allophane led Eusterhues et al. (2005) to conclude that almost all mineral-associated organic matter in soils is bound to Fe oxides. Churchman and Burke (1991) generally found that the strength of the relationship between a soil property and surface area depended upon the technique chosen to measure surface area (they compared N_2 gas sorption, with sorption of polar liquids ethylene glycol monoethyl ether [EGME] and water) and the relevance of the adsorbate to the soil property of interest. A good relationship between a particular soil property and surface area deduced from the sorption of a particular adsorbate shows that the property and the capacity of the surface for the adsorbate have similar causes but does not show that the property relates to a generic measure of surface area, which most probably does not exist.

To further complicate the nature of the relationship between minerals and organic C in soils, it is noted that a study of the energetics of organic C sorption by Mayer and Xing (2001) revealed that minerals appeared to be occluded by OM, rather than the OM being adsorbed by minerals, and that micropores were likely to be responsible for some of the associations. Kaiser and Guggenberger (2003) suggested that OM was preferentially adsorbed within, or at the mouths of, micropores <2 nm in diameter in iron oxides in soils. The results of density fractionations combined with TEM led Chenu and Plante (2006) to conclude that most OM was stabilized in soils by close associations with clays in very small microaggregates, either through adsorption or by entrapment. Wan et al. (2007) mapped organic carbon along with other elements in different soils using scanning transmission X-ray microscopy (STXM) and found that OM exists as distinct particles within microaggregates more often than as coatings on minerals. Using small-angle X-ray scattering to directly observe pores and their constituents, McCarthy et al. (2008) found that most OM was held within pores, although not necessarily very small pores, in formerly cultivated soils that had been restored to tallgrass prairie. Their main conclusion was that OM was encapsulated, rather than adsorbed, by minerals. According to Richards et al. (2009), it is the occurrence of aggregates, rather than the abundance of oxides of Fe and Al, that enabled the stabilization of OM in oxide-rich soils in the Australian subtropics. As a result, more organic C was

held in soils under rainforest and pastures than in those under pine plantations. Furthermore, Spielvogel et al. (2008), like many others (e.g., Kleber et al., 2005; Mikutta et al., 2005; Rasmussen et al., 2005), examined the causes of OM stabilization in acidic forest soils, but arrived at the conclusion that, while Fe and Al oxides, and, particularly, their poorly crystalline to nanocrystalline varieties, were responsible for stabilization, they are bound to, and stabilize, only a specific fraction of the OM. This fraction was concentrated in O/N-alkyl C and microbially derived sugars compared with the nonstabilized remainder (Spielvogel et al., 2008).

It, therefore, appears likely that the properties of the mineral fraction, which control the uptake and retention of organic C by soils, vary greatly between soils. Any generalization is likely to founder when applied to a particular case. Indeed, following Kaiser and Guggenberger (2003), Chenu and Plante (2006), Wan et al. (2007), McCarthy et al. (2008), and Richards et al. (2009), we suggest that it may be that the properties of minerals that are most important are those that enable them to form pores or aggregates for encapsulating OM, and their identification probably defies a simple summary. At the most, it can probably be concluded only that a number of possible characteristics of the mineral phases are likely to influence the capacity of soils to stabilize C and it remains for practice to show which are more important than the others in each instance. A possible list of such characteristics is given in Table 20.7. A number of studies have attempted, like Churchman and Burke (1991) and Saggar et al. (1996) with specific surface area, to explain soil properties by the sum of the contributions from their different mineral (and also, sometimes, organic) components. Properties that were proposed for explanation in this way have included CEC (Seybold et al., 2005). CECs could be largely explained within mineralogical groupings from *Soil Taxonomy* when organic C was included. Churchman and Burke (1991) also found good relationships between CECs and a measure of water retention within mineralogical groupings of soils with low carbon contents.

Together, these results suggested that the use of classical clay mineralogy to group mineralogically similar soils may be reasonably successful for predicting CECs. In addition, close comparisons of experimental and theoretical titration curves indicated

TABLE 20.7 Characteristics of Mineral Phases That Have Helped to Explain the Capacity of Various Soils for the Uptake and Retention of Organic Carbon

Most Relevant Mineral Phase and/or Property	Soil	References
Fe and Al oxides	Silty soils	Hamblin and Greenland (1977)
Allophane + pyrophosphate–Al	167 soils (from throughout New Zealand)	Percival et al. (2000)
Specific surface area + total Fe oxides	Neutral illitic topsoils	Kahle et al. (2002b)
Pedogenic oxides, mainly of Fe	Range of forest and arable soils with range of mineralogy	Kahle et al. (2004)
Fe oxides	Cambisol and podzol under forest (acid)	Eusterhues et al. (2005)
Poorly crystalline (or nanocrystalline) Fe and Al oxides, especially ferrihydrite	Acid subsoils under forest	Kleber et al. (2005) and Mikutta et al. (2005)
Allophane	Volcanic ash soils under forest	Basile-Doelsch et al. (2005)
Al-substituted Fe oxides	Tropical soils	Barthès et al. (2008)
Fe and Al oxides	Acidic forest soils	Spielvogel et al. (2008)

that surface charge densities of some soils from Argentina were apparently successfully modeled by the sum of those of their constituent mineral phases when the mineralogical compositions of the soils were determined by a Rietveld approach (Taubaso et al., 2004). This result suggests further that the charges on soils may be largely additive of those on their constituent clay-size minerals. Nonetheless, this and earlier work by a group from the same laboratory (Torres Sanchez et al., 2001) showed that there was a mismatch between the point of zero net charge (PZNC) and the isoelectric point for those soils in which Fe oxides were adsorbed on to phyllosilicate clays. Hence, the diffusion behavior of these soils, which relates to their isoelectric points (IEPs), cannot be simply regarded as additive of their component colloidal minerals but must account also for their associations with one another.

While there has been some success with the modeling of charge characteristics, both permanent (Seybold et al., 2005) and variable (Taubaso et al., 2004) from the mineral composition of soils, mineral composition patently could not describe organic matter stabilization by minerals in soils. However, in some cases, at least (Kahle et al., 2004; Kleber et al., 2005; Mikutta et al., 2005; see also Table 20.7), it appears that some functional groups, singly coordinated hydroxyl groups in particular, play a crucial role in binding natural organic compounds—albeit perhaps only a fraction of the total soil organic matter, following Spielvogel et al. (2008). Therefore, the question arises as to whether more accurate descriptions of mineral–organic associations, and of reactions of other species, for example, anions such as phosphates and arsenates with mineral surfaces, might be better made on the basis of the amount of the important functional groups present, by an approach that is analogous to that taken by modern soil organic matter studies using, for example, ^{13}C-NMR and FTIR spectral analyses (e.g., Baldock and Nelson, 2000).

Gustafsson (2001) used a "component additivity" approach in an attempt to predict, or model the adsorption of arsenate in competition with sulfate, silicic acid, and phosphate on allophane and also ferrihydrite, and tested the results against experimental data for the adsorption on the spodic horizon of a Swedish soil that contained both ferrihydrite and allophane but little organic matter. Although this approach led to results that were qualitatively realistic, Gustafsson (2001) nevertheless concluded that the approach is probably impractical, partly, at least, because it relies on the assumption that the properties of allophane and ferrihydrite in real soils can be approximated by those of gibbsite and ferrihydrite synthesized in the laboratory. Among many others, Gérard et al. (2007) have shown that this assumption is simplistic and unrealistic. In the particular case that these last authors studied, allophane formed on basalt in the Azores was found to have an especially complex genesis, leading to allophanic products with a large range of compositions. Gérard et al. (2007) claimed that not all of the allophane formed was the product of recrystallization from solution, describing some of it as "alteromorphs after lapilli or pumice," which result from the leaching of Si and cations from these materials. The products encompassed a compositional range from pure aluminosilicates to Fe- and Ti-enriched aluminosilicates and included

varieties of both Al-rich and Si-rich allophanes, while the soils also contained many different iron phases, ferrihydrite, hematite, and also "iddingsite" (see Section 20.2.4.1), among them.

The component additivity modeling approach adopted by Gustafsson (2001) foundered because allophanes and ferrihydrite, as they are found in soils, differ in many important respects from either natural minerals from nonsoil environments such as the "stream deposit allophane" (Childs et al., 1990; Parfitt, 1990) or synthetic minerals of the same name, in the case of ferrihydrite. They are unique to soils, and they also show variations within their type. A number of other minerals are unique, or almost unique, to soil environments, and the properties and reactivities of these, also, cannot be modeled upon those of minerals found in nonsoil environments. These include fully or partially expanded 2:1 Si:Al aluminosilicates with hydroxyl-Al interlayers (see Section 20.2.4.3), which may show even seasonal changes in the soil biological system (see Section 20.2.5.3). They also include interstratified phases in their various manifestations. Velde and Meunier (2008) have characterized those of illite and smectite as being "peculiar to the soil environment" (Section 20.2.5.3.1), as well as being "the most widespread and representative soils clays." By the same token, smectites found in soils have been characterized as being commonly ferribeidellitic in composition and, therefore, are quite different from the bentonites from which typical nonsoil montmorillonites have usually been extracted to serve as models for smectites in soils (Section 20.3.5).

However, it is not just the particular composition, structure, and arrangement of layers of the minerals in soils that often make them different from so-called typical minerals taken from nonsoil sources. They differ from the latter partly because important physical properties of secondary minerals in soils are generally different from those of nonsoil minerals. Studies of soil kaolinites from soils on three continents by Singh, Gilkes, and coworkers (Section 20.3.1), a number of studies of soil halloysites (Section 20.3.2), and a study of iron oxides by Trakoonyingcharoen et al. (2006) (Section 20.3.7), have focused specifically on the sizes and sometimes also surface areas of the particles. The studies generally concluded that those of the soil particles are almost invariably much smaller than those of their "reference" nonsoil counterparts with the same mineral names.

Soil clay minerals also differ from most, at least, of their nonsoil counterparts because of their associations with other species in soils. Apart from their ready association with organic matter that has already been discussed in this section, they are very commonly associated also with other minerals. Almost universally, they are associated with iron oxides (see Table 20.5 and Section 20.3.7) and, often, also with aluminum oxides. While the content of Fe in its various forms varies widely (Table 20.5), it is probably only in very young soils that it is nearly negligible in affecting soil properties. The youngest soil in a chronosequence on andesitic lava studied by Nieuwenhuyse et al. (2000) contained only 0.4% total Fe in its lowest, CB horizon. This horizon also contained 2.2% oxalate-extractable Al and, hence, ~8% allophane, but no phyllosilicate minerals (Nieuwenhuyse et al., 2000).

Table 20.5 also shows that soils are almost certainly never mono-mineralic in their secondary minerals. Only one of the soils shown there (an Orthoxic Tropohumult) is apparently monomineralic in phyllosilicates (kaolinite, in this case), but clearly also contains goethite, at least, in addition.

20.4.4 Nature of Soil Minerals and Relationship to Their Mode of Formation

Clearly, properties of soils relate to properties of clays besides those arising from their particular mineralogical characterization. Properties such as particle size and associations that minerals form with other minerals derive from their mode of formation. In general terms, minerals with a neogenetic origin from solutions in the soil occur in smaller particles than those formed by either transformation in the solid phase from primary minerals or even those that may be inherited from a source of minerals of authigenic origin from beyond the soil environment, for example, hydrothermally formed kaolinite and halloysite, or smectite in bentonite deposits of marine or lacustrine origins. A dramatic illustration of the differences that are wrought by pedogenesis in comparison with a nonsoil genesis (in this case, saprolitic) is given by the observation that the area of the 001 faces of kaolinite crystals that formed in the saprolite below 58 m depth in Brazil were more than five times larger than those that formed in the soil above it (Varajão et al., 2001). It is recalled that Kahle et al. (2002b) found that a combination of specific surface area and total (i.e., dithionite-extractable) Fe oxide content predicted C content almost completely in a set of illitic soils whereas explanations of C contents of other soils from mineral properties were more complex and usually involved poorly crystalline or nanocrystalline phases (see also Section 20.4.3). Even though the land-use history of the illitic soils and their pH may have played a role in simplifying the relationships between the inorganic (mineral) and organic components of the illitic soils (and, incidentally, in altering the course of mineral genesis), it is also most likely that the probable formation of the illite in the soils by transformation rather than neogenesis has led to mineral–organic relationships that are different from those in soils with higher proportions of neogenetic minerals, including poorly crystalline/nanocrystalline metal oxides (e.g., Kleber et al., 2005; Mikutta et al., 2005). Furthermore, it has been found that the nature of aggregation within soils dominated by kaolinites in close associations with Fe and Al oxides differs from that in soils that are more illitic. Both in a study in which aggregates were broken down mechanically (Oades and Waters, 1991) and another in which they were given the opportunity to build up through plant growth and inputs of plant residues (Denef and Six, 2005), aggregates in the kaolinitic soils were shown to comprise largely self-associations of minerals, while those in the illitic soils included close associations of minerals with organic matter.

The application of analytical techniques with higher resolution than XRD, and particularly electron microscopy, has shown that (1) "many clay particles in soils consist of complex inter-growths of different structural types rather than being the ideal monomineralic species described in textbooks" (Gilkes, 1990, p. 72); and (2) the extent and nature (e.g., charge) of two types of kaolinites (in different Australian soils) were quite different (op. cit., p. 67). Soils are highly heterogeneous in composition and are not simply mixtures of nonassociated entities. In particular, clay minerals are commonly coated by other minerals or organic matter, or both, so that the properties of their reactive surfaces reflect the characteristics of the coating material rather than the mineral substrate. For instance, it was found that "without exception, the suspended particles in rivers and gesturing waters were negatively and quite uniformly charged" (R.J. Hunter, 1981, in Mills, 2003, p. 11). It is almost certain that these particles had a soil source. Furthermore, "Since the particles themselves varied widely as to composition, Hunter concluded that this was likely due to a coating of organic matter or metal (iron, aluminum, and/or manganese) oxide" (Mills, 2003, p. 11).

These observations all suggest that the colloidal "particles" controlling soil properties are generally heterogeneous mixtures of various types of materials, both inorganic and organic. Predictions of their properties from those of well-crystallized minerals are likely to fall short of the mark. Grouping of soils by their dominant idealized mineral types can generally give only poor, or, at best, only qualitative, predictions of their properties. A structural alternative of grouping soils by the nature of their associations of, for example, particular aluminosilicate minerals, metal oxides, and organic matter together would provide the hopeless case of too many categories; indeed, each association would be very likely unique. A better alternative is suggested by the example of research into the nature of purely organic colloids in soils.

Organic colloids in soils are alternatively known as "humic macromolecules." Hayes et al. (1989, p. 16) have concluded, on theoretical grounds, that "it is highly unlikely that there are two humic macromolecules on earth … which are exactly the same." Even so, the field of the chemical and analytical study of organic colloids in soils is a very active one. It has had continuing success because instead of seeking definitive general structures of inevitably heterogeneous molecules, it has sought instead to characterize organic colloids by their "functional groups" (e.g., aromatic, carbohydrate, alkyl, carboxyl) —in other words, by their capacities to participate in reactions with other entities in soils.

Since a structural characterization of the "ultimate" particles or groupings of minerals would be similarly fraught with difficulties as that of organic colloids per se, it may be useful to classify soil minerals as contributors to soil properties (mostly via complex heterogeneous mixtures) through their characterization and appropriate grouping according to their most relevant capacities. These might include (on a weight basis) their surface area—as measured by various techniques, their charge and its variability with pH, their content of plant-available potassium, their shrink–swell capabilities, and their concentrations of reactive hydroxyl groups. Together, these properties reflect many of the most important for soil applications and may be correlated with other properties, for example, their affinity with organic matter and their capacity to adsorb phosphates or chlorides and other anions.

Nevertheless, selection of a set of useful properties by which to characterize mineral properties does not provide advances in understanding the origin of these properties. As has always been the case in mineralogical studies, this uncertainty will be reduced by advances in new instrumental techniques and in their application to soils. Undoubtedly, electron microscopy at higher resolutions, both for viewing and also for quantitative analyses, will be among these. Such techniques of surface analysis as XPS might also help. For example, Gerin et al. (2003) were able to analyze the nature of soil surface coatings, both of organic matter and also Fe and Al species, in order to better understand mineral-organic associations, and they projected more similar work with XPS. With the advent of synchrotron radiation, X-ray spectroscopic techniques, which include and enhance XPS and also XRD and XRF, but also encompass extended X-ray absorption fine structure (EXAFS), X-ray absorption near-edge structure (XANES), and other related methods have become available to determine the sites on mineral surfaces that are involved in interactions with other species (Gates, 2006).

Continuing developments in instrumental techniques undoubtedly enable better descriptions of processes involving minerals in soils, but the foregoing discussion nonetheless leads us to conclude that in order to be able to provide the best predictions or explanations of soil behavior from its clay mineralogy, the origins of the clays themselves must be thoroughly understood. The earlier parts of this chapter on the alteration, formation, and occurrence of minerals in soils thereby gains relevance to the understanding of soils both as useful materials and providers of essential ecological and human services.

Acknowledgments

References

Adamo, P., C. Colombo, and P. Violante. 1997. Iron oxides and hydroxides in the weathering interface between *Stereocaulon vesuvianum* and volcanic rock. Clay Miner. 32:453–461.

Adamo, P., and P. Violante. 2000. Weathering of rocks and neogenesis of minerals associated with lichen activity. Appl. Clay Sci. 16:229–256.

Adamo, P., P. Violante, and M.J. Wilson. 2001. Tubular and spheroidal halloysite in pyroclastic deposits of the Roccamonfina volcano (Southern Italy). Geoderma 99:295–316.

Adams, S.N., J.L. Honeysett, K.G. Tiller, and K. Norrish. 1969. Factors controlling the increase of cobalt in plants following the addition of a cobalt fertilizer. Aust. J. Soil Res. 7:29–42.

Ahn, J.-H., and D.R. Peacor. 1987. Kaolinization of biotite—TEM data and implications for an alteration mechanism. Am. Mineral. 72:353–356.

Aldridge, L.P., and G.J. Churchman. 1991. The role of iron in the weathering of a climosequence of soils derived from schist. Aust. J. Soil Res. 29:387–398.

Alekseev, A., T. Alekseeva, V. Ostroumov, C. Siegert, and B. Gradusov. 2005. Mineral transformations in permafrost-affected soils, North Kolyma lowland, Russia. Soil Sci. Soc. Am. J. 67:596–605.

Alexandre, A., J.-D. Meunier, F. Colin, and J.-M. Koud. 1997. Plant impact on the biogeochemical cycle of silicon and related weathering processes. Geochim. Cosmochim. Acta 61:677–682.

Alimova, A., A. Katz, N. Steiner, E. Rudolph, H. Wei, J.C. Steiner, and P. Gottleib. 2009. Bacteria-clay interaction: Structural changes in smectite induced during biofilm formation. Clays Clay Miner. 57:205–212.

Allen, B.L., and B.F. Hajek. 1989. Mineral occurrence in soil environments, p. 199–278. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Alloway, B.V., M.S. McGlone, V.E. Neall, and C.G. Vucetich. 1992. The role of Egmont-sourced tephra in evaluating the paleoclimatic correspondence between the bio- and soil-stratigraphic records of central Taranaki, New Zealand. Quat. Int. 13–14:187–194.

Altschuler, Z.S., E.J. Dwornik, and H. Kramer. 1963. Transformation of montmorillonite to kaolinite during weathering. Science 141:148–152.

Amrhein, C., and D.L. Suarez. 1988. The use of a surface complexation model to describe the kinetics of ligand-promotion dissolution of anorthite. Geochim. Cosmochim. Acta 52:2785–2793.

Amundson, R. 2001. The carbon budget in soils. Annu. Rev. Earth Planet Sci. 29:535–562.

Anand, R.R., and R.J. Gilkes. 1984. Weathering of ilmenite in a lateritic pallid zone. Clays Clay Miner. 32:363–374.

Anand, R.R., and R.J. Gilkes. 1987. The association of maghemite and corundum in Darling Range laterites, Western Australia. Aust. J. Soil Res. 25:303–311.

Anand, R.R., R.J. Gilkes, T.M. Armitage, and J.W. Hillyer. 1985. Feldspar weathering in lateritic saprolite. Clays Clay Miner. 33:31–43.

Aomine, S., and C. Mizota. 1973. Distribution and genesis of imogolite in volcanic ash soils of northern Kanto, Japan, p. 207–213. *In* J.M. Serratosa (ed.) Proc. Int. Clay Conf. June 25–30, 1972. Madrid, Spain.

Aomine, S., and K. Wada. 1962. Differential weathering of volcanic ash and pumice resulting in formation of hydrated halloysite. Am. Mineral. 47:1024–1048.

Aoudjit, H., F. Elsass, D. Righi, and M. Robert. 1996. Mica weathering in acidic soils by analytical electron microscopy. Clay Miner. 31:319–332.

April, R., and D. Keller. 1990. Mineralogy of the rhizosphere in forest soils of the eastern United States. Biogeochemistry 9:1–18.

April, R.H., D. Keller, and C.T. Driscoll. 2004. Smectite in Spodosols from the Adirondack mountains of New York. Clay Miner. 39:99–113.

Aragoneses, F.J., and M.T. García-González. 1991. High-charge smectite in Spanish "Raña" soils. Clay Miner. 39:211–218.

Ashman, M.R., and G. Puri. 2002. Essential soil science. Blackwell, Oxford, U.K.

Aspandiar, M.F., and R.A. Eggleton. 2002a. Weathering of chlorite I: Reactions and products in microsystems controlled by primary minerals. Clays Clay Miner. 50:685–698.

Aspandiar, M.F., and R.A. Eggleton. 2002b. Weathering of chlorite II: Reactions and products in microsystems controlled by solution avenues. Clays Clay Miner. 50:699–709.

Bailey, S.W. 1990. Halloysite—A critical assessment. *In* V.C. Farmer and Y. Tardy (eds.) Proc. 9th Int. Clay Conf., 28 August–2 September 1989, Strasbourg, Germany, Vol. 2. Sci. Géologiq. 86:89–98.

Bain, D.C. 1977. The weathering of ferruginous chlorite in a podzol from Argyllshire, Scotland. Geoderma 17:193–208.

Bain, D.C., and J.D. Russell. 1981. Swelling minerals in a basalt and its weathering products from Morven, Scotland: II. Swelling chlorite. Clay Miner. 16:203–212.

Baisden, W.T., R.L. Parfitt, and C.W. Ross. 2010. Radiocarbon evidence for contrasting soil carbon dynamics in an Andisol and non-Andisol pasture soil comparison. J. Integr. Field Sci. 7:59–64.

Bakker, L., D.J. Lowe, and A.G. Jongmans. 1996. A micromorphological study of pedogenic processes in an evolutionary soil sequence formed on late Quaternary rhyolitic tephra deposits, North Island, New Zealand. Quat. Int. 34–36:249–261.

Baldock, J.A., and P.N. Nelson. 2000. Soil organic matter, p. B25–B84. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Banfield, J.F., and W.W. Barker. 1994. Direct observation of reactant–product interface formed in natural weathering of exsolved, defective amphibole to smectite: Evidence of episodic, isovolumetric reactions involving structural inheritance. Geochim. Cosmochim. Acta 58:1419–1429.

Banfield, J.F., and R.A. Eggleton. 1988. A transmission electron microscope study of biotite weathering. Clays Clay Miner. 36:46–70.

Banfield, J.F., and R.A. Eggleton. 1989. Apatite replacement and rare earth mobilization, fractionation and fixation during weathering. Clays Clay Miner. 37:113–127.

Banfield, J.F., G.G. Ferruzzi, W.H. Casey, and H.R. Westrich. 1995. HRTEM study comparing naturally and experimentally weathered pyroxenoids. Geochim. Cosmochim. Acta 59:19–31.

Banfield, J.F., B.J. Jones, and D.R. Veblen. 1991. An AEM-TEM study of weathering and diagenesis, Albert Lake, Oregon. Parts I and II. Geochim. Cosmochim. Acta 55:2781–2810.

Banfield, J.F., and T. Murakami. 1998. Atomic-resolution transmission electron microscope evidence for the mechanism by which chlorite weathers to 1:1 semi-regular chlorite–vermiculite. Am. Mineral. 83:348–357.

Barker, W.W., S.A. Welch, and J.F. Banfield. 1997. Biogeochemical weathering of silicate minerals. Rev. Miner. Geochem. 35:391–428.

Barnhisel, R.I., and P.M. Bertsch. 1989. Chlorites and hydroxy-interlayered vermiculite and smectite, p. 729–788. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Barré, P., G. Berger, and B. Velde. 2009. How element translocation by plants may stabilize illitic clays in the surface of temperate soils. Geoderma 151:22–30.

Barré, P., B. Velde, C. Fontaine, N. Catel, and L. Abbadie. 2008. Which 2:1 clay minerals are involved in the soil potassium reservoir? Insights from potassium addition or removal experiments on three temperate grassland soil clay assemblages. Geoderma 146:216–223.

Barshad, I., and F.M. Kishk. 1969. Chemical composition of soil vermiculite clays as related to their genesis. Contrib. Miner. Petrol. 24:136–155.

Barthès, B.G., E. Kouakoua, M.-C. Larré-Larrouy, T.M. Razafimbelo, E.F. de Luca, A. Azontonde, C.S.V.J. Neves, P.L. de Freitas, and C.L. Feller. 2008. Texture and sesquioxide effects on water-stable aggregates and organic matter in some tropical soils. Geoderma 143:14–25.

Basile-Doelsch, I., R. Amundson, W.E.E. Stone, C.A. Masiello, J.Y. Bottero, F. Colin, D. Borschneck, and J.D. Meunier. 2005. Mineralogical control of organic carbon dynamics in a volcanic ash soil on La Réunion. Eur. J. Soil Sci. 56:689–703.

Bassett, W.A. 1959. The origin of the vermiculite deposit at Libby, Montana. Am. Mineral. 44:282–299.

Batchelder, M., J.D. Mather, and J.B. Joseph. 1998. The stability of the Oxford clay as a mineral liner for landfill. Water Environ. J. 12:92–97.

Bates, T.E. 1962. Halloysite and gibbsite formation in Hawaii. Clays Clay Miner. 9:315–328.

Bennett, P.C., J.R. Rogers, W.J. Choi, and F.K. Hiebert. 2001. Silicates, silicate weathering, and microbial ecology. Geomicrobiol. J. 18:3–19.

Berg, A., and S. Banwart. 2000. Carbon dioxide mediated dissolution of Ca-feldspar: Implications for silicate weathering. Chem. Geol. 163:25–52.

Bergaya, F., B.K.G. Theng, and G. Lagaly (eds.). 2006. Handbook of clay science. Vol. 1. Developments in clay science. Elsevier, Amsterdam, the Netherlands.

Berkgaut, V., A. Singer, and K. Stahr. 1994. Palagonite reconsidered: Paracrystalline illite–smectites from regoliths on basic pyroclastics. Clays Clay Miner. 42:582–592.

Berner, R.A., and G.R. Holdren, Jr. 1979. Mechanism of feldspar weathering. II. Observations of feldspars from soils. Geochim. Cosmochim. Acta 43:1173–1186.

Bétard, F., L. Caner, and Y. Gunnell. 2009. Illite neoformation in plagioclase during weathering: Evidence from semi-arid northeast Brazil. Geoderma 152:53–62.

Bhattacharyya, T., D.K. Pal, and P. Srivastava. 2000. Formation of gibbsite in the presence of 2:1 minerals: An example from Ultisols of northeast India. Clay Miner. 35:827–840.

Bigham, J.M., R.W. Fitzpatrick, and D.G. Schulze. 2002. Iron oxides, p. 323–366. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA Book Series 7. SSSA, Madison, WI.

Birkeland, P.W. 1999. Soils and geomorphology. 3rd edn. Oxford University Press, New York.

Bishop, J.L., E.Z. Noe Dobrea, N.K. McKeown, M. Parente, B.L. Ehlmann, J.R. Michalski, R.E. Milliken, F. Poulet, G.A. Swayze, J.F. Mustard, S.L. Murchie, and J.-P. Bibring. 2008. Phyllosilicate diversity and past aqueous activity revealed at Mawrth Vallis, Mars. Science 321:830–833.

Bleeker, P., and R.L. Parfitt. 1974. Volcanic ash and its clay mineralogy at Cape Hoskins, New Britain, Paua New Guinea. Geoderma 11:123–135.

Blum, A.E., and L.L. Stillings. 1995. Feldspar dissolution kinetics. Rev. Miner. Geochem. 31:291–351.

Blum, W.E.H., B.R. Warkentin, and E.E. Frossard. 2006. Soil, human society and the environment, p. 1–8. *In* E. Frossard, W.E.H. Blum, and B.P. Warkentin (eds.) Function of soils for human societies. Special Publication No. 266. Geological Society, London, U.K.

Bockheim, J.G. 1997. Properties and classification of cold desert soils from Antarctica. Soil Sci. Soc. Am. J. 61:224–231.

Boettinger, J.L., and D.W. Ming. 2002. Zeolites, p. 585–610. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Boettinger, J.L., and R.J. Southard. 1995. Phyllosilicate distribution and origin in Aridisols on a granitic pediment, western Mojave desert. Soil Sci. Soc. Am. J. 59:1189–1198.

Bonifacio, E., E. Zanini, V. Boero, and M. Franchini-Angela. 1997. Pedogenesis in a soil catena on serpentinite in north-western Italy. Geoderma 75:33–51.

Borchardt, G. 1989. Smectites, p. 675–727. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Borggaard, O.K. 1983. Iron oxides in relation to aggregation of soil particles. Acta Agric. Scand. 33:257–260.

Bouza, P.J., M. Simón, J. Aguilar, K. Del Valle, and M. Rostagno. 2007. Fibrous-clay mineral formation and soil evolution in Aridisols of northeastern Patagonia, Argentina. Geoderma 139:38–50.

Bowen, N.L. 1922. The reaction principle in petrogenesis. J. Geol. 30:177–198.

Boyer, S.J. 1975. Chemical weathering of rocks on the Lassiter Coast, Antarctic Peninsula, Antarctica. N.Z. J. Geol. Geophys. 18:623–628.

Brinkman, R. 1970. Ferrolysis, a hydromorphic soil forming process. Geoderma 3:199–206.

Bühmann, C., and P.L.C. Grubb. 1991. A kaolin–smectite interstratification sequence from a red and black complex. Clay Miner. 26:343–358.

Buol, S.W. 1965. Present soil-forming factors and processes in arid and semiarid regions. Soil Sci. 99:45–49.

Buol, S.W., R.J. Southard, R.C. Graham, and P.A. McDaniel. 2003. Soil genesis and classification. 5th edn. Iowa State Press, Ames, IA.

Burton, E.D., R.T. Bush, and L.A. Sullivan. 2006. Sedimentary iron geochemistry in acidic waterways associated with coastal lowland acid sulfate soils. Geochim. Cosmochim. Acta 70:5455–5468.

Burton, E.D., R.T. Bush, L.A. Sullivan, R.K. Hocking, D.R.G. Mitchell, S.G. Johnston, R.W. Fitzpatrick, M. Raven, S. McClure, and L.Y. Yang. 2009. Iron-monosulfide oxidation in natural sediments: Resolving microbially mediated S transformations using XANES, electron microscopy, and selective extractions. Environ. Sci. Technol. 43:3128–3134.

Burton, E.D., R.T. Bush, L.A. Sullivan, and D.R.G. Mitchell. 2008. Schwertmannite transformation to goethite via the Fe(II) pathway: Reaction rates and implications for iron-sulfide formation. Geochim. Cosmochim. Acta 72:4551–4564.

Buurman, P., and A.G. Jongmans. 2005. Podzolisation and soil organic matter dynamics. Geoderma 125:71–83.

Cady, J.G., and K.W. Flach. 1997. History of soil mineralogy in the United States Department of Agriculture. Adv. Geoecol. 29:211–240.

Caillaud, J., D. Proust, D. Righi, and F. Martin. 2004. Fe-rich clays in a weathering profile developed from serpentinite. Clays Clay Miner. 52:779–791.

Calvaruso, C., L. Mareschal, M.-P. Turpault, and E. Leclerc. 2009. Rapid clay weathering in the rhizosphere of Norway spruce and oak in an acid forest ecosystem. Soil Sci. Soc. Am. J. 73:331–338.

Calvert, C.S., S.W. Buol, and S.B. Weed. 1980. Mineralogical transformations of a vertical rock–saprolite–soil sequence in the North Carolina Piedmont. Soil Sci. Soc. Am. J. 44:1096–1112.

Carnicelli, S., A. Mirabella, G. Cecchini, and G. Sanesi. 1997. Weathering of chlorite to a low-charge expandable mineral in a Spodosol on the Apennine Mountains, Italy. Clays Clay Miner. 45:28–41.

Carroll, D. 1970. Rock weathering. Monographs in geoscience. Plenum Press, New York.

Carson, J.K., L. Campbell, D. Rooney, N. Clipson, and D.B. Gleeson. 2009. Minerals in soil select distinct bacterial communities in their microhabitats. FEMS Microbiol. Ecol. 67:381–388.

Carson, C.D., and G.W. Kunze. 1970. New occurrence of tabular halloysite. Soil Sci. Soc. Am. Proc. 34:538–540.

Cas, R.A.F., and J.V. Wright. 1987. Volcanic successions—Modern and ancient. Allen and Unwin, London, U.K.

Casey, W.H. 1995. Surface chemistry during the dissolution of oxides and silicate minerals, p. 185–217. *In* D.J. Vaughan and R.A.D. Pattrick (eds.) Mineral surfaces. Chapman and Hall, London, U.K.

Casey, W.H., M.F. Hochella, and H.R. Westrich. 1993. The surface chemistry of manganiferous silicate minerals as inferred from experiments on tephroite (Mn_2SiO_4). Geochim. Cosmochim. Acta 57:785–793.

Casey, W.H., H.R. Westrich, and G.W. Arnold. 1988. Surface chemistry of labradorite feldspar reacted with aqueous solutions at pH = 2, 3 and 12. Geochim. Cosmochim. Acta 52:821–832.

Casey, W.H., H.R. Westrich, G.W. Arnold, and J.F. Banfield. 1989. The surface chemistry of dissolving labradorite feldspar. Geochim. Cosmochim. Acta 53:2795–2807.

Certini, G., S. Hillier, E. McMurray, and A.C. Edwards. 2003. Weathering of sandstone clasts in a forest soil in Tuscany (Italy). Geoderma 116:357–372.

Certini, G., and R. Scalenghe. 2006. Soil formation on Earth and beyond: The role of additional soil-forming factors, p. 193–210. *In* G. Certini and R. Scalenghe (eds.) Soils: Basic concepts and future challenges. Cambridge University Press, Cambridge, U.K.

Certini, G., M.J. Wilson, S.J. Hillier, A.R. Fraser, and E. Delbos. 2006. Mineral weathering in trachydacitic-derived soils and saprolites involving formation of embryonic halloysite and gibbsite at Mt. Amiata, central Italy. Geoderma 133:173–190.

Chadwick, O.A., and J. Chorover. 2001. The chemistry of pedogenic thresholds. Geoderma 100:321–353.

Chadwick, O.A., R.T. Gavenda, E.F. Kelly, K. Ziegler, C.G. Olson, W.C. Elliott, and D.M. Hendricks. 2003. The impact of climate on the biogeochemical functioning of volcanic soils. Chem. Geol. 202:195–223.

Chadwick, O.A., D.M. Hendricks, and W.D. Nettleton. 1987. Silica in duric soils: I. A depositional model. Soil Sci. Soc. Am. J. 51:975–982.

Chamley, H. 1989. Clay sedimentology. Springer-Verlag, Berlin, Germany.

Chan, C.S., S.C. Fakra, D.C. Edwards, D. Emerson, and J.F. Banfield. 2009. Iron oxyhydroxide mineralization on microbial extracellular polysaccharides. Geochim. Cosmochim. Acta 73:3807–3818.

Chen, J., H.-P. Blume, and L. Beyer. 2000. Weathering of rocks induced by lichen colonization—A review. Catena 39:121–146.

Chenu, C., and A.F. Plante. 2006. Clay-sized organo-mineral complexes in a cultivation chronosequence: Revisiting the concept of the 'primary organo-mineral complex.' Eur. J. Soil Sci. 57:596–607.

Childs, C.W. 1992. Ferrihydrite: A review of structure, properties and occurrence in relation to soils. Z. Pflanzenernähr. Bodenkd. 155:41–448.

Childs, C.W., R.L. Parfitt, and R.H. Newman. 1990. Structural studies of Silica Springs allophane. Clay Miner. 25:329–341.

Cho, H.D., and A.R. Mermut. 1992. Evidence for halloysite formation from weathering of ferruginous chlorite. Clays Clay Miner. 40:608–619.

Churchman, G.J. 1978. Studies on a climosequence of soils in tussock grasslands. 21. Mineralogy. N.Z. J. Sci. 21:467–480.

Churchman, G.J. 1980. Clay minerals formed from micas and chlorites in some New Zealand soils. Clay Miner. 15:59–76.

Churchman, G.J. 1990. Relevance of different intercalation tests for distinguishing halloysite from kaolinite in soils. Clays Clay Miner. 38:591–599.

Churchman, G.J. 2000. The alteration and formation of soil minerals by weathering, p. F3–F76. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Churchman, G.J. 2006. Soil phases: The inorganic solid phase, p. 23–44. *In* G. Certini and R. Scalenghe (eds.) Soils: Basic concepts and future challenges. Cambridge University Press, Cambridge, U.K.

Churchman, G.J. 2010. Is the geological concept of clay minerals appropriate for soil science? Phys. Chem. Earth. 35:922–940.

Churchman, G.J., and J.G. Bruce. 1988. Relationships between loess deposition and mineral weathering in some soils in Southland, New Zealand, p. 11–31. *In* D.N. Eden and R.J. Furkert (eds.) Loess: Its distribution, geology and soils. A.A. Balkema, Rotterdam, the Netherlands.

Churchman, G.J., and C.M. Burke. 1991. Properties of subsoils in relation to various measures of surface area and moisture contents. J. Soil Sci. 42:463–478.

Churchman, G.J., and R.M. Carr. 1975. The definition and nomenclature of halloysites. Clays Clay Miner. 23:382–388.

Churchman, G.J., R.N. Clayton, K. Sridhar, and M.L. Jackson. 1976. Oxygen isotopic composition of aerosol-sized quartz in shales. J. Geophys. Res. 81:381–386.

Churchman, G.J., and R.J. Gilkes. 1989. Recognition of intermediates in the possible transformation of halloysite to kaolinite. Clay Miner. 24:579–590.

Churchman, G.J., I.R. Pontifex, and S.G. McClure. 2010. Factors influencing the formation and characteristics of halloysites or kaolinites in granitic and tuffaceous saprolites in Hong Kong. Clays Clay Miner. 58:122–139.

Churchman, G.J., P.G. Slade, P.G. Self, and L.J. Janik. 1994. Nature of interstratified kaolin–smectites in some Australian soils. Aust. J. Soil Res. 32:805–822.

Churchman, G.J., and K.R. Tate. 1987. Stability of aggregates of different size grades in allophanic soils from volcanic ash in New Zealand. J. Soil Sci. 38:19–27.

Churchman, G.J., and B.K.G. Theng. 1984. Interactions of halloysites with amides: Mineralogical factors affecting complex formation. Clay Miner. 19:161–175.

Claridge, G.G.C. 1965. The clay mineralogy and chemistry of some soils from Ross dependency, Antarctica. N.Z. J. Sci. 8:186–220.

Claridge, G.G.C., and I.B. Campbell. 1968. Soils of the Shackleton Glacier, Queen Maud Range, Antarctica. N.Z. J. Sci. 11:171–218.

Claridge, G.G.C., and I.B. Campbell. 2008. Zeolites in Antarctic soils: Examples from Coombs Hills and Marble Point. Geoderma 144:66–72.

Clayton, R.N., M.L. Jackson, and K. Sridhar. 1978. Resistance of quartz silt to isotopic exchange under burial and intense weathering conditions. Geochim. Cosmochim. Acta 42:1517–1522.

Coleman, D.C., and D.A. Crossley, Jr. 2003. Fundamental of soil ecology. Academic Press, Amsterdam, the Netherlands.

Colombo, C., and J. Torrent. 1991. Relationships between aggregation and iron oxides in Terra Rossa soils from southern Italy. Catena 18:51–59.

Cornell, R.M., and U. Schwertmann. 1996. The iron oxides. VCH, Weinheim, Germany.

Courchesne, F. 2006. Factors of soil formation: Biota. As exemplified by case studies on the direct imprint of trees on trace metal concentrations, p. 165–179. *In* G. Certini and R. Scalenghe (eds.) Soils: Basic concepts and future challenges. Cambridge University Press, Cambridge, U.K.

Cradwick, P.D.G., V.C. Farmer, J.D. Russell, C.R. Masson, K. Wada, and N. Yoshinaga. 1972. Imogolite, a hydrated aluminium silicate of tubular structure. Nat. Phys. Sci. 240:187–189.

Craig, D.C., and F.C. Loughnan. 1964. Chemical and mineralogical transformations accompanying the weathering of basic volcanic rocks from New South Wales. Aust. J. Soil Res. 2:218–234.

Cronin, S.J., V.E. Neall, and A.S. Palmer. 1996. Investigation of an aggrading paleosol developed into andesitic ring-plain deposits, Ruapehu volcano, New Zealand. Geoderma 69:119–135.

Cuadros, J., A. Delgado, A. Cardenete, E. Reyes, and J. Linares. 1994. Kaolinite/montmorillonite resembles beidellite. Clays Clay Miner. 42:643–651.

Dahlgren, R.A., J.P. Dragoo, and F.C. Ugolini. 1997. Weathering of Mt. St. Helens tephra under a cryic–udic climatic regime. Soil Sci. Soc. Am. J. 61:1519–1525.

Dahlgren, R.A., M. Saigusa, and F.C. Ugolini. 2004. The nature, properties, and management of volcanic soils. Adv. Agron. 82:113–182.

Daux, V., J.L. Crovisier, C. Hemond, and J.C. Petit. 1994. Geochemical evolution of basaltic rocks subjected to weathering: Fate of the major elements, rare earth elements, and thorium. Geochim. Cosmochim. Acta 58:4941–4954.

De Vleeschouwer, F., B. Van Vliët Lanoé, and N. Fagel. 2008. Long term mobilisation of chemical elements in tephra-rich peat (NE Iceland). Appl. Geochem. 23:3819–3839.

Delmonte, B., I. Basile-Doelsch, J.-R. Petit, V. Maggi, M. Revel-Rolland, A. Michard, E. Jagoutz, and F. Grousset. 2004. Comparing the Epica and Vostok dust records during the last 220,000 years: Stratigraphical correlation and provenance in glacial periods. Earth Sci. Rev. 66:63–87.

Delvaux, B., and A.J. Herbillon. 1995. Pathways of mixed-layer kaolin–smectite formation in soils, p. 457–461. *In* G.J. Churchman, R.W. Fitzpatrick, and R.A. Eggleton (eds.) Clays: Controlling the environment. Proc. 10th Int. Clay Conf., 18–23 July 1993. Adelaide, Australia. CSIRO Publishing, Melbourne, Australia.

Delvigne, J. 1998. Atlas of micromorphology of mineral alteration and weathering. The Canadian Mineralogist. Special Publication No. 3. Mineralogical Association of Canada, Ottawa, Canada.

Denef, K., and J. Six. 2005. Clay mineralogy determines the importance of biological versus abiotic processes for macroaggregate formation and stabilization. Eur. J. Soil Sci. 56:469–479.

Dent, D.L., and L.J. Pons. 1995. A world perspective on acid sulphate soils. Geoderma 67:263–276.

Derry, L.A., A.C. Kurtz, K. Ziegler, and O.A. Chadwick. 2005. Biological control of terrestrial silica cycling and export fluxes to watershed. Nature 433:728–731.

Deshpande, T.L., D.J. Greenland, and J.P. Quirk. 1968. Changes in soil properties associated with the removal of iron and aluminium oxides. J. Soil Sci. 19:108–122.

Dias, I., I. Gonzalez, S. Prates, and E. Galán. 1997. Palygorskite occurrences in the Portuguese sector of the Tagus basin: A preliminary report. Clay Miner. 32:323–328.

Dixon, J.B., and T.R. McKee. 1974. Internal and external morphology of tubular and spheroidal halloysite particles. Clays Clay Miner. 22:127–137.

Dixon, J.B., and G.N. White. 2002. Manganese oxides, p. 367–388. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Do Nascimento, N.R., G.T. Bueno, E. Fritsch, A.J. Herbillon, T. Allard, A.J. Melfi, R. Astolfo, and Y. Li. 2004. Podzolization as a deferralitization process: A study of an Acrisol–Podzol sequence derived from palaeozoic sandstones in the northern Amazon basin. Eur. J. Soil Sci. 55:523–538.

Doner, H.E., and P.R. Grossl. 2002. Carbonates and evaporates, p. 199–228. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Doner, H.E., and W.C. Lynn. 1989. Carbonate, halide, sulfate, and sulfide minerals, p. 279–330. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Dong, H., D.P. Jaisi, J. Kim, and G. Zhang. 2009. Microbe-clay mineral interactions. Am. Mineral. 94:1505–1519.

Douglas, L.A. 1989. Vermiculites, p. 635–674. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Drees, L.R., L.P. Wilding, N.E. Smeck, and A.L. Senkayi. 1989. Silica in soils: Quartz and disordered silica polymorphs, p. 913–974. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA Book Series 1. SSSA, Madison, WI.

Drief, A., and P. Schiffman. 2004. Very low-temperature alteration of sideromelane in hyaloclastites and hyalotuffs from Kilauea and Mauna Kea volcanoes: Implications for the mechanism of palagonite formation. Clays Clay Miner. 52:622–634.

Ducloux, J., Y. Guero, P. Sardini, and A. Decarreau. 2002. Xerolysis: A hypothetical process of clay particles weathering under Sahelian climate. Geoderma 105:83–110.

Ducloux, J., A. Meunier, and B. Velde. 1976. Smectite, chlorite and a regular interlayered chlorite–vermiculite in soils developed on a small serpentinite body, Massif Central, France. Clay Miner. 11:121–135.

Dudas, M.J., and M.E. Harward. 1975. Weathering and authigenic halloysite in soil developed in Mazama ash. Soil Sci. Soc. Am. Proc. 39:561–566.

Duiker, S.W., F.E. Rhoton, J. Torrent, N.E. Smeck, and R. Lal. 2003. Iron (hydr)oxide crystallinity effects on soil aggregation. Soil Sci. Soc. Am. J. 67:606–611.

Egashira, K., and S. Tsuda. 1983. High-charge smectite found in weathered granitic rocks of Kyushu. Clay Sci. 6:67–81.

Eger, A., and A.E. Hewitt. 2008. Soils and their relationship to aspect and vegetation history in the eastern Southern Alps, Canterbury High Country, New Zealand. Catena 75:297–307.

Eggleton, R.A. 1984. Formation of iddingsite rims on olivine: A transmission electron microscopy study. Clays Clay Miner. 32:1–11.

Eggleton, R.A., and J.N. Boland. 1982. Weathering of enstatite to talc through a sequence of transitional phases. Clays Clay Miner. 30:11–20.

Eggleton, R.A., and P.R. Buseck. 1980. High resolution electron microscopy of feldspar weathering. Clays Clay Miner. 28:173–178.

Egli, M., A. Mirabella, A. Mancabelli, and G. Sartori. 2004. Weathering of soils in alpine areas as influenced by climate and parent material. Clays Clay Miner. 52:287–303.

Egli, M., M. Nater, M.A. Mirabella, S. Raimondi, M. Plötze, and L. Alioth. 2008. Clay minerals, oxyhydroxide formation, element leaching and humus development in volcanic soils. Geoderma 143:101–114.

Ehlmann, B.L., J.F. Mustard, G.A. Swayze, R.N. Clark, J.L. Bishop, F. Poulet, D.J.D. Marais, L.H. Roach, R.E. Milliken, J.J. Wray, O. Barnouin-Jha, and S.L. Murchie. 2009. Identification of hydrated silicate minerals on Mars using MRO-CRISM: Geologic context near Nili Fossae and implications for aqueous alteration. J. Geophys. Res. 114:E00D08. doi:10.1029/2009JE003339.

Elsass, F., D. Dubreoucq, and M. Thiry. 2000. Diagenesis of silica minerals from clay minerals in volcanic soils of Mexico. Clay Miner. 35:477–489.

Erickson, G.E. 1983. The Chilean nitrate deposit. Am. Sci. 71:366–374.

Estoule-Choux, J., J. Estoule, and D. Hallalouche. 1995. Congruent dissolution of microcline and epitaxial growth of "skeletal" quartz during weathering of a granite from the central Hoggar (Algeria), p. 373–377. *In* G.J. Churchman, R.W. Fitzpatrick, and R.A. Eggleton (eds.) Clays controlling the environment. Proc. 10th Int. Clay Conf. CSIRO Publishing, Melbourne, Australia.

Eswaran, H., and C.B. Wong. 1978. A study of a deep weathering profile on granite in Peninsular Malaysia. Parts I, II, and III. Soil Sci. Soc. Am. J. 42:144–158.

Eswaran, H., and Y.H. Yeow. 1976. The weathering of biotite in a profile on gneiss in Malaysia. Geoderma 16:9–20.

Etame, J., M. Gerard, C.E. Suh, and P. Bilong. 2009. Halloysite neoformation during the weathering of nephelinitic rocks under humid tropical conditions at Mt Etinde, Cameroon. Geoderma 154:59–68.

Eusterhues, K., C. Rumpel, and I. Kögel-Knabner. 2005. Organo-mineral associations in sandy acid forest soils: Importance of specific surface area, iron oxides and micropores. Eur. J. Soil Sci. 56:753–763.

Fairbridge, R.W., and J. Bourgeois. 1978. The encyclopedia of sedimentology. Dowden, Hutchinson and Ross, Inc., Stroudsberg, PA.

Fanning, D.S., V.Z. Keramidas, and M.A. El-Desoky. 1989. Micas, p. 551–634. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Fanning, D.S., M.C. Rabenhorst, S.N. Burch, K.R. Islam, and S.A. Tangren. 2002. Sulfides and sulfates, p. 229–260. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Farmer, V.C. 1982. Significance of the presence of allophane and imogolite in podzol Bs horizons for podzolization mechanisms: A review. Soil Sci. Plant Nutr. 28:571–578.

Farmer, V.C. 1997. Conversion of ferruginous allophanes to ferruginous beidellites at 95°C under alkaline conditions with alternating oxidation and reduction. Clays Clay Miner. 45:591–597.

Farmer, V.C. 2005. Forest vegetation does recycle substantial amounts of silicon from and back to the soil solution with phytoliths as an intermediate phase, contrary to recent reports. Eur. J. Soil Sci. 56:271–272.

Farmer, V.C., E. Delbos, and J.D. Miller. 2005. The role of phytolith formation and dissolution in controlling concentrations of silica in soil solutions and streams. Geoderma 127:71–79.

Farmer, V.C., and A.R. Fraser. 1982. Chemical and colloidal stability of sols in the $Al_2O_3–Fe_2O_3–SiO_2–H_2O$ system: Their role in podzolization. J. Soil Sci. 33:737–742.

Farmer, V.C., and D.G. Lumsdon. 2001. Interactions of fulvic acid with aluminium and a proto-imogolite sol: The contribution of E-horizon eluates to podzolization. Eur. J. Soil Sci. 52:177–188.

Farmer, V.C., W.J. McHardy, L. Robertson, A. Walker, and M.J. Wilson. 1985. Micromorphology and sub-microscopy of allophane and imogolite in a podzol Bs horizon: Evidence for translocation and origin. J. Soil Sci. 36:87–95.

Farmer, V.C., J.D. Russell, and M.L. Berrow. 1980. Imogolite and proto-imogolite allophane in spodic horizons: Evidence for a mobile aluminium silicate complex in podzol formation. J. Soil Sci. 31:673–684.

Farmer, V.C., J.D. Russell, W.J. McHardy, A.C.D. Newman, J.L. Ahlrichs, and J.Y.H. Rimsaite. 1971. Evidence for loss of protons and octahedral iron from oxidised biotites and vermiculites. Mineral Mag. 38:121–137.

Farmer, V.C., J.D. Russell, and B.F.L. Smith. 1983. Extraction of inorganic forms of translocated Al, Fe and Si from a podzol Bs horizon. J. Soil Sci. 34:571–576.

Ferris, F.G. 1997. Formation of authigenic minerals by bacteria, p. 187–208. *In* J.M. McIntosh and L.A. Groat (eds.) Biological–mineralogical interactions. Mineralogical Association of Canada. Short Course. Vol. 25. Mineralogical Association of Canada, Ottawa, Canada.

Fieldes, M. 1955. Clay mineralogy of New Zealand soils. Part II: Allophane and related mineral colloids. N.Z. J. Sci. Technol. 37:336–350.

Fieldes, M. 1968. Clay mineralogy, In Soils of New Zealand, Part 2. N.Z. Soil Bur. Bull. 26:22–39.

Fieldes, M., and L.D. Swindale. 1954. Chemical weathering of silicates in soil formation. N.Z. J. Sci. Technol. B36:140–154.

Fisher, C.B., and P.C. Ryan. 2006. The smectite-to-disordered kaolinite transition in a tropical soil chronosequence, Pacific coast, Costa Rica. Clays Clay Miner. 54:571–586.

Fisher, R.V., and H.-U. Schminke. 1984. Pyroclastic rocks. Springer-Verlag, Berlin, Germany.

Fitzpatrick, R.W., and D.J. Chittleborough. 2002. Titanium and zirconium minerals, p. 667–690. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Fitzpatrick, R.W., E. Fritsch, and P.G. Self. 1996. Interpretation of soil features produced by ancient and modern processes in degraded landscapes. V. Development of saline sulfidic features in non-tidal seepage areas. Geoderma 69:1–29.

Fortin, D., and S. Langley. 2005. Formation and occurrence of biogenic iron-rich minerals. Earth Sci. Rev. 72:1–19.

Franke, W.A., and R. Teschner-Steinhardt. 1994. An experimental approach to the sequence of the stability of rock-forming minerals towards chemical weathering. Catena 21:279–290.

Fritsch, E., G. Morin, A. Bedidi, D. Bonnin, E. Balan, S. Caquineau, and G. Calas. 2005. Transformation of haematite and Al-poor goethite to Al-rich goethite and associated yellowing in a ferralitic clay soil profile of the middle Amazon Basin (Manaus, Brazil). Eur. J. Soil Sci. 56:575–588.

Fu, F.F., T. Akagi, and S. Yabuki. 2002. Origin of silica particles found in the cortex of Matteuccia roots. Soil Sci. Soc. Am. J. 66:1265–1271.

Furian, S., L. Barbiéro, R. Boulet, P. Curmi, M. Grimaldi, and C. Grimaldi. 2002. Distribution and dynamics of gibbsite and kaolinite in an Oxisol of Serra do Mar, southeastern Brazil. Geoderma 106:83–100.

Furquim, S.A.C., R.C. Graham, L. Barbiero, J.P. de Queiroz Neto, and V. Valles. 2008. Mineralogy and genesis of smectites in an alkaline–saline environment of Pantanal wetland, Brazil. Clays Clay Miner. 56:579–595.

Gaines, R.V., H.W. Skinner, E.F. Foord, B. Mason, and A. Rosenzweig. 1997. Dana's new mineralogy. John Wiley & Sons, New York.

Galán, E., J.M. Brell, A. La Iglesia, and H.S. Robertson. 1975. The Cáceras palygorskite deposit, Spain, p. 81–94. *In* S.W. Bailey (ed.) Proc. Int. Clay Conf., Mexico. Applied Publishing, Wilmette, IL.

Galán, E., and A. Castillo. 1984. Sepiolite–palygorskite in Spanish tertiary basins: Genetical patterns in continental environments. Dev. Sedimentol. 37:87–124.

Gallez, A., A.S.R. Juo, and A.J. Herbillon. 1976. Surface and charge characteristics of selected soils in the tropics. Soil Sci. Soc. Am. J. 40:601–608.

Gardam, M., A.J. Mason, A.F. Reid, G.J. Churchman, and M. Raven. 2008. Arumpo bentonite deposits: Distinctive indicators of past volcanic events in the Murray Basin, southeastern Australia. Aust. J. Earth Sci. 55:183–194.

Garrels, R.M., and F.T. Mackenzie. 1971. Evolution of sedimentary rocks. W.W. Norton and Co, New York.

Gates, W.P. 2006. X-ray absorption spectroscopy, p. 789–864. *In* F. Bergaya, B.K.G. Theng, and G. Lagaly (eds.) Handbook of clay science. Developments in clay science 1. Elsevier, Amsterdam, the Netherlands.

Geiss, C.E., and C.W. Zanner. 2006. How abundant is pedogenic magnetite? Abundance and grain size estimates for loessic soils based on rock magnetic analyses. J. Geophys. Res. 111:B12S21. doi:10.1029/2006JB004564.

Gérard, M., S. Caquineau, J. Pinheiro, and G. Stoops. 2007. Weathering and allophane neoformation in soils developed on volcanic ash in the Azores. Eur. J. Soil Sci. 58:496–515.

Gerin, P.A., M.J. Genet, A.J. Herbillon, and B. Delvaux. 2003. Surface analysis of soil material by x-ray photoelectron spectroscopy. Eur. J. Soil Sci. 54:589–603.

Giesler, R., H. Ilvesniemi, L. Nyberg, P. van Hees, M. Starr, K. Bishop, T. Kareinen, and U.S. Lundström. 2000. Distribution and mobilization of Al, Fe and Si in three podzolic soil profiles in relation to the humus layer. Geoderma 94:249–263.

Gilkes, R.J. 1990. Mineralogical insights into soil productivity: An anatomical perspective, p. 63–73. *In* Transactions, 14th Int. Cong. Soil Sci., 12–18 August 1990, Kyoto, Japan.

Gilkes, R.J., and A. Suddhiprakarn. 1979. Biotite alteration in deeply weathered granite. I. Morphological, mineralogical, and chemical properties. Clays Clay Miner. 27:249–360.

Gillot, F., D. Righi, and M.L. Räisänen. 2001. Layer-charge evaluation of expandable clays from a chronosequence of podzols in Finland using an alkylammonium method. Clay Miner. 36:571–584.

Gislason, S.R., and E.H. Oelkers. 2003. Mechanism, rates, and consequences of basaltic glass dissolution: II. An experimental study of the dissolution rates of basaltic glass as a function of pH and temperature. Geochim. Cosmochim. Acta 67:3817–3832.

Gjems, O. 1970. Mineralogical composition and pedogenic weathering of the clay fraction in podzol weathering profiles in Zalesine, Yugoslavia. Soil Sci. 110:237–243.

Glassmann, J.R., and G.M. Simonson. 1985. Alteration of basalt in soils of western Oregon. Soil Sci. Soc. Am. J. 49:262–273.

Goldich, S.S. 1938. A study in rock weathering. J. Geol. 46:17–58.

Goodman, B.A., J.D. Russell, B. Montez, E. Oldfield, and R.J. Kirkpatrick. 1985. Structural studies of imogolite and allophanes by aluminum-27 and silicon-29 nuclear magnetic resonance spectroscopy. Phys. Chem. Miner. 12:342–346.

Green, B.E. 1987. Weathering of buried paleosols on late Quaternary rhyolitic tephras, Rotorua region, New Zealand. Unpublished M.Sc. thesis. University of Waikato. Hamilton, New Zealand.

Greenland, D.J. 1965. Interactions between clays and organic compounds in soils. Part 1. Mechanisms of interaction between clays and defined organic compounds. Soils Fertil. 28:415–425.

Grey, I.E., and A.F. Reid. 1975. The structure of pseudorutile and its role in the natural alteration of ilmenite. Am. Mineral. 60:898–906.

Grim, R.E. 1968. Clay mineralogy. McGraw-Hill, New York.

Grim, R.E., R.H. Bray, and W.F. Bradley. 1937. The mica in argillaceous sediments. Am. Mineral. 22:813–829.

Gustafsson, J.P. 2001. Modelling competitive anion adsorption on oxide minerals and an allophane-containing soil. Eur. J. Soil Sci. 52:639–653.

Hall, P.L., G.J. Churchman, and B.K.G. Theng. 1985. Size distribution of allophane unit particles in aqueous suspensions. Clays Clay Miner. 33:345–349.

Hamblin, A.P., and D.J. Greenland. 1977. Effect of organic constituents and complexed metal ions on aggregate stability of some East Anglian soils. J. Soil Sci. 28:410–416.

Hamilton, V.E., P.R. Christensen, and J.L. Bandfield. 2003. Volcanism or aqueous alteration on Mars? Nature 421:711–712.

Harder, H. 1972. The role of magnesium in the formation of smectite minerals. Chem. Geol. 10:31–39.

Harder, H. 1977. Clay mineral formation under lateritic weathering conditions. Clay Miner. 12:281–288.

Harlan, P.W., D.P. Franzmeier, and C.B. Roth. 1977. Soil formation on loess in southwestern Indiana. II. Distribution of clay and free iron oxides and fragipan formation. Soil Sci. Soc. Am. J. 42:99–103.

Harris, W.G. 2002. Phosphate minerals, p. 637–665. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Hart, R.D., R.J. Gilkes, S. Siradz, and B. Singh. 2002. The nature of soil kaolins from Indonesia and Western Australia. Clays Clay Miner. 50:198–207.

Hart, R.D., W. Wiriyakitnateekul, and R.J. Gilkes. 2003. Properties of soil kaolins from Thailand. Clay Miner. 38:71–94.

Hartemink, A.E., A.B. McBratney, and J.A. Cattle. 2001. Developments and trends in soil science: 100 volumes of Geoderma (1967–2001). Geoderma 100:217–268.

Hay, R.L. 1960. Rate of clay formation and mineral alteration in a 4000-year-old volcanic ash soil on St. Vincent, B.W.I. Am. J. Sci. 258:354–368.

Hayes, M.H.B., P. MacCarthy, R.L. Malcolm, and R.S. Swift. 1989. The search for structure: Setting the scene, p. 3–31. *In* M.H.B. Hayes, P. MacCarthy, R.L. Malcolm, and R.S. Swift (eds.) Humic substances. II. John Wiley & Sons, Chichester, U.K.

He, Y., D.C. Li, B. Velde, Y.F. Yang, C.M. Huang, Z.T. Gong, and G.I. Zhang. 2008. Clay minerals in a soil chronosequence derived from basalt on Hainan Island, China and its implication for pedogenesis. Geoderma 148:206–212.

Hellmann, R., C.H. Egglestone, M.F. Hochella, Jr., and D.A. Crerar. 1990. The formation of leached layers on albite surfaces during dissolution under hydrothermal conditions. Geochim. Cosmochim. Acta 54:1267–1282.

Hemley, J.J. 1959. Some mineralogical equilibria in the system $K_2O–Al_2O_3–SiO_2–H_2O$. Am. J. Sci. 257:241–270.

Hendricks, S.B., and W.H. Fry. 1930. The results of X-ray and microscopical examinations of soil colloids. Soil Sci. 29:457–479.

Henriet, C., N. De Jaeger, M. Dore, S. Opfergelt, and B. Delvaux. 2008. The reserve of weatherable primary silicates impacts the accumulation of biogenic silicon in volcanic ash soils. Biogeochemistry 90:209–223.

Herbillon, A.J., R. Frankart, and L. Vielvoye. 1981. An occurrence of interstratified kaolinite–smectite minerals in a red-black soil toposequence. Clay Miner. 16:195–201.

Herbillon, A.J., and M.N. Makumbi. 1975. Weathering of chlorite in a soil derived from a chlorito-schist under humid tropical conditions. Geoderma 13:89–104.

Herbillon, A.J., M.M. Mestdagh, L. Vielvoye, and E. Derouane. 1976. Iron in kaolinite with special reference to kaolinite from tropical soils. Clay Miner. 11:201–220.

Hillier, S., and A.L. Pharande. 2008. Contemporary pedogenic formation of palygorskite in irrigation-induced saline–sodic, shrink–swell soils of Maharashtra, India. Clays Clay Miner. 56:531–548.

Hinsinger, P., O.N.F. Barros, M.F. Benedetti, Y. Noack, and G. Callot. 2001. Plant-induced weathering of a basaltic rock. Geochim. Cosmochim. Acta 65:137–152.

Hinsinger, P., F. Elsass, B. Jaillard, and M. Robert. 1993. Root-induced irreversible transformation of a trioctahedral mica in the rhizosphere of rape. J. Soil Sci. 44:535–545.

Hinsinger, P., and B. Jaillard. 1993. Root-induced release of interlayer potassium and vermiculitization of phlogopite as related to potassium depletion in the rhizosphere of ryegrass. J. Soil Sci. 44:525–534.

Hinsinger, P., B. Jaillard, and J.E. Dufey. 1992. Rapid weathering of a trioctahedral mica by the roots of ryegrass. Soil Sci. Soc. Am. J. 56:977–982.

Hiradate, S., H. Hirai, and H. Hashimoto. 2006. Characterisation of allophanic Andisols by solid-state ^{13}C, ^{27}Al, and ^{29}Si NMR and by C stable isotopic ratio, δ^{13}C. Geoderma 136:696–707.

Hiradate, S., T. Nakadai, H. Shindo, and T. Yoneyama. 2004. Carbon source of humic substances in some Japanese volcanic ash soils determined by carbon stable isotopic ratio, δ^{13}C. Geoderma 119:133–141.

Hiradate, S., and S.-I. Wada. 2005. Weathering processes of volcanic glass to allophane determined by ^{27}Al and ^{29}Si solid-state NMR. Clays Clay Miner. 53:401–408.

Hochella, M.F., Jr. 2008. Nanogeoscience: From origins to cutting-edge applications. Elements 4:373–379.

Hochella, M.F., Jr., and J.F. Banfield. 1995. Chemical weathering of silicates in nature: A microscopic perspective with theoretical considerations, p. 353–406. *In* A.F. White and S.L. Brantley (eds.) Chemical weathering rates of silicate minerals. Reviews in mineralogy. Vol. 31. Mineralogical Society of America, Washington, DC.

Hodder, A.P.W., P.J. de Lange, and D.J. Lowe. 1991. Dissolution and depletion of ferromagnesian minerals from Holocene tephras in an acid bog, New Zealand, and implications for tephra correlation. J. Quat. Sci. 6:195–208.

Hodder, A.P.W., B.E. Green, and D.J. Lowe. 1990. A two-stage model for the formation of clay minerals from tephra-derived volcanic glass. Clay Miner. 25:313–327.

Hodder, A.P.W., T.R. Naish, and D.J. Lowe. 1996. Towards an understanding of thermodynamic and kinetic controls on the formation of clay minerals from volcanic glass under various environmental conditions, p. 1–11. *In* S.G. Pandalai (ed.) Recent research developments in chemical geology. Research Signpost, Trivandrum, India.

Hodder, A.P.W., T.R. Naish, and C.S. Nelson. 1993. A two-stage model for the formation of smectite from detrital volcanic glass under shallow-marine conditions. Mar. Geol. 109:279–285.

Hogg, A.G., T.F.G. Higham, D.J. Lowe, J.G. Palmer, P. Reimer, and R.M. Newnham. 2003. A wiggle-match date for Polynesian settlement of New Zealand. Antiquity 77:116–125.

Hogg, A.G., D.J. Lowe, J.G. Palmer, G. Boswijk, C. Bronk Ramsey, and R. Sparks. 2011. Definitive high-precision calender date for the Taupo eruption derived by [14]C Wiggle-match dating using a New Zealand kauri-derived ^{14}C calibration curve. The Holocene (in review).

Holdren, G.R., Jr., and P.M. Speyer. 1987. Reaction rate-surface area relationships during the early stages of weathering: II Data on eight additional feldspars. Geochim. Cosmochim. Acta 51:2311–2318.

Hong, H.-L., and J.-X. Mi. 2006. Characteristics of halloysite associated with rectorite from Hubei, China. Mineral. Mag. 70:257–264.

Hseu, Z.Y., H. Tsai, H.C. Hsi, and Y.C. Chen. 2007. Weathering sequences of clay minerals in soils along a serpentinitic toposequence. Clays Clay Miner. 55:389–401.

Huang, P.M. 1989. Feldspars, olivines, pyroxenes, and amphiboles, p. 975–1050. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Huang, P.M., and M. Schnitzer (eds.). 1986. Interactions of soil minerals with natural organics and microbes. SSSA Special Publication No. 17. SSSA, Madison, WI.

Huang, P.M., M.K. Wang, N. Kämpf, and D.G. Schulze. 2002. Aluminum hydroxides, p. 261–290. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Hue, N.V., G.R. Craddock, and F. Adams. 1986. Effect of organic acids on aluminum toxicity in subsoils. Soil Sci. Soc. Am. J. 50:28–34.

Hughes, J.C., and G. Brown. 1979. A crystallinity index for soil kaolins and its relation to parent rock, climate and soil maturity. J. Soil Sci. 30:557–563.

Hunter, R.J. 1981. The zeta potential in colloid science. Academic Press, London, U.K.

Hurowitz, J.A., and S.M. McLennan. 2007. A ~3.5 Ga record of water-limited, acidic weathering conditions on Mars. Earth Planet Sci. Lett. 260:432–443.

Ildefonse, P., E. Copin, and B. Velde. 1979. A soil vermiculite formed from a meta-gabbro, Liore-Atlantique, France. Clay Miner. 14:201–210.

Inoue, K., and T. Sase. 1996. Paleoenvironmental history of post-Toya Ash tephric deposits and paleosols at Iwate Volcano, Japan, using aeolian dust content and phytolith composition. Quat. Int. 34–36:127–137.

Inskeep, W.P., J.L. Clayton, and D.W. Mogk. 1993. Naturally weathered plagioclase grains from the Idaho Batholith: Observations using scanning electron microscopy. Soil Sci. Soc. Am. J. 57:851–860.

Jackson, M.L. 1959. Frequency distribution of clay minerals in major great soil groups as related to the factors of soil formation. Clays Clay Miner. 6:133–143.

Jackson, M.L. 1963. Aluminum bonding in soils: A unifying principle in soil science. Soil Sci. Soc. Am. Proc. 27:1–10.

Jackson, M.L. 1964. Chemical composition of soils, p. 71–141. *In* F.E. Bear (ed.) Chemistry of the soil. Reinhold Publishing Corporation, New York.

Jackson, M.L., Y. Hseung, R.B. Corey, E.J. Evans, and R.C. Van den Heuval. 1952. Weathering of clay-size minerals in soils and sediments II. Chemical weathering of layer silicates. Soil Sci. Soc. Am. Proc. 16:3–6.

Jackson, M.L., S.A. Tyler, A.L. Willis, G.A. Bourbeau, and R.P. Pennington. 1948. Weathering sequence of clay-size minerals in soils and sediments. I. Fundamental generalizations. J. Phys. Colloid Chem. 52:1237–1260.

Janik, L.J., and J.L. Keeling. 1993. FT-IR partial least-squares analysis of tubular halloysite in kaolin samples from the Mount Hope kaolin deposit. Clay Miner. 28:265–378.

Jansen, B., G.J. Nierop, and J.M. Verstraten. 2005. Mechanisms controlling the mobility of dissolved organic matter, aluminium and iron in podzol B horizons. Eur. J. Soil Sci. 56:537–550.

Jaynes, W.F., J.M. Bigham, N.E. Smeck, and M.J. Shipitalo. 1989. Interstratified 1:1–2:1 mineral formation in a polygenetic soil from southern Ohio. Soil Sci. Soc. Am. J. 53:1888–1894.

Johnson, L.J. 1964. Occurrence of regularly interstratified chlorite–vermiculite as a weathering product of chlorite in a soil. Am. Mineral. 49:556–572.

Johnson, D.L. 2002. Darwin would be proud: Bioturbation, dynamic denudation, and the power of theory in science. Geoarchaeology 17:7–40.

Johnson, S.L., S. Guggenheim, and A.F. Koster van Groos. 1990. Thermal stability of halloysite by high pressure differential thermal analysis. Clays Clay Miner. 38:477–484.

Johnson-Maynard, J.L., P.A. McDaniel, D.E. Ferguson, and A.L. Falen. 1997. Chemical and mineralogical conversion of Andisols following invasion by bracken fern. Soil Sci. Soc. Am. J. 61:549–555.

Jolicoeur, S., P. Ildefonse, and M. Bouchard. 2000. Kaolinite and gibbsite weathering of biotite within saprolites and soils of central Virginia. Soil Sci. Soc. Am. J. 64:1118–1119.

Jongkind, A.G., and P. Buurman. 2006. The effect of kauri (*Agathis australis*) on grain size distribution and clay mineralogy of andesitic soils in the Waitakere Ranges, New Zealand. Geoderma 134:171–186.

Jongmans, A.G., N. Van Breeman, U. Lundström, P.A.W. van Hees, R.D. Finlay, M. Srinivasan, T. Unestam, R. Giesler, P.-A. Melkerud, and M. Olsson. 1997. Rock-eating fungi. Nature 389:682–683.

Jongmans, A.G., F. Van Oort, P. Buurman, A.M. Jaunet, and J.D.J. van Doesburg. 1994. Morphology, chemistry and mineralogy of isotropic aluminosilicate coatings in a Guadeloupe Andisol. Soil Sci. Soc. Am. J. 58:501–507.

Jouquet, P., L. Mamou, M. Lepage, and B. Velde. 2002. Effect of termites on clay minerals in tropical soils: Fungus-growing termites as weathering agents. Eur. J. Soil Sci. 53:521–528.

Joussein, E., S. Petit, G.J. Churchman, B. Theng, D. Righi, and B. Delvaux. 2005. Halloysite clay minerals—A review. Clay Miner. 40:383–426.

Kadir, S., and M. Eren. 2008. The occurrence and genesis of clay minerals associated with Quaternary caliches in the Mersin area, southern Turkey. Clays Clay Miner. 56:244–258.

Kahle, M., M. Kleber, and R. Jahn. 2002a. Review of XRD-based quantitative analyses of clay minerals in soils: The suitability of mineral intensity factors. Geoderma 109:191–205.

Kahle, M., M. Kleber, and R. Jahn. 2002b. Predicting carbon content in illitic clay fractions from surface area, cation exchange capacity and dithionite-extractable iron. Eur. J. Soil Sci. 53:639–644.

Kahle, M., M. Kleber, and R. Jahn. 2004. Retention of dissolved organic matter by phyllosilicate and soil clay fractions in relation to mineral properties. Org. Geochem. 35:269–276.

Kaiser, K., and G. Guggenberger. 2000. The role of DOM sorption to mineral surfaces in the preservation of organic matter in soils. Org. Geochem. 31:711–724.

Kaiser, K., and G. Guggenberger. 2003. Mineral surfaces and organic matter. Eur. J. Soil Sci. 54:219–236.

Kämpf, N., and U. Schwertmann. 1982. Goethite and hematite in a climosequence in southern Brazil and their application in classification of kaolinitic soils. Geoderma 29:27–39.

Kanket, W., A. Suddhiprakern, I. Kheoruenromne, and R.J. Gilkes. 2005. Chemical and crystallographic properties of kaolin from ultisols in Thailand. Clays Clay Miner. 53:478–489.

Kapoor, B.S. 1973. The formation of 2:1–2:2 intergrade clays in some Norwegian podzols. Clay Miner. 10:79–86.

Kautz, C.Q., and P.C. Ryan. 2003. The 10 Å to 7 Å halloysite transition in a tropical soil sequence, Costa Rica. Clays Clay Miner. 51:252–263.

Kawano, M., and K. Tomita. 1994. Growth of smectite from leached layer during experimental alteration of albite. Clays Clay Miner. 42:7–17.

Kawano, M., and K. Tomita. 2001. Microbial biomineralization in weathered volcanic ash deposit and formation of biogenic minerals by experimental incubation. Am. Mineral. 86:400–410.

Kawano, M., and K. Tomita. 2002. Microbial formation of silicate minerals in the weathering environment of a pyroclastic deposit. Clays Clay Miner. 50:99–110.

Kendrick, K.J., and R.C. Graham. 2004. Pedogenic silica accumulation in chronosequence soils, southern California. Soil Sci. Soc. Am. J. 68:1295–1303.

Khademi, H., and J.M. Arocena. 2008. Kaolinite formation from palygorskite and sepiolite in rhizosphere soils. Clays Clay Miner. 56:429–436.

Khormali, F., and A. Abtahi. 2003. Origin and distribution of clay minerals in calcareous arid and semi-arid soils of Fars province, southern Iran. Clay Miner. 38:511–527.

Kirkman, J.H. 1976. Clay mineralogy of Rotomahana sandy loam soil, North Island, New Zealand. N.Z. J. Geol. Geophys. 19:35–41.

Kirkman, J.H. 1981. Morphology and structure of halloysite in New Zealand tephras. Clays Clay Miner. 29:1–9.

Kirkman, J.H., and W.J. McHardy. 1980. A comparative study of the morphology, chemical composition and weathering of rhyolitic and andesitic glass. Clay Miner. 15:165–173.

Kittrick, J.A. 1967. Gibbsite–kaolinite equilibria. Soil Sci. Soc. Am. Proc. 31:314–316.

Kittrick, J.A. 1973. Mica-derived vermiculites as unstable intermediates. Clays Clay Miner. 21:479–488.

Kleber, M., R. Mikutta, M.S. Torn, and R. Jahn. 2005. Poorly crystalline mineral phases protect organic matter in acid subsoil horizons. Eur. J. Soil Sci. 56:717–725.

Kleber, M., L. Schwendenmann, E. Veldkamp, J. Röbner, and R. Jahn. 2007. Halloysite versus gibbsite: Silicon cycling as a pedogenetic process in two lowland rain forest soils of La Selva, Costa Rica. Geoderma 138:1–11.

Kohut, C., K. Muehlenbachs, and M.J. Dudas. 1995. Authigenic dolomite in a saline soil in Alberta, Canada. Soil Sci. Soc. Am. J. 59:1499–1504.

Kondo, R., C.W. Childs, and I. Atkinson. 1994. Opal phytoliths in New Zealand. Manaaki Whenua Press, Lincoln, New Zealand.

Laird, D.A., P. Barak, E.A. Nater, and R.H. Dowdy. 1991. Chemistry of smectitic and illitic phases in interstratified soil smectite. Soil Sci. Soc. Am. J. 55:1499–1504.

Laird, D.A., and E.A. Nater. 1993. Nature of the illitic phase associated with randomly interstratified smectite/illite in soils. Clays Clay Miner. 41:280–287.

Langmuir, D., and D.O. Whittemore. 1971. Variations in the stability of precipitated ferric oxyhydroxides, p. 209–234. *In* R.F. Gould (ed.) Nonequilibrium systems in natural water chemistry. Advances in Chemistry Series No. 106. American Chemical Society, Washington, DC.

Lanson, B. 1997. Decomposition of experimental X-ray diffraction patterns (profile fitting): A convenient way to study clays. Clays Clay Miner. 45:132–146.

Lapidus, D.F. 1987. Collins dictionary of geology. Collins, London, U.K.

Lee, B.D., S.K. Sears, R.C. Graham, C. Amrhein, and H. Vali. 2003. Secondary mineral genesis from chlorite and serpentine in an ultramafic soil toposequence. Soil Sci. Soc. Am. J. 65:1183–1196.

Leifeld, J. 2006. Soils as sources and sinks of greenhouse gases, p. 23–44. Geological Society of London Special Publication No. 266.

Leonard, R.A., and S.B. Weed. 1967. Influence of exchange ions on the b-dimension of dioctahedral vermiculite. Clays Clay Miner. 15:149–161.

Li, Z., B. Velde, and D. Li. 2003. Loss of K-bearing clay minerals in flood-irrigated, rice-growing soils in Jiangxi province, China. Clays Clay Miner. 51:75–82.

Lilienfein, J., R.G. Qualls, S.M. Uselman, and S.D. Bridgham. 2003. Soil formation and organic matter accretion in a young andesitic chronosequence at Mt. Shasta, California. Geoderma 116:249–264.

Lim, C.H., M.L. Jackson, R.D. Koons, and P.A. Helmke. 1980. Kaolins: Sources of differences in cation-exchange capacities and cesium retention. Clays Clay Miner. 28:223–229.

Lin, C.-W., Z.-Y. Hseu, and Z.-S. Chen. 2002. Clay mineralogy of Spodosols with high clay contents in the subalpine forests of Taiwan. Clays Clay Miner. 50:726–735.

Lindsay, W.L., P.L.G. Vlek, and S.H. Chien. 1989. Phosphate minerals, p. 1089–1130. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Loughnan, F.C. 1969. Chemical weathering of the silicate minerals. Elsevier, New York.

Loveland, P.J. 1984. The soil clays of Great Britain: 1. England and Wales. Clay Miner. 19:681–707.

Loveland, P.J., and P. Bullock. 1975. Crystalline and amorphous components of the clay fractions in brown podzolic soils. Clay Miner. 10:451–469.

Loveland, P.J., I.G. Wood, and A.H. Weir. 1999. Clay mineralogy at Rothamsted: 1934–1988. Clay Miner. 34:165–183.

Lowe, D.J. 1986. Controls on the rates of weathering and clay mineral genesis in airfall tephras: A review and New Zealand case study, p. 265–330. *In* S.M. Colman and D.P. Dethier (eds.) Rates of chemical weathering of rocks and minerals. Academic Press, Orlando, FL.

Lowe, D.J. 1995. Teaching clays: From ashes to allophane, p. 19–23. *In* G.J. Churchman, R.W. Fitzpatrick, and R.A. Eggleton (eds.) Clays: Controlling the environment. Proc. 10th Int. Clay Conf., 18–23 July 1993, Adelaide, Australia. CSIRO Publishing, Melbourne, Australia.

Lowe, D.J. (ed.). 2008. Guidebook for pre-conference North Island field trip A1 "ashes and issues." Australian and New Zealand 4th Joint Soils Conf., Massey University, Palmerston North, p. 1–194. New Zealand Society of Soil Science, Christchurch, New Zealand (ISBN 978-0-473-14476-0).

Lowe, D.J., G.J. Churchman, R.H. Merry, R.W. Fitzpatrick, M.J. Sheard, and W.H. Hudnall. 1996. Holocene basaltic volcanogenic soils of the Mt. Gambier area, South Australia, are unusual globally: What do they tell us? p. 153–154. Australian and New Zealand National Soils Conf. 1996. Vol. 2. Oral papers. Australian Society of Soil Science, University of Melbourne, Australia.

Lowe, D.J., and D.J. Palmer. 2005. Andisols of New Zealand and Australia. J. Integr. Field Sci. 2:39–65.

Lowe, D.J., and H.J. Percival. 1993. Clay mineralogy of tephras and associated paleosols and soils, and hydrothermal deposits, North Island. Guidebook for New Zealand pre-conference field trip F1, p. 1–110. 10th Int. Clay Conf., 18–23 July 1993, Adelaide, Australia.

Lowe, D.J., P.A.R. Shane, B.V. Alloway, and R.M. Newnham. 2008. Fingerprints and age models for widespread New Zealand tephra marker beds erupted since 30,000 years ago: A framework for NZ-INTIMATE. Quat. Sci. Rev. 27:95–126.

Lowe, D.J., J.M. Tippett, P.J.J. Kamp, I.J. Liddell, R.M. Briggs, and J.L. Horrocks. 2001. Ages on weathered Plio-Pleistocene tephra sequences, western North Island, New Zealand. Les Dossiers de l'Archeo-Logis 1:45–60.

Lucas, Y. 2001. The role of plant in controlling rates and products of weathering: Importance of biological pumping. Annu. Rev. Earth Planet. Sci. 29:135–163.

Lucas, Y., F.J. Luizao, A. Chauvel, J. Rouiller, and D. Nahon. 1993. The relation between biological activity of the rain forest and mineral composition of soils. Science 260:521–523.

Lundström, U.S., N. Van Breeman, and D. Bain. 2000. The podzolisation process. A review. Geoderma 94:91–107.

Ma, C., and R.A. Eggleton. 1999. Surface layer types of kaolinite: A high-resolution transmission electron microscope study. Clays Clay Miner. 47:181–191.

Madden, M.E.M., R.J. Bodnar, and J.D. Rimstidt. 2004. Jarosite as an indicator of water-limited chemical weathering on Mars. Nature 431:821–823.

Maher, K., C.I. Steefel, A.F. White, and D.A. Stonestrom. 2009. The role of reaction affinity and secondary minerals in regulating chemical weathering rates at the Santa Cruz Soil Chronosequence, California. Geochim. Cosmochim. Acta 73:2804–2831.

Maher, B.A., and R.M. Taylor. 1988. Formation of ultrafine-grained magnetite in soils. Nature 336:368–370.

Mahjoory, R.A. 1975. Clay mineralogy, physical, and chemical properties of some soils in arid regions of Iran. Soil Sci. Soc. Am. Proc. 39:1157–1164.

Marsan, F.A., and J. Torrent. 1989. Fragipan bonding by silica and iron oxides in a soil from northwestern Italy. Soil Sci. Soc. Am. J. 53:1140–1145.

Marx, S.K., H.A. McGowan, and B.S. Kamber. 2009. Long-range dust transport from eastern Australia: A proxy for Holocene aridity and ENSO-type climate variability. Earth Planet Sci. Lett. 282:167–177.

Mayer, L.M., and B. Xing. 2001. Organic matter-surface area relationships in acid soils. Soil Sci. Soc. Am. J. 65:250–258.

McCarthy, J.F., J. Ilavsky, J.D. Jastrow, L.M. Mayer, E. Perfect, and J. Zhuang. 2008. Protection of organic matter in soil microaggregates via restructuring of aggregate porosity and filling of pores with accumulating organic matter. Geochim. Cosmochim. Acta 72:4725–4744.

McDaniel, P.A., D.J. Lowe, O. Arnalds, and C.-L. Ping. 2012. Andisols. *In* Li, Y. and Sumner, M.E. (eds.) Handbook of soil science. 2nd edn. CRC Press (Taylor & Francis), London, U.K.

McGahan, D.G., R.J. Southard, and R.J. Zasoki. 2003. Mineralogical comparison of agriculturally acidified and naturally acidic soils. Geoderma 114:355–368.

McIntosh, P.D. 1979. Halloysite in a New Zealand tephra and paleosol less than 2500 years old. N.Z. J. Sci. 22:49–54.

McIntosh, P.D. 1980. Weathering products in Vitrandept profiles under pine and manuka, New Zealand. Geoderma 24:225–239.

McKeague, J.A., and J.E. Brydon. 1970. Mineralogical properties of ten reddish brown soils from the Atlantic provinces in relation to parent materials and pedogenesis. Can. J. Soil Sci. 50:47–55.

McKenzie, R.M. 1989. Manganese oxides and hydroxides, p. 439–465. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Mehra, O.P., and M.L. Jackson. 1960. Iron oxide removal from soils and clays by a dithionite-citrate system buffered with sodium bicarbonate. Clays Clay Miner. 7:317–327.

Mejia, G., H. Kohnke, and J.L. White. 1968. Clay mineralogy of certain soils of Columbia. Soil Sci. Soc. Am. Proc. 32:665–670.

Melkerud, P.-A., D.C. Bain, A.G. Jongmans, and T. Tarvainen. 2000. Chemical, mineralogical and morphological characterization of three podzols developed on glacial deposits in northern Europe. Geoderma 94:125–148.

Melo, V.F., B. Singh, C.E.G.R. Schaefer, R.F. Novais, and M.P.F. Fontes. 2001. Chemical and mineralogical properties of kaolinite-rich Brazilian soils. Soil Sci. Soc. Am. J. 65:1324–1333.

Mestdagh, M.M., L. Vielvoye, and A.J. Herbillon. 1980. Iron in kaolinite: II. The relationship between kaolinite crystallinity and iron content. Clay Miner. 15:1–13.

Meunier, A., and B. Velde. 1976. Mineral reactions at grain contacts in early stages of granite weathering. Clay Miner. 11:235–240.

Meunier, A., and B. Velde. 1979. Weathering mineral facies in altered granites: The importance of local small-scale equilibria. Mineral Mag. 43:261–268.

Meysman, F.J.R., J.J. Middelburg, and C.H.R. Heip. 2006. Bioturbation: A fresh look at Darwin's last idea. Trends Ecol. Evol. 21:688–695.

Mikutta, R., M. Kleber, and R. Jahn. 2005. Poorly crystalline minerals protect organic carbon in clay subfractions from acid subsoil horizons. Geoderma 128:106–115.

Millot, G. 1970. Geology of clays. Springer-Verlag, New York.

Mills, A.L. 2003. Keeping in touch: Microbial life on soil particle surfaces. Adv. Agron. 78:1–43.

Milnes, A.R., and R.W. Fitzpatrick. 1989. Titanium and zirconium minerals, p. 1131–1205. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Ming, D.W., and F.A. Mumpton. 1989. Zeolites in soils, p. 873–911. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Mirabella, A., and M. Egli. 2003. Structural transformations of clay minerals in soils of a climosequence in an Italian alpine environment. Clays Clay Miner. 51:264–278.

Mirabella, A., M. Egli, S. Raimondi, and D. Giaccai. 2005. Origin of clay minerals in soils on pyroclastic deposits in the island of Lipari (Italy). Clays Clay Miner. 53:409–421.

Mizota, C., and M. Itoh. 1993. Volcanic origin of a cristobalite in the Te Ngae tephric loess from North Island, New Zealand. Clays Clay Miner. 41:755–756.

Mizota, C., and Y. Takahashi. 1982. Eolian origin of quartz and mica in soils developed on basalts in northwestern Kyushu and Sanoin, Japan. Soil Sci. Plant Nutr. (Tokyo) 28:369–378.

Mizota, C., N. Toh, and Y. Matsuhisa. 1987. Origin of cristobalite in soils derived from volcanic ash in temperate and tropical regions. Geoderma 39:323–330.

Mogk, D.W., and W.W. Locke, III. 1988. Application of Auger electron spectroscopy (AES) to naturally weathered hornblende. Geochim. Cosmochim. Acta 52:2537–2542.

Mokma, D.L., M. Yli-Halla, and K. Lindqvist. 2004. Podzol formation in sandy soils of Finland. Geoderma 120:259–272.

Monger, H.C., and E.F. Kelly. 2002. Silica minerals, p. 611–636. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Morgan, C.P., and M.H. Stolt. 2006. Soil morphology—water table cumulative duration relationships in southern New England. Soil Sci. Soc. Am. J. 70:816–824.

Mortland, M.M. 1958. Kinetics of potassium release from biotite. Soil Sci. Soc. Am. Proc. 22:503–508.

Mortland, M.M., K. Lawton, and G. Uehara. 1956. Alteration of biotite to vermiculite by plant growth. Soil Sci. 82:477–481.

Mossin, L., M. Mortensen, and P. Nørnberg. 2002. Imogolite related to podzolization processes in Danish podzols. Geoderma 109:103–116.

Muir, I.J., G.M. Bancroft, and H.W. Nesbitt. 1989. Characteristics of altered labradorite surfaces by SIMS and XPS. Geochim. Cosmochim. Acta 53:1235–1241.

Muir, I.J., G.M. Bancroft, W. Shotyk, and H.W. Nesbitt. 1990. A SIMS and XPS study of dissolving plagioclase. Geochim. Cosmochim. Acta 54:2247–2256.

Muir, I.J., and H.W. Nesbitt. 1997. Reactions of aqueous anions and cations at the labradorite–water interface: Coupled effects of surface processes and diffusion. Geochim. Cosmochim. Acta 61:265–274.

Müller, B. 2009. Impact of the bacterium *Pseudomonas fluorescens* and its genetic derivatives on vermiculite: Effects on trace metals contents and clay mineralogical properties. Geoderma 153:94–103.

Munk, L.P., and R.J. Southard. 1993. Pedogenic implications of opaline pendants in some California late-Pleistocene Palexeralfs. Soil Sci. Soc. Am. J. 57:149–154.

Murakami, T., H. Isobe, T. Sato, and T. Ohnuki. 1996. Weathering of chlorite in a quartz–chlorite schist: I. Mineralogical and chemical changes. Clays Clay Miner. 44:244–256.

Mustard, J.F., S.L. Murchie, S.M. Pelkey, B.L. Ehlmann, R.E. Milliken, J.A. Grant, J.-P. Bibring, et al. 2008. Hydrated silicate minerals on Mars observed by the mars reconnaissance orbiter CRISM instrument. Nature 454:305–309.

Nadeau, P.H., M.J. Wilson, W.J. McHardy, and J.M. Tait. 1984. Interparticle diffraction: A new concept for interstratification of clay minerals. Clay Miner. 19:757–769.

Nagashima, K., R. Tada, A. Tani, S. Toyoda, Y. Sun, and Y. Isozaki. 2007. Contribution of aeolian dust in Japan sea sediments estimated from ESR signal intensity and crystallinity of quartz. Geochem. Geophys. Geosyst. 8:Q02Q04. doi:10.1029/2006GC001364.

Nahon, D., and F. Colin. 1982. Chemical weathering of orthopyroxenes under lateritic conditions. Am. J. Sci. 282:1232–1243.

Nahon, D., F. Colin, and Y. Tardy. 1982. Formation and distribution of Mg, Fe, Mn–smectites in the first stages of the lateritic weathering of forsterite and tephroite. Clay Miner. 17:339–348.

Naish, T.R., C.S. Nelson, and A.P.W. Hodder. 1993. Evolution of Holocene sedimentary bentonite in a shallow-marine embayment, Firth of Thames, New Zealand. Mar. Geol. 109:267–278.

Nakao, A., S. Funakawa, T. Watanebe, and T. Kosaki. 2009. Pedogenic alterations of illitic minerals represented by Radiocaesium Interception Potential in soils with different soil moisture regimes in humid Asia. Eur. J. Soil Sci. 60:139–152.

Nanzyo, M. 2007. Introduction to studies on volcanic ash soils in Japan and international collaboration. J. Integr. Field Sci. 4:71–77.

Ndayiragije, S., and B. Delvaux. 2003. Coexistence of allophane, gibbsite, kaolinite and hydroxyl-Al-interlayered 2:1 clay minerals in a perudic Andosol. Geoderma 117:203–214.

Neall, V.E. 1977. Genesis and weathering of Andosols in Taranaki, New Zealand. Soil Sci. 123:400–408.

Nesbitt, H.W. 1997. Bacterial and inorganic weathering processes and weathering of crystalline rocks, p. 113–142. *In* J.M. McIntosh and L.A. Groat (eds.) Biological–mineralogical interactions. Mineralogical Association of Canada. Short Course. Vol. 25. Mineralogical Association of Canada, Ottawa, Canada.

Nettleton, W.D., K.W. Flach, and R.E. Nelson. 1970. Pedogenic weathering of tonalite in southern California. Geoderma 4:387–402.

Newnham, R.M., D.J. Lowe, and P.W. Williams. 1999. Quaternary environmental change in New Zealand: A review. Prog. Phys. Geog. 23:567–610.

Newsom, H. 2005. Clays in the history of Mars. Nature 438:570–571.

Nieuwenhuyse, A., P.S.J. Verburg, and A.G. Jongmans. 2000. Mineralogy of a soil chronosequence on andesitic lava in humid tropical Costa Rica. Geoderma 98:61–82.

Nizeyimana, E., T.J. Bicki, and P.A. Agbu. 1997. An assessment of colloidal constituents and clay mineralogy of soils derived from volcanic materials along a toposequence in Rwanda. Soil Sci. 162:361–371.

Norfleet, M.L., A.D. Karathanasis, and B.R. Smith. 1993. Soil solution composition relative to mineral distribution in Blue Ridge mountain soils. Soil Sci. Soc. Am. J. 57:1375–1380.

Noro, H. 1986. Hexagonal platy halloysite in an altered tuff bed, Komaki City, Aichi prefecture, Central Japan. Clay Miner. 21:401–415.

Norrish, K. 1968. Some phosphate minerals of soils, p. 713–723. Trans. 9th Int. Congr. Soil Sci., 6–16 August 1968, Adelaide, Australia.

Norrish, K. 1973. Factors in the weathering of mica to vermiculite, p. 417–432. *In* J.M. Serratosa (ed.) Proc. 1972 Int. Clay Conf., Div. de Ciencas., 25–30 June 1972, Madrid, Spain.

Norrish, K., and J.G. Pickering. 1983. Clay minerals, p. 281–308. *In* Soils: An Australian viewpoint. CSIRO, Melbourne, Australia.

Norrish, K., and H. Rosser. 1983. Mineral phosphate, p. 335–361. *In* Soils: An Australian viewpoint. CSIRO, Melbourne, Australia.

Oades, J.M., and A.G. Waters. 1991. Aggregate hierarchy in soils. Aust. J. Soil Res. 29:815–828.

Oelkers, E.H., and S.R. Gislason. 2001. The mechanism, rates and consequences of basaltic glass dissolution: I. An experimental study of the dissolution rates of basaltic glass as a function of aqueous Al, Si and oxalic acid concentration at 25°C and pH = 3 and 11. Geochim. Cosmochim. Acta 65:3671–3681.

Oelkers, E.H., and J. Schott. 1995. Experimental study of anorthite dissolution and the relative mechanism of feldspar hydrolysis. Geochim. Cosmochim. Acta 59:5039–5053.

Ogura, Y., R. Tanaka, and H. Takesako. 2008. Unique clay mineral formation in Andisols derived from Holocene tephras of Mt Fuji, Japan, p. 105. Abstracts, Joint Conf. Australia and New Zealand Soc. Soil Sci., 1–5 December 2008, Palmerston North, New Zealand.

Ojanuga, A.G. 1973. Weathering of biotite in soils of a humid tropical climate. Soil Sci. Soc. Am. Proc. 37:644–646.

Owliaie, H.R., A. Abtahi, and R.J. Heck. 2006. Pedogenesis and clay mineralogical investigation of soils formed on gypsiferous and calcareous materials, on a transect, southwestern Iran. Geoderma 134:62–81.

Pai, C.W., M.K. Wang, and C.Y. Chiu. 2007. Clay mineralogical characterization of a toposequence of perhumid subalpine forest soils in northeastern Taiwan. Geoderma 138:177–184.

Pai, C.W., M.K. Wang, H.B. King, C.Y. Chiu, and J.-L. Hwong. 2004. Hydroxy-interlayered mineral of forest soils. Geoderma 123:245–255.

Papoulis, D., P. Tsolis-Katagas, and C. Katagas. 2004. Progressive stages in the formation of kaolin minerals of different morphologies in the weathering of plagioclase. Clays Clay Miner. 52:275–286.

Paquet, H., and G. Millot. 1972. Geochemical evolution of clay minerals in the weathered products in soils of Mediterranean climate, p. 199–206. *In* J.M. Serratosa (ed.) Proc. Int. Clay Conf., 25–30 June 1972, Madrid, Spain.

Parfitt, R.L. 1990. Allophane in New Zealand—A review. Aust. J. Soil Res. 28:343–360.

Parfitt, R.L. 2009. Allophane and imogolite: Role in soil biogeochemical processes. Clay Miner. 44:135–155.

Parfitt, R.L., and G.J. Churchman. 1988. Clay minerals and humus complexes in five Kenyan soils derived from volcanic ash—A discussion. Geoderma 42:365–367.

Parfitt, R.L., and T. Henmi. 1980. Structure of some allophanes from New Zealand. Clays Clay Miner. 28:285–294.

Parfitt, R.L., M. Russell, and G.E. Orbell. 1983. Weathering sequence of soils from volcanic ash involving allophane and halloysite, New Zealand. Geoderma 29:41–57.

Parfitt, R.L., M. Saigusa, and J.D. Cowie. 1984. Allophane and halloysite formation in a volcanic ash bed under differing moisture conditions. Soil Sci. 138:360–364.

Parfitt, R.L., and A.D. Wilson. 1985. Estimation of allophane and halloysite in three sequences of volcanic soils, New Zealand. Catena Suppl. 7:1–8.

Paton, T.R., G.S. Humphreys, and P.B. Mitchell. 1995. Soils: A new global view. UCL Press, London, U.K.

Pauling, L. 1929. The principles determining the structure of complex ionic crystals. J. Am. Chem. Soc. 51:1010–1026.

Percival, H.J. 1985. Soil solutions, minerals, and equilibria. New Zealand Soil Bureau Scientific Report 69. Department of Scientific and Industrial Research, Lower Hutt, New Zealand.

Percival, H.J., R.L. Parfitt, and N.A. Scott. 2000. Factors controlling soil carbon levels in New Zealand grasslands: Is clay content important? Soil Sci. Soc. Am. J. 64:1623–1630.

Pernes-Debuyser, A., M. Pernes, B. Velde, and D. Tessier. 2003. Soil mineralogy evolution in the INRA 42 plots experiment. Clays Clay Miner. 51:577–584.

Petit, S., T. Prot, A. Decarreau, C. Mosser, and M.C. Toledo-Groke. 1992. Crystallochemical study of a population of particles in smectites from a lateritic weathering profile. Clays Clay Miner. 40:436–445.

Pochet, G., M. Van der Velde, M. Vanclooster, and B. Delvaux. 2007. Hydric properties of high charge, halloysite clay soils from the tropical South Pacific region. Geoderma 138:96–109.

Poncelot, G.M., and G.W. Brindley. 1967. Experimental formation of kaolinite from montmorillonite at low temperatures. Am. Mineral. 52:1161–1173.

Poulenard, J., P. Podwojewski, and A.J. Herbillon. 2003. Characteristics of non-allophanic Andisols with hydric properties from the Ecuadorian páramos. Geoderma 117:267–281.

Poulet, F., J.-P. Bibring, J.F. Mustard, A. Gendrin, N. Mangold, Y. Langevin, R.E. Arvidson, B. Gondet, and C. Gomez. 2005. Phyllosilicates on Mars and implications for the early Mars history. Nature 438:632–627.

Proust, D., J. Caillaud, and C. Fontaine. 2006. Clay minerals in early amphibole weathering: Tri- to dioctahedral sequence as a function of crystallization sites in the amphibole. Clays Clay Miner. 54:351–362.

Quantin, P., T. Becquer, J.H. Rouiller, and J. Berthelin. 2001. Oxide weathering and trace metal release by bacterial reduction in a New Caledonia Ferralsol. Biogeochemistry 53:323–340.

Quantin, P., A.J. Herbillon, C. Janot, and G. Siefferman. 1984. L' "halloysite" blanche riche en fer de Vate (Vanuatu); hypothese d'un edifice interstratifie halloysite–hisingerite. Clay Miner. 19:629–643.

Raman, K.V., and M.L. Jackson. 1966. Layer charge reduction in clay minerals of micaceous soils and sediments. Clays Clay Miner. 14:53–68.

Rampazzo, N., and W.E.H. Blum. 1992. Changes in chemistry and mineralogy of forest soils by acid rain. Water Air Soil Pollut. 61:209–220.

Rasmussen, C., N. Matsuyama, R.A. Dahlgren, R.J. Southard, and N. Brauer. 2007. Soil genesis and mineral transformation across an environmental gradient on andesitic lahar. Soil Sci. Soc. Am. J. 71:225–237.

Rasmussen, C., R.J. Southard, and W.R. Horwath. 2005. Soil mineralogy affects conifer forest soil carbon source utilization and microbial priming. Soil Sci. Soc. Am. J. 71:1141–1150.

Rebertus, R.A., S.B. Weed, and S.W. Buol. 1986. Transformations of biotite to kaolinite during saprolite-soil weathering. Soil Sci. Soc. Am. J. 50:810–819.

Reid, D.A., R.C. Graham, L.A. Douglas, and C. Amrhein. 1996. Smectite mineralogy and charge characteristics along an arid geomorphic transect. Soil Sci. Soc. Am. J. 60:1602–1611.

Rex, R.W., J.K. Syers, M.L. Jackson, and R.N. Clayton. 1969. Eolian origin of quartz in soils of Hawaiian islands and in Pacific pelagic sediments. Science 163:277–279.

Rice, C.M., and J.M. Williams. 1969. A Mössbauer study of biotite weathering. Miner. Mag. 37:210–215.

Rich, C.I. 1968. Hydroxy interlayers in expansible layer silicates. Clays Clay Miner. 16:15–30.

Richards, A.E., R.C. Dalal, and S. Schmidt. 2009. Carbon storage in a Ferrasol under subtropical rainforest, tea plantations, and pasture is linked to soil aggregation. Aust. J. Soil Res. 47:341–350.

Righi, D., K. Huber, and C. Keller. 1999. Clay formation and podzol development from postglacial moraines in Switzerland. Clay Miner. 34:319–332.

Righi, D., S. Petit, and A. Bouchet. 1993. Characterization of hydroxy-interlayered vermiculite and illite/smectite interstratified minerals from the weathering of a chlorite in a Cryorthod. Clays Clay Miner. 41:484–495.

Righi, D., F. Terribile, and S. Petit. 1995. Low-charge to high-charge beidellite conversion in a Vertisol from south Italy. Clays Clay Miner. 43:495–502.

Ritsema, C.J., M.E.F. van Mensvoort, D.L. Dent, Y. Tan, H. van den Bosch, and A.L.M. van Wijk. 2000. Acid sulfate soils, p. G121–G154. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Rivers, J.M., J.E. Nyquist, Y. Roh, D.O. Terry, Jr., and W.E. Doll. 2004. Investigation into the origin of magnetic soils on the Oak Ridge reservation, Tennessee. Soil Sci. Soc. Am. J. 68:1772–1779.

Robert, M., and J. Berthelin. 1986. Role of biological and biochemical factors in soil mineral weathering, p. 453–495. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and microbes. SSSA Special Publication No. 17. SSSA, Madison, WI.

Robert, M., and C. Chenu. 1992. Interactions between soil minerals and microorganisms, p. 307–404. *In* G. Stotsky and J.-M. Bollag (eds.) Soil biochemistry. Marcel Dekker, New York.

Robert, M., M. Hardy, and F. Elsass. 1991. Crystallochemistry, properties and organization of soil clays derived from major sedimentary rocks in France. Clay Miner. 26:409–420.

Robertson, I.D.M., and R.A. Eggleton. 1991. Weathering of granitic muscovite to kaolinite and halloysite and of plagioclase-derived kaolinite to halloysite. Clays Clay Miner. 39:113–126.

Robinson, D.A., I. Lebro, and H. Vereecken. 2009. On the definition of the natural capital of soils: A framework for description, evaluation, and monitoring. Soil Sci. Soc. Am. J. 73:1904–1911.

Ross, G.J., and M.M. Mortland. 1966. A soil beidellite. Soil Sci. Soc. Am. Proc. 39:337–343.

Ross, G.J., C. Wang, A.I. Ozkan, and H.W. Rees. 1982. Weathering of chlorite and mica in New Brunswick podzol developed on till derived from chlorite–mica schist. Geoderma 27:255–267.

Russell, E.W. 1973. Soil conditions and plant growth. 10th edn. Longman, London, U.K.

Ruxton, B.P. 1988. Towards a weathering model of Mount Lamington Ash, New Guinea. Earth Sci. Rev. 25:387–397.

Ryan, P.C., and F.J. Huertas. 2009. The temporal evolution of pedogenic Fe–smectite to Fe–kaolin via interstratified kaolin–smectite in a moist tropical soil chronosequence. Geoderma 151:1–15.

Saggar, S., A. Parshotam, G.P. Sparling, C.W. Feltham, and P.B.S. Hart. 1996. 14C-labelled ryegrass turnover and residence times in soils varying in clay content and mineralogy. Soil Biol. Biochem. 28:1677–1686.

Saigusa, M., S. Shoji, and T. Kato. 1978. Origin and nature of halloysite in Ando soils from Towada tephra. Geoderma 20:115–129.

Sakharov, B.A., and V. Drits. 1973. Mixed-layer kaolinite-montmorillonite: A comparison of observed and calculated diffraction patterns. Clays Clay Miner. 21:15–17.

Sase, T., M. Hosono, T. Utsugawa, and K. Aoki. 1988. Opal phytolith analysis of present and buried volcanic ash soils at Te Ngae Road tephra section, Rotorua Basin, North Island, New Zealand. Quat. Res. (Japan) 27:153–163.

Sawhney, B.L. 1972. Selective sorption and fixation of cations by clay minerals: a review. Clays Clay Miner. 20:93–100.

Sawhney, B.L., and M.L. Jackson. 1958. Soil montmorillonite formulas. Soil Sci. Soc. Am. Proc. 22:115–118.

Scalenghe, R., E. Bonifacio, L. Celi, F.C. Ugolini, and E. Zanini. 2002. Pedogenesis in disturbed alpine soils (NW Italy). Geoderma 109:207–224.

Schaefer, C.E.G.R., R.J. Gilkes, and R.B.A. Fernandes. 2004. EDS/SEM study on microaggregates of Brazilian Latosols, in relation to P adsorption and clay fraction attributes. Geoderma 123:69–81.

Schaetzl, R.J., and S. Anderson. 2005. Soils—Genesis and geomorphology. Cambridge University Press, Cambridge, U.K.

Schott, J., and R.A. Berner. 1985. Dissolution mechanisms of pyroxenes and olivines during weathering, p. 35–53. In J.I. Drever (ed.) The chemistry of weathering. Reidel, New York.

Schwertmann, U. 2008. Iron oxides, p. 363–369. In W. Chesworth (ed.) Encyclopaedia of soil science. Springer, Dordrecht, the Netherlands.

Schwertmann, U., and R.M. Taylor. 1989. Iron oxides, p. 379–438. In J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Sedov, S., E. Solleiro-Rebolledo, P. Morales-Puente, A. Arias-Herrìe, E. Vallejo-Gòmez, and C. Jasso-Castañeda. 2003. Mineral and organic components of the buried paleosols of the Nevado de Toluca, central Mexico as indicators of paleoenvironments and soil evolution. Quat. Int. 106–107:169–184.

Senkayi, A.L., J.B. Dixon, L.R. Hossner, and B.E. Viani. 1983. Mineralogical transformations during weathering of lignite overburden in east Texas. Clays Clay Miner. 31:49–56.

Seyama, H., M. Soma, and A. Tanaka. 1996. Surface characterization of acid-leached olivines by X-ray photoelectron spectroscopy. Chem. Geol. 129:209–216.

Seybold, C.A., R.B. Grossman, and T.G. Reinsch. 2005. Predicting cation exchange capacity for soil survey using linear methods. Soil Sci. Soc. Am. Proc. 69:856–863.

Shepherd, T.G. 1984. A pedological study of the Hamilton Ash Group at Welches Road, Mangawara, north Waikato. Unpublished M.Sc. thesis. University of Waikato, Hamilton, New Zealand.

Shikazono, N., A. Takino, and H. Ohtani. 2005. An estimate of dissolution rate constant of volcanic glass in volcanic ash soil from the Mt. Fuji area, central Japan. Geochem. J. 39:185–196.

Shoji, S., M. Nanzyo, and R.A. Dahlgren. 1993a. Volcanic ash soils: Genesis, properties and utilization. Developments in soil science. Vol. 21. Elsevier, Amsterdam, the Netherlands.

Shoji, S., M. Nanzyo, Y. Shirato, and T. Ito. 1993b. Chemical kinetics of weathering in young Andisols from northeastern Japan using soil age normalized to 10°C. Soil Sci. 155:53–60.

Shoji, S., M. Nanzyo, and T. Takahashi. 2006. Factors of soil formation: Climate. As exemplified by volcanic ash soils, p. 131–149. In G. Certini and R. Scalenghe (eds.) Soils: Basic concepts and future challenges. Cambridge University Press, Cambridge, U.K.

Siffert, B. 1978. Genesis and synthesis of clays and clay minerals: Recent developments and future prospects, p. 337–347. In M.M. Mortland and V.C. Farmer (eds.) Developments in sedimentology. Vol. 27. Int. Clay Conf. 1978. Elsevier, Amsterdam, the Netherlands.

Sigfusson, B., S.R. Gislason, and G.I. Paton. 2008. Pedogenesis and weathering rates of a Histic Andosol in Iceland: Field and experimental soil solution study. Geoderma 144:572–592.

Simas, F.N.B., C.E.G.R. Schaefer, E.I.F. Filho, A.C. Chagas, and P.C. Brandão. 2005. Chemistry, mineralogy and micropedology of highland soils on crystalline rocks of Serra da Mantiqueira, southeastern Brazil. Geoderma 125:187–201.

Simas, F.N.B., C.E.G.R. Schaefer, V.F. Melo, M.B.B. Guerra, M. Saunders, and R.J. Gilkes. 2006. Clay-sized minerals in permafrost-affected soils (Cryosols) from King George Island, Antarctica. Clays Clay Miner. 54:721–736.

Simonsson, M., S. Hillier, and I. Öborn. 2009. Changes in clay minerals and potassium fixation capacity as a result of release and fixation of potassium in long-term field experiments. Geoderma 151:109–120.

Singer, A. 1989. Palygorskite and sepiolite group minerals, p. 829–872. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Singer, A. 2002. Palygorskite and sepiolite, p. 555–584. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Singer, A., and A. Banin. 1990. Characteristics and mode of formation of palagonite: A review, p. 173–181. *In* V.C. Farmer and Y. Tardy (eds.) Proc. 9th Int. Clay Conf., 28 August–2 September 1989, Strasbourg, France.

Singer, A., and K. Norrish. 1974. Pedogenic palygorskite occurrences in Australia. Am. Mineral. 59:508–517.

Singer, A., M. Zarei, F.M. Lange, and K. Stahr. 2004. Halloysite characteristics and formation in the northern Golan Heights. Geoderma 123:279–295.

Singh, B., and R.J. Gilkes. 1991. A potassium-rich beidellite from a lateritic pallid zone in Western Australia. Clay Miner. 26:233–244.

Singh, B., and R.J. Gilkes. 1992a. Properties of soil kaolinites from south-western Australia. J. Soil Sci. 43:645–667.

Singh, B., and R.J. Gilkes. 1992b. The electron-optical investigation of the alteration of kaolinite to halloysite. Clays Clay Miner. 40:212–229.

Singh, B., and R.J. Gilkes. 1993. The recognition of amorphous silica in indurated soil profiles. Clay Miner. 28:461–474.

Singleton, P.L., M. Mcleod, and H.J. Percival. 1989. Allophane and halloysite content and soil solution silicon in soils from rhyolitic volcanic material, New Zealand. Aust. J. Soil Res. 27:67–77.

Skiba, M. 2007. Clay mineral formation during podzolization in an alpine environment of the Tatra Mountains, Poland. Clays Clay Miner. 55:618–637.

Smeck, N.E., J.M. Bigham, W.F. Guertal, and G.F. Hall. 2002. Spatial distribution of lepidocrocite in a soil hydrosequence. Clay Miner. 37:687–697.

Smith, W.W. 1962. Weathering of some Scottish basic igneous rocks with reference to soil formation. J. Soil Sci. 13:202–215.

Smith, K.L., A.R. Milnes, and R.A. Eggleton. 1987. Weathering of basalt: Formation of iddingsite. Clays Clay Miner. 35:418–428.

Smits, M.M., E. Hoffland, A.G. Jongmans, and N. Van Breeman. 2005. Contribution of mineral tunnelling to total feldspar weathering. Geoderma 125:59–69.

Soma, M., G.J. Churchman, and B.K.G. Theng. 1992. X-ray photoelectron spectroscopic analysis of halloysites with different composition and particle morphology. Clay Miner. 27:413–421.

Sousa, E.C., and H. Eswaran. 1975. Alteration of micas in the saprolite of a profile from Angola. A morphological study. Pedologie 25:71–79.

Sparling, G.P., D. Wheeler, E.-T. Vesely, and L.A. Schipper. 2006. What is soil organic matter worth? J. Environ. Qual. 35:548–557.

Spielvogel, S., J. Preitzel, and I. Kögel-Knabner. 2008. Soil organic matter stabilization in acidic forest soils is preferential and soil type-specific. Eur. J. Soil Sci. 59:674–692.

Sposito, G. 1989. The chemistry of soils. Oxford University Press, New York

Sridhar, K., and M.L. Jackson. 1974. Layer charge decrease by tetrahedral cation removal and silicon incorporation during natural weathering of phlogopite to saponite. Soil Sci. Soc. Am. Proc. 38:847–851.

Sridhar, K., M.L. Jackson, and R.N. Clayton. 1975. Quartz oxygen isotopic stability in relation to isolation from sediments and diversity of source. Soil Sci. Soc. Am. Proc. 42:1209–1213.

Starr, M., and A.-J. Lindroos. 2006. Changes in the rate of release of Ca and Mg and normative mineralogy due to weathering along a 5300-year chronosequence of boreal forest soils. Geoderma 133:269–280.

Stephen, I., and D.M.C. MacEwan. 1951. Some chlorite minerals of unusual type. Clay Miner. Bull. 1:157–161.

Stevens, K.F., and C.G. Vucetich. 1985. Weathering of Upper Quaternary tephras in New Zealand. 2. Clay minerals and their climatic interpretation. Chem. Geol. 53:237–247.

Stewart, R.B., V.E. Neall, and J.K. Syers. 1984. Occurrence and source of quartz in six basaltic soils from Northland, New Zealand. Aust. J. Soil Res. 22:365–377.

Stoch, L., and W. Sikora. 1976. Transformation of micas in the process of kaolinitization of granites and gneisses. Clays Clay Miner. 24:156–162.

Stroncik, N.A., and H.-U. Schmincke. 2001. Evolution of palagonite: Crystallization, chemical changes, and element budget. Geochem. Geophys. Geosyst. 2:1017. doi:10.1029/2000GC000102.

Stroncik, N.A., and H.-U. Schmincke. 2002. Palagonite—A review. Int. J. Earth Sci. 91:680–697.

Stucki, J.W., and C.B. Roth. 1977. Oxidation reduction mechanism for structural iron in nontronite. Soil Sci. Soc. Am. J. 41:808–814.

Sudo, T., and H. Yatsumoto. 1977. The formation of halloysite tubes from spherulitic halloysite. Clays Clay Miner. 25:155–159.

Surapaneni, A., A.S. Palmer, R.W. Tillman, J.H. Kirkman, and P.E.H. Gregg. 2002. The mineralogy and potassium supplying power of some loessial and related soils of New Zealand. Geoderma 110:191–204.

Swindale, L.D., and G. Uehara. 1966. Ionic relationships in the pedogenesis of Hawaiian soils. Soil Sci. Soc. Am. Proc. 30:726–730.

Syers, J.K., V.E. Berkheiser, M.L. Jackson, R.N. Clayton, and R.W. Rex. 1969. Eolian sediment influence on pedogenesis during the Quaternary. Soil Sci. 107:421–427.

Takahashi, T., R.A. Dahlgren, B.K.G. Theng, J.S. Whitton, and M. Soma. 2001. Potassium-selective, halloysite-rich soils formed in volcanic materials from northern California. Soil Sci. Soc. Am. J. 65:516–526.

Takahashi, T., R. Dahlgren, and P. van Susteren. 1993. Clay mineralogy and chemistry of soils formed in volcanic materials in the xeric moisture regime of northern California. Geoderma 59:131–150.

Takesako, H., D.J. Lowe, G.J. Churchman, and D.J. Chittleborough. 2010. Holocene volcanic soils in the Mt Gambier region, South Australia, p. 47–50. Proceedings published on DVD and at http://www.iuss.org, IUSS 19th World Soil Cong., Brisbane (August 1–6, 2010), Symposium 1.3.1 Pedogenesis: Ratio and range of influence.

Tamm, O. 1922. Eine methode Zur Bestimmung de anorganischen Komponente des Gelcomplexes im Boden. Meddelanden fran Statens skogsforsokanstalt Stockholm 19:387–404.

Taubaso, C., M. Dos Santos Afonso, and R.M. Torres Sánchez. 2004. Modelling soil surface charge density using mineral composition. Geoderma 121:123–133.

Taylor, R.M. 1988. Proposed mechanism for the formation of soluble Si–Al and Fe(III)–Al hydroxyl complexes in soils. Geoderma 42:65–77.

Tazaki, K. 1979. Micromorphology of halloysite produced by weathering of plagioclase in volcanic ash. Dev. Sedimentol. 27:415–422.

Tazaki, K. 1997. Biomineralization of layer silicates and hydrated Fe/Mn oxides in microbial mats: An electron microscopical study. Clays Clay Miner. 45:203–212.

Tazaki, K. 2005. Microbial formation of a halloysite-like mineral. Clays Clay Miner. 53:224–233.

Theng, B.K.G. 1995. On measuring the specific surface area of clays and soils by adsorption of para-nitrophenol: Use and limitations, p. 304–310. *In* G.J. Churchman, R.W. Fitzpatrick, and R.A. Eggleton (eds.) Clays: Controlling the environment. Proc. 10th Int. Clay Conf. Adelaide, Australia. CSIRO Publishing, Melbourne, Australia.

Theng, B.K.G., M. Russell, G.J. Churchman, and R.L. Parfitt. 1982. Surface properties of allophane, halloysite, and imogolite. Clays Clay Miner. 30:143–149.

Theng, B.K.G., and G. Yuan. 2008. Nanoparticles in the soil environment. Elements 4:395–399.

Thiry, M., and R. Simon-Coinçon. 1996. Tertiary paleoweatherings and silcretes in the southern Paris Basin. Catena 26:1–26.

Thompson, A., O.A. Chadwick, D.G. Rancourt, and J. Chorover. 2006. Iron-oxide crystallinity increases during soil redox oscillations. Geochim. Cosmochim. Acta 70:1710–1727.

Thompson, M.L., and L. Ukrainczyk. 2002. Micas, p. 431–466. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Thorseth, I.H., H. Furnes, and O. Tumyr. 1995. Textural and chemical effects of bacterial-activity on basaltic glass—An experimental approach. Chem. Geol. 119:139–160.

Tonkin, P.J. 1970. Contorted stratification with clay lobes in volcanic ash beds, Raglan-Hamilton region, New Zealand. Earth Sci. J. 4:129–140.

Torres, S.R.M., M. Okamura, and R.C. Mercader. 2001. Charge properties of red Argentine soils as an indicator of iron oxide/clay associations. Aust. J. Soil Res. 39:423–434.

Trakoonyingcharoen, P., I. Kheoruenromne, A. Suddhiprakarn, and R.J. Gilkes. 2006. Properties of iron oxides in red Ultisols as affected by rainfall and soil parent material. Aust. J. Soil Res. 44:63–70.

Trolard, F., J.-M. Génin, M. Abdelmoula, G. Bourrié, B. Humbert, and A. Herbillon. 1997. Identification of a green rust mineral in a reductomorphic soil by Mössbauer and Raman spectroscopies. Geochim. Cosmochim. Acta 61:1107–1111.

Turpault, M.-P., D. Righi, and C. Utérano. 2008. Clay minerals: Precise markers of the spatial and temporal variability of the biogeochemical soil environment. Geoderma 147:108–115.

Tye, A.M., S.J. Kemp, and P.R. Poulton. 2009. Responses of soil clay mineralogy in the Rothamsted classical experiments to management practice and changing land use. Geoderma 153:136–146.

Ueshima, M., and Tazaki, K. 2001. Possible role of microbial polysaccharides in nontronite formation. Clays Clay Miner. 49:292–299.

Ugolini, F.C., R. Dahlgren, J. LaManna, W. Nuhn, and J. Zachara. 1991. Mineralogy and weathering processes in recent and Holocene tephra deposits of the Pacific Northwest, USA. Geoderma 51:277–299.

Ugolini, F.C., and R.S. Sletten. 1991. The role of proton donors in pedogenesis as revealed by soil solution studies. Soil Sci. 151:59–75.

Urrutia, M.M., and T.J. Beveridge. 1994. Formation of fine-grained silicate minerals and metal precipitates by a bacterial surface (*Bacillus subtilis*). Chem. Geol. 116:261–280.

Vacca, A., P. Adamo, M. Pigna, and P. Violante. 2003. Genesis of tephra-derived soils from the Roccamonfina volcano, south central Italy. Soil Sci. Soc. Am. J. 67:198–207.

van Breeman, N. 1992. Soil: Biotic construction in a Gaian sense, p. 189–207. *In* A. Teller, P. Muthy, and J.N.R. Jeffers (eds.) Responses of forest ecosystems to environmental change. Elsevier, Amsterdam, the Netherlands.

van Breemen, N., R.D. Findlay, U.S. Lundström, A.G. Jongmans, R. Giesler, and M. Olsson. 2000a. Mychorrhizal weathering: A true case of mineral nutrition? Biogeochemistry 49:53–67.

van Breemen, N., and A.C. Finzi. 1998. Plant-soil interactions: Ecological aspects and evolutionary implications. Biogeochemistry 42:1–19.

van Breemen, N., U.S. Lundström, and A.G. Jongmans. 2000b. Do plants drive podzolisation via rock-eating mycorrhizal fungi? Geoderma 94:163–171.

van Raij, B., and M. Peech. 1972. Electrochemical properties of some Oxisols and Alfisols of the tropics. Proc. Soil Sci. Soc. Am. 36:587–593.

Van Ranst, E., and F. De Coninck. 2002. Evaluation of ferrolysis in soil formation. Eur. J. Soil Sci. 53:513–519.

van Wesemael, B., J.M. Verstraten, and J. Sevink. 1995. Pedogenesis by clay dissolution on acid, low-grade metamorphic rocks under Mediterranean forests in southern Tuscany (Italy). Catena 24:105–125.

Varajão, A.F.D.C., R.J. Gilkes, and R.D. Hart. 2001. The relationships between kaolinite crystal properties and the origin of materials for a Brazilian kaolin deposit. Clays Clay Miner. 49:44–59.

Velbel, M.A. 1985. Geochemical mass balances and weathering rates in forested watersheds of the southern Blue Ridge. Am. J. Sci. 285:904–930.

Velbel, M.A., and W.W. Barker. 2008. Pyroxene weathering to smectite: Conventional and cryo-field emission scanning microscopy, Koua Bocca ultramafic complex, Ivory Coast. Clays Clay Miner. 56:112–127.

Velde, B. (ed.). 1995. Origin and mineralogy of clays. Springer-Verlag, New York.

Velde, B., and P. Barré. 2010. Soils, plants and clay minerals: Mineral and biologic interactions. Springer-Verlag, Berlin, Germany.

Velde, B., B. Goffé, and A. Hoellard. 2003. Evolution of clay minerals in a chronosequence of poldered sediments under the influence of a natural pasture development. Clays Clay Miner. 51:205–217.

Velde, B., and A. Meunier. 2008. The origin of clay minerals in soils and weathered rocks. Springer-Verlag, Berlin, Germany.

Velde, B., and T. Peck. 2002. Clay mineral changes in the Morrow experimental plots, University of Illinois. Clays Clay Miner. 50:364–370.

Vepraskas, M.J. 1992. Redoximorphic features for identifying aquic conditions. North Carolina Agricultural Research Service Technical Bulletin No. 301. North Carolina State University, Raleigh, NC.

Vepraskas, M.J., X. He, D.L. Lindbo, and R.W. Skaggs. 2004. Calibrating hydric soil field indicators to long-term wetland hydrology. Soil Sci. Soc. Am. J. 68:146–1469.

Verrecchia, E.P., and M.-N. Le Coustumer. 1996. Occurrence and genesis of palygorskite and associated clay minerals in a Pleistocene calcrete complex, Sde Boqer, Negev Desert, Israel. Clay Miner. 31:183–202.

Viana, J.H.M., P.R.C. Couceiro, M.C. Pereira, J.D. Fabris, E.I. Fernandes Filho, C.E.G.R. Schaefer, H.R. Rechenberg, W.A.P. Abrahão, and E.C. Mantovani. 2006. Occurrence of magnetite in the sand fraction of an oxisol in the Brazilian savanna ecosystem, developed from a magnetite-free lithology. Aust. J. Soil Res. 44:71–83.

Vicente, M.A., M. Razzaghe, and M. Robert. 1977. Formation of aluminium hydroxy vermiculite (intergrade) and smectite from mica under acidic conditions. Clay Miner. 12:101–112.

Vingiani, S., D. Righi, S. Petit, and F. Terribile. 2004. Mixed-layer kaolinite–smectite minerals in a red-black soil sequence from basalt in Sardinia (Italy). Clays Clay Miner. 52:473–483.

Wada, K. 1987. Minerals formed and mineral formation from volcanic ash by weathering. Chem. Geol. 60:17–28.

Wada, K. 1989. Allophane and imogolite, p. 1051–1087. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA Book Series No. 1. SSSA, Madison, WI.

Wada, K., and Y. Kakuto. 1983. Intergradient vermiculite-kaolin mineral in a Korean Ultisol. Clays Clay Miner. 31:183–190.

Wada, K., and Y. Kakuto. 1989. "Chloritized" vermiculite in a Korean Ultisol studied by ultramicrotomy and transmission electron microscopy. Clays Clay Miner. 37:263–268.

Wada, K., Y. Kakuto, and F.N. Muchena. 1987. Clay minerals and humus complexes in five Kenyan soils derived from volcanic ash. Geoderma 39:307–321.

Wada, K., and N. Yoshinaga. 1969. The structure of imogolite. Am. Mineral. 54:50–71.

Wallace, R.C. 1991. New Zealand theses in earth sciences. The mineralogy of the Tokomaru silt loam and the occurrence of cristobalite and tridymite in selected North Island soils. N.Z. J. Geol. Geophys. 34:113.

Wan, J., T. Tyliszczak, and T.K. Tokunga. 2007. Organic carbon distribution, speciation, and elemental correlations within soil microaggregates: Applications of STXM and NEXAFS spectroscopy. Geochim. Cosmochim. Acta 71:5439–5449.

Wang, C., J.A. McKeague, and H. Kodama. 1986. Pedogenic imogolite and soil environments: Case study of spodosols in Quebec, Canada. Soil Sci. Soc. Am. J. 50:711–718.

Ward, W.T. 1967. Volcanic ash beds of the lower Waikato Basin, North Island, New Zealand. N.Z. J. Geol. Geophys. 10:1109–1135.

Watanabe, T., Y. Sawada, J.D. Russell, W.J. McHardy, and M.J. Wilson. 1992. The conversion of montmorillonite to interstratified halloysite–smectite by weathering in the Omi acid clay deposit. Clay Miner. 27:159–173.

Waychunas, G.A., and H. Zhang. 2008. Structure, chemistry, and properties of mineral nanoparticles. Elements 4:381–387.

Weaver, C.E. 1989. Clays, muds and shales. Developments in sedimentology. Vol. 44. Elsevier, Amsterdam, the Netherlands.

Welch, S.A., and J.F. Banfield. 2002. Modification of olivine surface morphology and reactivity by microbial activity during chemical weathering. Geochim. Cosmochim. Acta 66:213–221.

Wentworth, S.A. 1970. Illite. Clay Sci. 3:140–155.

Wentworth, S.J., E.K. Gibson, M.A. Velbel, and D.S. McKay. 2005. Antarctic dry valleys and indigenous weathering in Mars meteorites: Implications for water and life on Mars. Icarus 174:383–395.

White, A.F., A.E. Blum, M.S. Schulz, T.D. Bullen, J.W. Harden, and M.L. Peterson. 1996. Chemical weathering rates of a soil chronosequence on granitic alluvium: 1. Quantification of mineralogical and surface area changes and calculation of primary silicate reaction rates. Geochim. Cosmochim. Acta 60:2553–2550.

White, A.F., and S.L. Brantley. 2003. The effect of time on the weathering of silicate minerals: Why do weathering rates differ in the laboratory and field? Chem. Geol. 202:479–506.

White, G.N., and J.B. Dixon. 2002. Kaolin–serpentine minerals, p. 389–414. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA Book Series No. 7. SSSA, Madison, WI.

White, A.F., M.S. Schulz, D.V. Vivit, A.E. Blum, D.A. Stonestrom, and J.W. Harden. 2005. Chemical weathering rates of a soil chronosequence on granitic alluvium: III. Hydrochemical evolution and contemporary solute fluxes and rates. Geochim. Cosmochim. Acta 69:1975–1996.

Wildman, W.E., M.L. Jackson, and L.D. Whittig. 1968. Serpentinite rock dissolution as a function of carbon dioxide pressure in aqueous solution. Am. Mineral. 53:1252–1263.

Wilson, M.J. 1970. A study of weathering in a soil derived from a biotite–hornblende rock. I. Weathering of biotite. Clay Miner. 8:291–303.

Wilson, M.J. 1987. Soil smectites and related interstratified minerals: Recent developments, p. 167–173. *In* L.G. Schultz, H. van Olphen, and F.A. Mumpton (eds.) Proc. Int. Clay Conf. Denver, 1995. The Clay Minerals Society, Bloomington, IN.

Wilson, M.J. 1999. The origin and formation of clay minerals in soils: Past, present and future perspectives. Clay Miner. 34:7–25.

Wilson, M.J. 2004. Weathering of the primary rock-forming minerals: Processes, products and rates. Clay Miner. 39:233–266.

Wilson, M.J. 2006. Factors of soil formation: Parent material. As exemplified by a comparison of granitic and basaltic soils, p. 113–129. *In* G. Certini and R. Scalenghe (eds.) Soils: Basic concepts and future challenges. Cambridge University Press, Cambridge, U.K.

Wilson, M.J., D.C. Bain, and D.M.L. Duthie. 1984. The soil clays of Great Britain. II. Scotland. Clay Miner. 19:709–735.

Wilson, M.J., G. Certini, C.D. Campbell, I.C. Anderson, and S. Hillier. 2008. Does the preferential microbial colonisation of ferromagnesian minerals affect mineral weathering in soil? Naturwissenschaften 95:851–858.

Wilson, M.J., and W.J. McHardy. 1980. Experimental etching of a microcline perthite and implications regarding natural weathering. J. Microsc. 120:291–302.

Wolff-Boenisch, D., S.R. Gislason, E.H. Oelkers, and C.V. Putnis. 2004. The dissolution rates of natural glasses as a function of their composition at pH 4 and 10.6, and temperatures from 25 to 74°C. Geochim. Cosmochim. Acta 68:4843–4858.

Worden, R.H., F.D.L. Walker, I. Parsons, and W.L. Brown. 1990. Development of microporosity, diffusion channels and deuteric coarsening in perthitic alkali feldspars. Contrib. Mineral Petrol. 104:507–515.

Wright, J.D., and N.A.J.M. Sommerdijk. 2001. Soil-gel materials, chemistry and applications. Advanced chemistry texts. CRC Press, Boca Raton, FL.

Yagasaki, Y., J. Mulder, and M. Okazaki. 2006. The role of soil organic matter and short-range ordered aluminosilicates in controlling the activity of aluminum in soil solutions of volcanic ash soils. Geoderma 137:40–57.

Yaron, B., I. Dror, and B. Berkowitz. 2010. Contaminant geochemistry—A new perspective. Naturwissenschaften 97:1–17.

Yoshinaga, N., and S. Aomine. 1962. Imogolite in some Ando soils. Soil Sci. Plant Nutr. 8:22–29.

Young, I.M., and J.W. Crawford. 2004. Interactions and self-organization in the soil-microbe complex. Science 304:1634–1637.

Zanelli, R., M. Egli, A. Mirabella, M. Abdelmoula, M. Plötze, and M. Nötzli. 2006. "Black" soils in the southern Alps: Clay mineral formation and transformation, X-ray amorphous Al phases and Fe forms. Clays Clay Miner. 54:703–720.

Zehetner, F., W.P. Miller, and L.T. West. 2003. Pedogenesis of volcanic ash soils in Andean Ecuador. Soil Sci. Soc. Am. J. 67:1797–1809.

Zelazny, L.W., and F.G. Calhoun. 1977. Palygorskite (attapulgite), sepiolite, talc, pyrophyllite and zeolites, p. 435–470. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Zelazny, L.W., P.J. Thomas, and C.L. Lawrence. 2002. Pyrophyllite-talc minerals, p. 415–430. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA Book Series No. 7. SSSA, Madison, WI.

Zelazny, L.W., and G.N. White. 1989. The pyrophyllite-talc group, p. 527–550. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Ziegler, K., J.C.C. Hsieh, O.A. Chadwick, E.F. Kelly, D.M. Hendricks, and S.M. Savin. 2003. Halloysite as a kinetically controlled end product of arid-zone basalt weathering. Chem. Geol. 202:461–478.

Zysset, M., P. Blaser, J. Luster, and A.U. Gehring, 1999. Aluminum solubility control in different horizons of a podzol. Soil Sci. Soc. Am. J. 63:1106–1115.

21

Phyllosilicates

Hideomi Kodama
Agriculture and Agri-Food Canada

21.1 Introduction

Phyllosilicates, in other words, layer silicates, occupy one of six subclasses under the silicates class (Table 21.1). Phyllo- is after the Greek word describing the state of leaves, like a bed or heap of leaves. Thus in science, often the word is used as a prefix to express thin layer or paper sheet like in a book. This reflects general crystal habits and morphological features of the phyllosilicates. The silicates are abundant on the earth and of greater importance than any other minerals, which is easily understood from the fact that oxygen and silicon constitute nearly 75% of the elements in the earth's crust (Figure 21.1). The fundamental unit in all silicate structures is the silicon–oxygen tetrahedron (Figure 21.2). As Bragg (1950) stated, the bonds between silicon and oxygen are so strong that the four oxygen atoms are always found at the corners of a tetrahedron of nearly constant dimensions and regular shape whatever the rest of the structure may be like. The Si–O distance is about 0.16 nm, and the O–O distance 0.26 nm. Besides the crustal abundance of these two elements, the ionic radius of Si^{4+}, 0.042 nm, is conveniently accommodated in surroundings with four oxygens to form the fundamental structural unit of all silicates, the tetrahedron having the chemical formula $(SiO_4)^{4-}$. This unit can occur independently with compensating cations like Mg^{2+} and Fe^{2+}. And also these SiO_4 tetrahedra may be linked together by having one or more oxygen atoms shared in common by neighboring tetrahedra to form indefinitely extended and more complicated structural units to cover the existing silicate minerals. This mechanism of linking together is known as polymerization, and the classified structural units with their compositions are given in Table 21.2.

21.2 General Structural Features

The crystal structure of phyllosilicates has been determined largely by x-ray diffraction (XRD) methods that were well established within two to three decades of Roentgen's discovery of x-rays in 1895. The essential features of the mineral, typified by mica-like minerals, are the continuous 2D tetrahedral sheets of composition, $Si_2O_5^{2-}$, in which SiO_4 tetrahedra (Figure 21.3) are linked by sharing three corners of each tetrahedron to form a hexagonal mesh pattern (Figure 21.4a). Frequently Si atoms of the tetrahedra are partially substituted by Al and, to a lesser extent, by Fe^{3+}. The apical oxygen at the fourth corner of the tetrahedra, which is usually directed normal to the sheet, forms part of an adjacent octahedral sheet in which octahedra are linked by sharing edges (Figure 21.5). The junction plane between tetrahedral and octahedral sheets consists of the shared apical oxygens of the tetrahedra and unshared hydroxyls that lie at the center of each hexagonal ring of tetrahedra and at the same level as the shared apical oxygen atoms (Figure 21.6). Further details of octahedral configurations will be seen in Section 21.2.2, where the explanation of polytypes is given using the phlogopite structure as the example. Common cations that coordinate the octahedral sheets are Al, Mg, Fe^{3+}, and Fe^{2+}; but occasionally, Li, V, Cr, Mn, Ni, Cu, and Zn substitute in considerable amounts

TABLE 21.1   Six Subclasses under the Silicate Mineral Class

Subclass (Commonly Used)	Subclass (in Other Name)
Nesosilicates	Orthosilicates
Sorosilicates	Disilicates
Cyclosilicates	Ring silicates
Inosilicates	Chain silicates—Single chain and double chain
Phyllosilicates	Layer silicates
Tectosilicates	Framework silicates
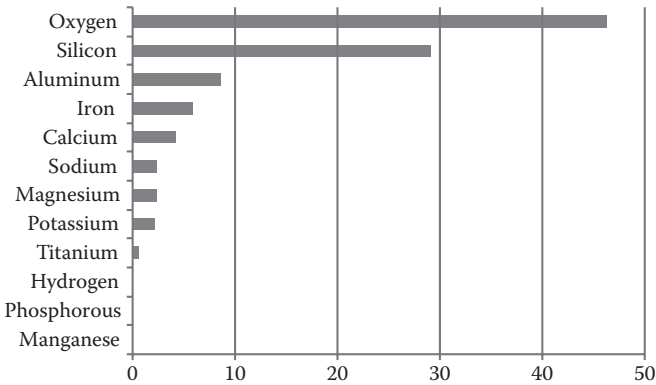
FIGURE 21.1   The 12 most abundant elements in the earth's crust. (Modified graph from Vanders, I., and P. Kerr. 1967. Mineral recognition. John Wiley & Sons, Inc., New York.)
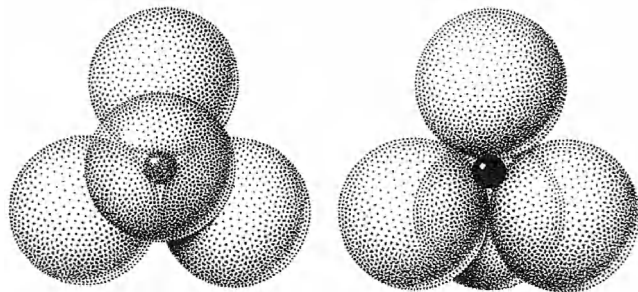
FIGURE 21.2   Close-packing representation of the SiO_4 tetrahedron. (From Klein, C., and C.S. Hurlbut, Jr., 1985. Manual of mineralogy. John Wiley & Sons, Inc., New York.)

TABLE 21.2   Unit Compositions of Six Subclasses, Mineral Group, and Species Examples

Subclass	Unit Composition	Mineral Group Example	Mineral Species Example
Nesosilicates	$(SiO_4)^{4-}$	Olivin	Forsterite
Sorosilicates	$(Si_2O_7)^{6-}$	Epidote	Zoisite
Cyclosilicates	$(Si_6O_{18})^{12-}$	Tourmaline	Elbaite
Inosilicates			
Single chain	$(SiO_3)^{2-}$	Pyroxene	Diopside
Double chain	$(Si_4O_{11})^{6-}$	Amphibole	Actinolite
Phyllosilicates	$(Si_2O_5)^{2-}$	Mica	Muscovite
Tectosilicates	$(SiO_2)^0$	Feldspar	Albite
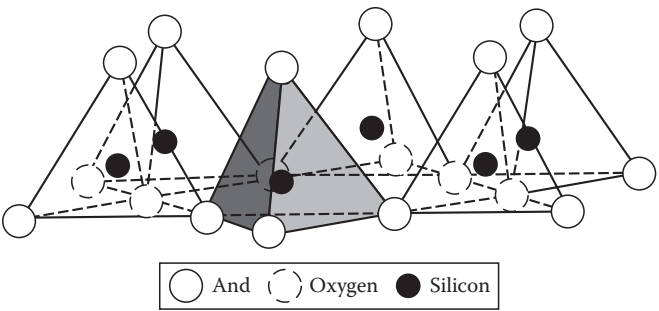
FIGURE 21.3 SiO_4 tetrahedra linking to form a hexagonal network. Single tetrahedron is shaded. (From Grim, R.E. 1968. Clay mineralogy. McGraw-Hill Book Co. Inc., New York. With permission.)
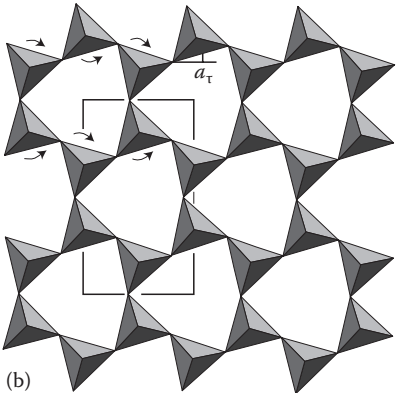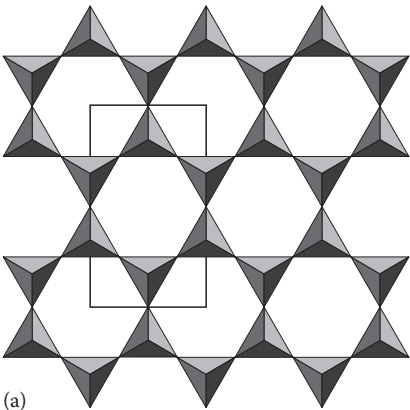
FIGURE 21.4   (a) Ideal hexagonal tetrahedral sheet. (b) Contracted sheet of ditrigonal symmetry by rotation of the tetrahedra due to a dimension misfit between tetrahedral and octahedral sheets. (With kind permission from Liebau, F. 1985. Structural chemistry of silicates: Structure, bonding, and classification. Springer-Verlag, New York.)

into the sheets. If divalent cations (M^{2+}) are in the octahedral sheets, the composition is $M_3^{2+}(OH)_2O_4$ and all the octahedra are occupied; if trivalent cations (M^{3+}), the composition is $M_2^{3+}(OH)_2O_4$ and two-thirds of the octahedra are occupied, with the absence of the third octahedron. The former type of octahedral sheet is called trioctahedral (Figure 21.7a) and the latter dioctahedral (Figure 21.7b). This distinction may occasionally be used as a subgrouping for certain mineral groups
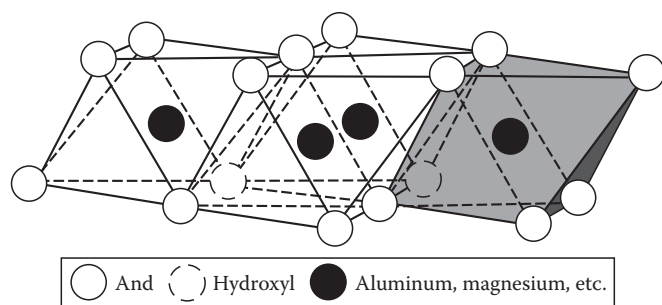
FIGURE 21.5 Single octahedron (shaded) and the sheet structure of octahedral units. (From Grim, R.E. 1968. Clay mineralogy. McGraw-Hill Book Co. Inc., New York. With permission.)
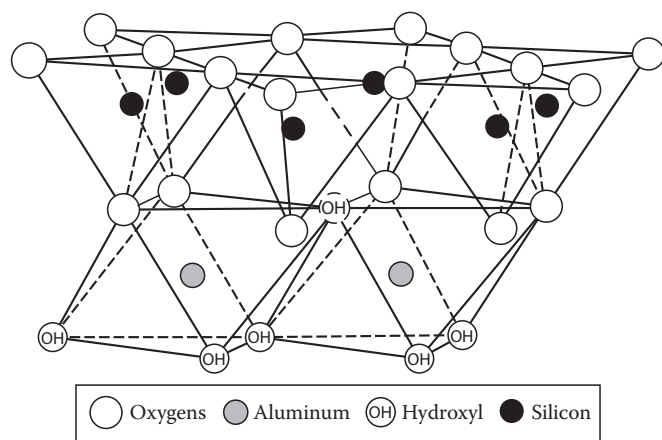
FIGURE 21.6 Structure of 1:1 layer silicate (kaolinite) illustrating the junction between tetrahedral and octahedral sheets. (From Grim, R.E. 1968. Clay mineralogy. McGraw-Hill Book Co. Inc., New York. With permission.)

FIGURE 21.7 Difference in the octahedral sheets of clay minerals: (a) trioctahedral, with dots of M^{2+}; (b) dioctahedral, with dots of M^{3+}. (With kind permission of Liebau, F. 1985. Structural chemistry of silicates: Structure, bonding, and classification. Springer-Verlag, New York.)

(e.g., mica). If all the anion groups are OH in the compositions of octahedral sheets, the resulting sheets may be expressed by $M^{2+}(OH)_2$ and $M^{3+}(OH)_3$, respectively. Such sheets called hydroxide sheets occur singly, alternating with silicate layers in some phyllosilicates, for example, in chlorite minerals. Brucite, $Mg(OH)_2$, and gibbsite, $Al(OH)_3$, are typical examples of minerals having similar structures.

There are two major types for structural "backbones" of phyllosilicates called silicate layers. The unit silicate layer formed by aligning one octahedral sheet to one tetrahedral sheet is referred to as a 1:1 silicate layer (Figure 21.8a), and the exposed surface of the octahedral sheet consists of hydroxyls. In another type, the unit silicate layer comprises one octahedral sheet sandwiched by two tetrahedral sheets, which are oriented in opposite directions and is termed a 2:1 silicate layer (Figure 21.8b). If hydroxide sheets in the chlorite structure are treated as a proxy for the octahedral sheet of the silicate layer and denoted as T for the tetrahedral sheet and O for the octahedral sheet, a majority of the phyllosilicates may be grouped by T–O type, T–O–T type, and T–O–T–O type, which correspond to 1:1, 2:1, and 2:2 (2:1:1 in some occasions) silicate layers, respectively.

Real structures of phyllosilicates, however, contain substantial crystal strains and distortions, which produce irregularities such as deformed octahedra and tetrahedra rather than polyhedra
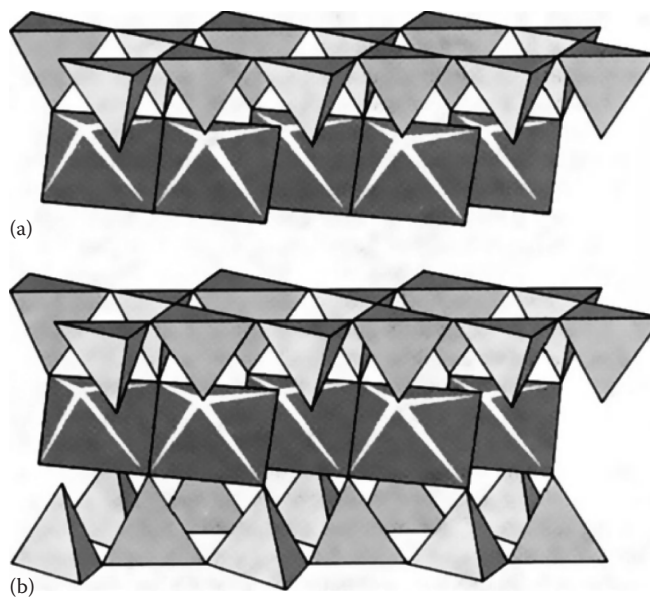
FIGURE 21.8 Two basic silicate layer types of phyllosilicates: (a) 1:1 or T–O type silicate layer structure, (b) 2:1 or T–O–T type silicate layer structure. (With kind permission of Springer Science + Business media: Liebau, F. 1985. Structural chemistry of silicates: Structure, bonding, and classification. Springer-Verlag, New York.)

with equilateral hexagonal faces, ditrigonal symmetry modified from the ideal hexagonal surface symmetry, and puckered surface instead of the flat plane made up by the basal oxygen atoms of the tetrahedral sheet. One of the major causes of such distortions is dimensional "misfits" between the tetrahedral and octahedral sheets. If the tetrahedral sheet contains only silicon in the cationic site and has an ideal hexagonal symmetry, the longer unit dimension within the basal dimension is 0.915 nm, which lies between the corresponding dimensions 0.86 nm of gibbsite and 0.94 nm of brucite. To fit the tetrahedral sheet into the dimension of the octahedral sheet, alternate SiO_4 tetrahedra rotate (up to a theoretical maximum of 30°) in opposite directions to distort the ideal hexagonal array into a doubly triangular (ditrigonal) array (Figure 21.4b). By this distortion mechanism, tetrahedral and octahedral sheets of a wide range of composition due to ionic substitutions can link together and maintain silicate layers. Among ionic substitutions, those between ions of distinctly different sizes most significantly affect geometric configurations of silicate layers.

21.2.1 Ionic Sizes and Coordination Number

The arrangement of anions around a cation (as normally anions are larger than cations), defined as the coordination number, would be expected to be the most symmetrical in the three dimensions, that is, 3, 4, 6, 8, and 12 ions would be arranged at the apices of an equilateral triangle, regular tetrahedron, octahedron, cube, and closest packing in order to minimize its overall potential energy. Pauling (1929) first elucidated that most minerals follow the principles that are now commonly known as Pauling's first rule to distinguish from his second rule regarding structure stabilization by the sum of the electrostatic strengths between an anion and adjacent cations. Thus, as far as geometry is concerned, the ionic radius ratio, R_{cation}/R_{anion}, is the key factor for determining a polyhedron suitable for an arrangement of anions around a cation for a specific ionic combination. Assuming that ions act as rigid spheres of fixed radii, the stable arrangements of cations and

TABLE 21.3 Relationship between Radius Ratio ($R_{cation}/R_{O^{2-}}$) and Coordination Number

Radius Ratio (R_{cation}/R_{anion})	Arrangement of Anions around a Cation	Coordination Number of Cation	Crystal Models Shown in Figure 21.9
0.15–0.22	Corners of an equilateral triangle	3	a
0.22–0.41	Corners of a tetrahedron	4	b
0.41–0.73	Corners of an octahedron	6	c
0.73–1	Corners of a cube	8	d
1	Closest packing	12	e

anions for particular radius ratios can be calculated from purely geometric considerations (Table 21.3) and the corresponding 3D coordination models are shown in Figure 21.9. Anions in phyllosilicate structures are mostly oxygen ions, O^{2-}, with radius of 0.132 nm. Although to a lesser extent the structures may contain F^- and $(OH)^-$, radii of these anions are similar to that of oxygen. Therefore, for phyllosilicates, it is practical to compare the relationship of coordination number with the size of cation in terms of $R_{cation}/R_{O^{2-}}$ (Table 21.4). The table allows making prediction for the coordination number of a cation in phyllosilicates and for a possible ionic substitution at a specific structural site.

21.2.2 Polytypes

Because phyllosilicates maintain hexagonal or near-hexagonal symmetry, the structures allow various ways to stack up atomic planes, sheets, and layers, which may be explained by crystallographic operations such as translation or shifting and rotation, thereby distinguishing them from polymorphs (e.g., diamond–graphite; calcite–aragonite). The former involves 1D variations but the latter generally 3D ones. With a fixed chemical composition, structural varieties that are resulted from different stacking sequences are
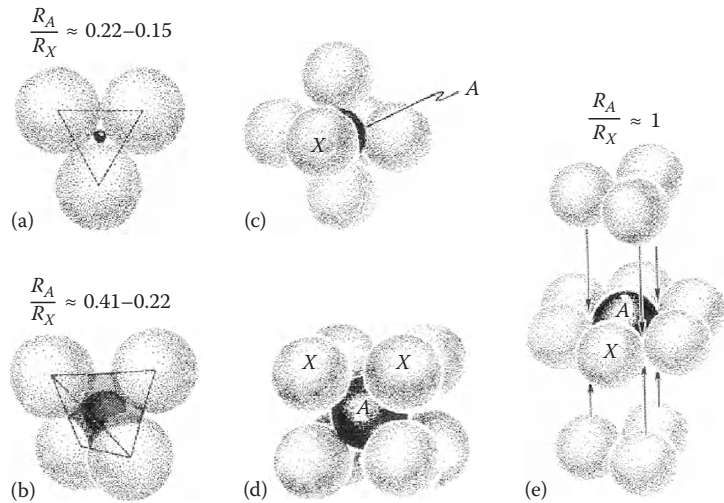
FIGURE 21.9 Coordination number (ligancy) See Table 21.3 for further details of each crystal model. (Modified after Bloss, F.D. 1971. Crystallography and crystal chemistry. Holt, Rinehart and Winston, Inc., New York.)

TABLE 21.4 Radius of Cations Commonly Present in Phyllosilicates, Radius Ratio ($R_{cation}/R_{O^{2-}}$), and Predicted and Observed Coordination Numbers

Cation	Radius (nm)	$R_{cation}/R_{O^{2-}}$	Predicted Coordination Number	Observed Coordination Number
Cs^+	0.167	1.26	12	12
Rb^+	0.147	1.11	12	8, 12
NH^{4+}	0.143	1.08	12	8, 12
Ba^{2+}	0.134	1.02	8	8, 12
K^+	0.133	1	8	8, 12
Sr^{2+}	0.112	0.84	8	8
Ca^{2+}	0.099	0.75	6	6, 8
Na^+	0.097	0.73	6	6, 8
Mn^{2+}	0.08	0.6	6	6
Fe^{2+}	0.074	0.56	6	6
V^{3+}	0.074	0.56	6	6
Zn^{2+}	0.074	0.56	6	6
Cu^{2+}	0.072	0.54	6	6
Co^{2+}	0.072	0.54	6	6
Ni^{2+}	0.069	0.52	6	6
Li^+	0.068	0.51	6	6
Ti^{4+}	0.068	0.51	6	6
Mg^{2+}	0.066	0.5	6	6
Fe^{3+}	0.064	0.48	6	6
Cr^{3+}	0.063	0.47	4	6
Al^{3+}	0.051	0.38	4	4, 6
Si^{4+}	0.042	0.31	4	4
P^{5+}	0.035	0.26	4	4
B^{3+}	0.023	0.17	3	3

termed polytypes. If such a variety is caused by ionic substitutions, which are minor but consistent, they are called polytypoids.

To understand this specific feature of phyllosilicates, consider the magnesium mica, phlogopite, as an example. Its silicate layer in terms of a sequence of atom planes is given in Figure 21.10. In the figure, T_L represents the atom plane at the level of apical oxygen atoms (●, solid circles) of the lower tetrahedral sheet and at the same level and hydroxyl (double circles with O and ●) fills the center of a hexagonal array made up by six apical oxygen atoms. In addition, the projection to the plane of basal oxygen atoms (○, open circles) of the tetrahedral sheet is also included. O_T is the octahedral cation plane where magnesium atoms are represented by small solid circles (●). T_U corresponds to the atom plane of the upper tetrahedral sheet. Unlike T_L, the direction of apical oxygen atoms is opposite to that of T_L. In T_U, the apical oxygen atoms, hydroxyls, and the projections of the basal oxygen atoms are expressed by an open circle with shaded circle, double circles, and solid circles, respectively. The lower tetrahedral sheet and the upper tetrahedral sheet having opposing vertices to each other are held together by magnesium ions. Each magnesium ion is linked to two oxygen anions and one hydroxyl from the lower tetrahedral sheet and two oxygen anions and one hydroxyl from the upper sheet to make a six coordination packing. This packing mechanism requires that the upper tetrahedral sheet be staggered by $-a/3$, where a is the unit cell dimension along the a-axis, relative to the lower tetrahedral sheet. The staggering between the two sheets, therefore, gives a monoclinic nature to the silicate layer, and this nature is common to all the basic silicate layers of phyllosilicates. The interlayer atom plane is denoted as I, in which solid circles represent potassium atoms. At the interlayer position, for example, between two silicate layers, each interlayer potassium ion keys the two layers without shift by coordinating 12 basal oxygens, 6 each from lower and upper tetrahedral sheets. Although due to the presence of interlayer potassium ions there is no shift between the two silicate layers, the orientation of the upper one
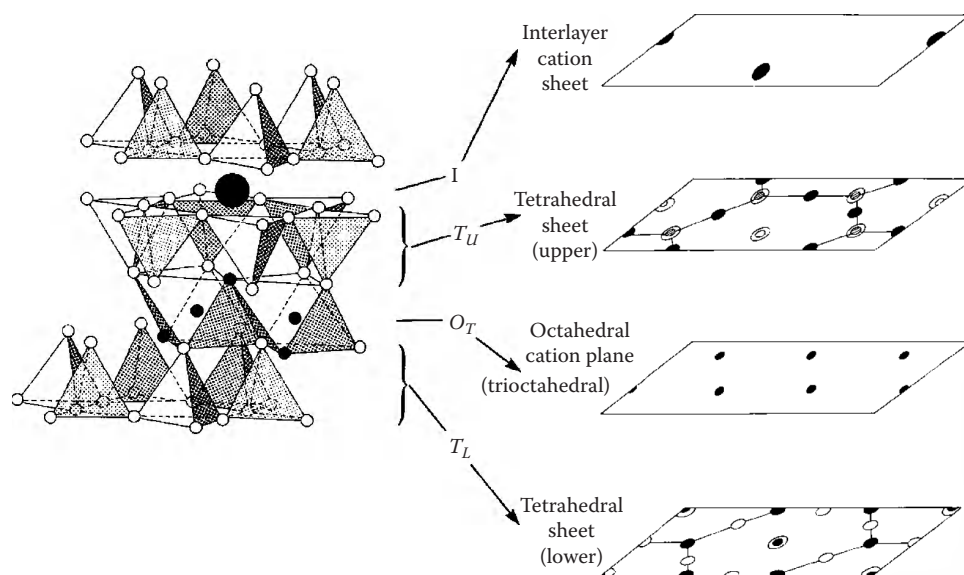
FIGURE 21.10 Idealized sketches of mica structure (phlogopite) and its structural segments (atomic planes) at four levels. (After Kodama, H. 1990. Use of color-coded transparencies for visualizing layer silicate structures, p. 169–175. *In* V.C. Farmer and Y. Tardy (eds.) Proc. 9th Int. Clay Conf. 1989. Strasbourg. Sci. Géol., Mém. Strasbourg, France.)

can be changed with respect to the lower one. Because these basal oxygen planes have a hexagonal or a pseudohexagonal symmetry, there are six or three possible orientations, which are related by $n \times 60°$ rotations, at maximum six angles: 0°, 60°, 120°, 180°, 240°, and 300°. Among them, angles 60° and 300° are equivalent, as also are angles 120° and 240°, because the silicate layer itself has the plane of symmetry, which is parallel to the layer stagger. Thus, 0°, 60°, 120°, and 180° rotation operations lead theoretically to six simple- ordered-stacking mica layers, and they were depicted by vectors from one potassium atom to a potassium atom in the next layer in the direction of increasing height of the mica layers (Smith and Yoder, 1956). Resulting layer structures having different layer-stacking sequences are called polytypes. Mica polytypes derived by them are $1M$ (0° rotation), $2M_1$ (a continuous alternation of ±120° rotation), $2M_2$ (a continuous alternation of ±60° rotation), $3T$ (±120° rotation), $2Or$ (±180° rotation), and $6H$ (±60° rotation), as expressed in an adapted simple nomenclature. The first symbol gives the number of layers in the repeat unit (subcell) and the second in italics gives the symmetry. In case of the same symmetry, the subscripts 1 and 2 are used for further distinction. Therefore, $1M$, $2M$, $3T$, $2Or$, and $6H$ indicate one-layer monoclinic, two-layer monoclinic, three-layer trigonal, two-layer orthorhombic, and six-layer hexagonal, respectively. Similar consideration can be applied to other types of layer silicates such as T–O and T–O–T–O. If polytypes are reported in nature, in the glossary of phyllosilicates given at the end of this chapter (compiled from Blackburn and Dennen, 1997, and Lalonde, 2003) a specific mineral name is given, for example, Phlogopite-$1M$, Kaolinite-$1Tc$, where the letter Tc represents triclinic. The descriptions stated above are based on the ideal or nearly ideal structure. In practice, however, there are many factors that affect the formation of polytypes, that is, ionic substitutions in the octahedral sheets, ionic substitutions for silicon atoms in the tetrahedral sheets, the degree of ordering of these substitutions, the vacancy distributions in the octahedral sheets, which normally exist in the dioctahedral layer silicates or may be caused by the oxidation of Fe^{2+} to Fe^{3+} in the sheets. These variations add more complexities to the form of polytypes.

21.2.3 1:1 Layer Type (T–O Type)

21.2.3.1 General

The structure of this type is shown in Figure 21.6 (also Figure 21.8), which was used for explaining the junction between tetrahedral and octahedral sheets. A 2D tetrahedral sheet, made up by sharing three corners of each SiO_4 tetrahedron, are attached to an octahedral sheet, thereby each cation in the octahedral sheet associates with two apical oxygen atoms and four hydroxyls to make up six coordination binding. Thus, the structure of this type of layer silicate consists of tetrahedral and octahedral sheets in which the anions at the exposed surface of the octahedral sheet are hydroxyls. General structural formula may be expressed by $Y_{2-3}Si_2O_5(OH)_4$, where Y are cations in the octahedral sheet such as Al^{3+} or Fe^{3+} for dioctahedral species and Mg^{2+}, Fe^{2+}, Mn^{2+}, or Ni^{2+} for trioctahedral species.

21.2.3.2 Serpentine–Kaolin Group

This group contains two subgroups, trioctahedral and dioctahedral. Serpentine minerals with the ideal formula of
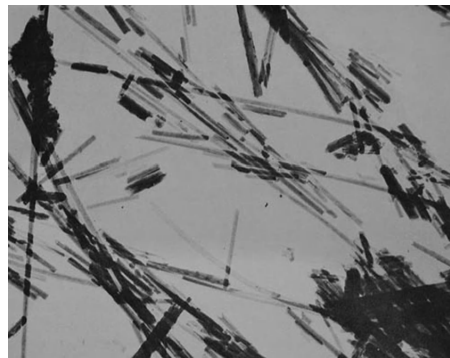
FIGURE 21.11 Electron micrograph of chrysotile. Most of the fibrils appear "solid," but a few show the appearance of hollow tubes, and a few show "open ends." (From Gard, J.A. (ed.). 1971. The electron-optical investigation of clays. Monograph No 3. Mineralogical Society, London, U.K. Reproduced with permission of The Mineralogical Society of Great Britain and Ireland.)

$Mg_3Si_2O_5(OH)_4$ represent the former, and kaolin minerals with $Al_2Si_2O_5(OH)_4$ represent the latter. Among trioctahedral magnesium species, chrysotile, antigorite, and lizardite are commonly known and their morphology and structure are uniquely different. (Wicks and Whittaker, 1975; Wicks and Zussman, 1975) Electron microscopic observations show that chrysotile crystals have a fibril appearance (Figure 21.11), but its cross sections clearly show cylindrical roll morphology (Figure 21.12a and b); whereas, antigorite crystals exhibit lath-shape morphology (Figure 21.13a), but the electron-diffraction photograph indicates an alternating wave structure as shown by the presence of segmental groups of superlattice spots (Figure 21.13b). The structure deduced by Kunze (1956) consists of curved layers elongated along the a-axis, but the layers change alternately their polarity (or the direction of apical oxygens in the tetrahedral sheets) at half a period of the wave to make up a full periodicity of 4.35 nm, which is the dimension of a of antigorite (Figure 21.14). These alternating wave structural characteristics may be attributed to the degree of fit between the lateral dimensions of the tetrahedral and octahedral sheets. On the other hand, lizardite crystals are platy (Figure 21.15) and have often a small amount of substitution of Al or Fe^{3+} for both Si and Mg. This substitution appears to be the main reason for the platy nature of lizardite. Obviously, morphology is one of the key characteristics to distinguish among serpentine minerals. A few selected chemical compositions of serpentine minerals are given in Table 21.5 for reference. Other than Mg, Fe^{2+}, Ni, and Mn can occupy the octahedral sites and form species such as berthierine, brindleyite, and kellyite. Chemical compositions of some such varieties are given in Table 21.6. Planar polytypes of the trioctahedral species are far more complicated than those of the dioctahedral ones, owing to the fact that the trioctahedral silicate layer has a higher symmetry because all octahedral cationic sites are occupied. Polytype structures, such as one-layer trigonal ($1T$), one-layer monoclinic ($1M$), two-layer hexagonal ($2H_1$, $2H_2$), two-layer trigonal ($2T$), two-layer orthorhombic ($2Or$), two-layer monoclinic ($2M_1$, $2M_2$), three-layer rhombohedral ($3R$),
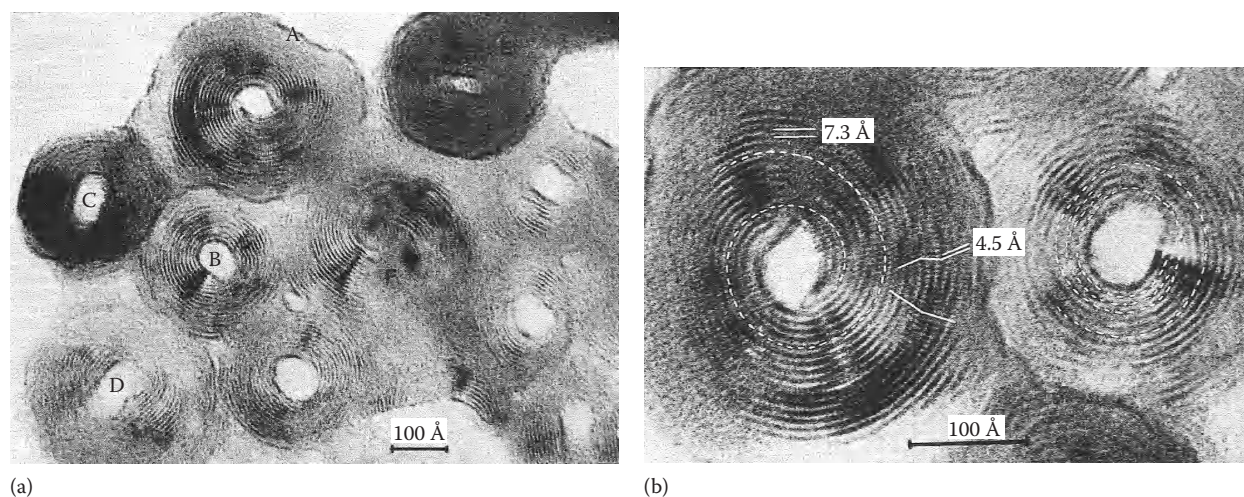
FIGURE 21.12 Electron micrographs of cross-sectioned chrysotile crystals: (a) The lattice image observed from the direction parallel to the fiber axis. (From Gard, J.A. (ed.). 1971. The electron-optical investigation of clays. Monograph No 3. Mineralogical Society, London, U.K.) (b) Enlarged image on crystallites A and B in micrograph (a) showing two kinds of fringe patterns, which correspond to d_{001}-spacing (0.73 nm) and d_{020}-spacing (0.45 nm). (From Gard, J.A. (ed.). 1971. The electron-optical investigation of clays. Monograph No 3. Mineralogical Society, London, U.K. Reproduced with permission of The Mineralogical Society of Great Britain and Ireland.)
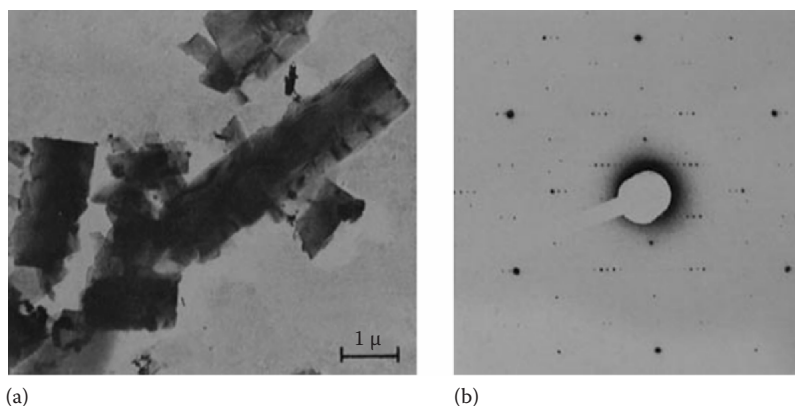
FIGURE 21.13 (a) Electron micrograph of antigorite (platy crystals), from Antigorio, Italy. (b) Electron-diffraction pattern from a single crystal of antigorite, showing the presence of a superlattice structure along the a-axis, with $a = \sim 3.85$ nm. (From Gard, J.A. (ed.). 1971. The electron-optical investigation of clays. Monograph No 3. Mineralogical Society, London, U.K. Reproduced with kind permission of The Mineralogical Society of Great Britain and Ireland.)
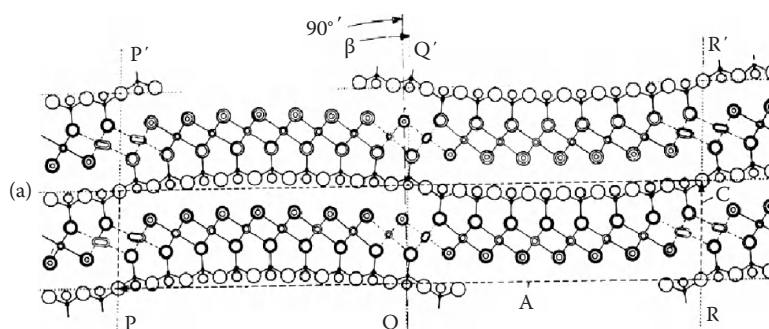
FIGURE 21.14 Structure of antigorite as viewed along the y-axis. The curved layers (radius of curvature 7.5 nm) reverse polarity at PP′, RR′, and near QQ′. (From Gard, J.A. (ed.). 1971. The electron-optical investigation of clays. Monograph No 3. Mineralogical Society, London, U.K. Reproduced with kind permission of The Mineralogical Society of Great Britain and Ireland.)
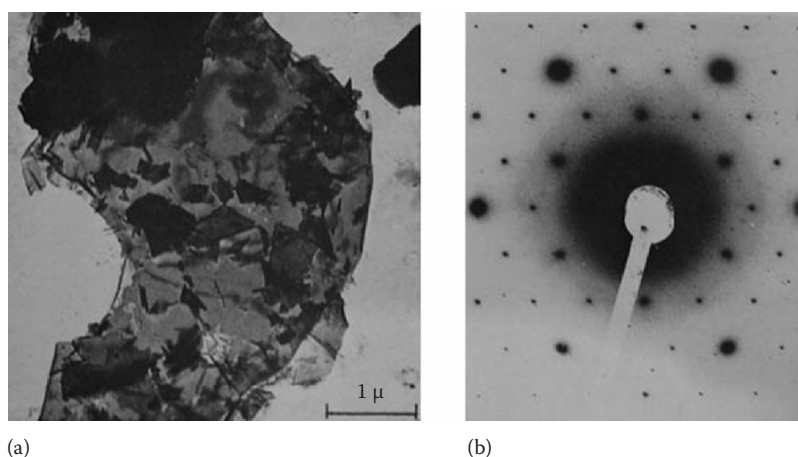
FIGURE 21.15 (a) Electron micrograph and (b) electron-diffraction pattern of lizardite, from Kennack, Cornwall (platy). (From Gard, J.A. (ed.). 1971. The electron-optical investigation of clays. Monograph No 3. Mineralogical Society, London, U.K. Reproduced with kind permission of The Mineralogical Society of Great Britain and Ireland.)

TABLE 21.5   Serpentine Mineral Analyses (1)

	1	2	3	4	5	6
SiO_2	43.37	41.25	41.65	42.02	42.44	43.45
Al_2O_3	0	0.54	0.1	0.52	0.64	0.81
TiO_2	0	0.02	nr	None	None	0.02
Fe_2O_3	0	1.32	2.88	0.19	0.19	0.88
FeO	0	0.09	0.16	0.11	0.03	0.69
MnO	0	0.07	0.05	0.03	0.03	None
MgO	43.64	41.84	41.06	41.44	42.76	41.9
CaO	0	0.02	nr	None	nr	nr
K_2O	0	nr	nr	0.49	0.08	0.02
Na_2O	0	nr	nr	0.36	0.06	0.05
H_2O^+	12.99	13.68	13.1	14.04	13.58	12.29
H_2O^-	0	0.97	nr	1.64	0.5	nr
Total	100	99.8	100.12	99.99	100.31	100.19
Numbers of cations on basis of $O_{10}(OH)_8$						
Si	4	3.92	3.94	4	3.96	3.99
Al	0	0.06	0.01		0.04	0.01
Fe^{3+}	0	0.02	0.05			
Σ_{tet}	4	4	4	4	4	4
Al	0			0.06	0.03	0.08
Fe^{3+}	0	0.07	0.15	0.01	0.01	0.06
Fe^{2+}	0	0.01	0.01	0.01		0.05
Mg	6	5.92	5.79	5.88	5.98	5.92
Σ_{oct}	6	6.01	5.95	5.96	5.98	5.92
Ca	0					
Na	0				0.01	
K	0				0.01	

1, Ideal composition based on the structural formula, $Mg_3Si_2O_5(OH)_4$; 2, lizardite, Transvaal, South Africa (Deer et al., 1962); 3, lizardite, Shetland Islands, United Kingdom (Brindley and von Knorring, 1954); 4, chrysotile, Gila County, Arizona (Faust and Fahey, 1962); 5, chrysotile, Montville, New Jersey (Faust and Fahey, 1962); 6, antigorite, Mikonui, New Zealand (Zussman, 1954). Total includes 0.04% NiO; 0.04% H_2O (—).

Analyses 2–6 are taken from Newman (1987) in which original data for those analyses were referred from respective papers cited above. See Newman (1987) for original source of each analysis.

TABLE 21.6   Serpentine Mineral Analyses (2)

	1	2	3	4
SiO_2	19.08	22.03	27.45	17.6
Al_2O_3	26.66	22.91	24.09	28.55
TiO_2	nr	3.63	0.99	<0.05
Fe_2O_3	4.29	0.46	nr	2.18
Cr_2O_3	nr	0.05	0.17	nr
FeO	34.52	36.68	1.15	
MnO	nr	0.04	nr	38.84
MgO	1.55	1.91	3.18	2.97
NiO	nr	nr	30.18	nr
CaO	nr	0.07	0.07	nr
K_2O	nr	0.03	nr	nr
Na_2O	nr	0.08	nr	nr
H_2O total	11.18			
H_2O^+	nr	10.65	nr	nr
H_2O^-	nr	0.63	nr	nr
Total	99.25	100.05	87.28	90.14
Numbers of cations on basis of $O_{10}(OH)_8$				
Si	2.2	3.23	3.01	1.96
Al	1.8	0.77	0.99	2.04
Fe^{3+}				
Σ_{tet}	4	4	4	4
Al	1.84	1.95	2.12	1.72
Fe^{3+}	0.37	0.03		0.18
Cr^{3+}			0.01	
Fe^{2+}	3.32	3.09	0.11	
Mn				3.67
Mg	0.27	0.29	0.52	0.49
Ni			2.66	
Σ_{oct}	5.8	5.36	5.42	6.06

1, Berthierine (Brindley, 1982); 2, berthierine (Brindley, 1982); 3, brindleyite, Marmara karstic bauxite deposits, Greece (Bish, 1978); 4, kellyite, Bald Knob, North Carolina (Peacor et al., 1974).

See Newman (1987) for original source of each analysis.

TABLE 21.7 Kaolin Mineral Analyses

	1	2	3	4	5	6
SiO_2	46.55	46.9	46.77	45.2	46.43	46.22
Al_2O_3	39.49	37.4	37.79	37.02	39.54	39.92
TiO_2	0	0.18	nr	1.26	Nil	0
Fe_2O_3	0	0.65	0.45	0.27	0.15	0
FeO	0	nr	0.11	0.06	nr	0
MgO	0	0.27	0.24	0.47	0.17	0
CaO	0	0.29	0.13	0.52	Nil	0
K_2O	0	0.84	1.49	0.49	0.02	0
Na_2O	0	0.44	0.05	0.36	0.03	0
H_2O^+	13.96	12.95	12.18	13.27	14.2	13.86
H_2O^-	0	nr	0.61	1.56	nr	0
Total	100	99.92	99.82	100.47	100.54	100
Numbers of cations on basis of $O_{10}(OH)_8$						
Si	4				3.982	3.969
Al	0				0.018	0.031
Σ_{tet}	4				4	4
Al	4				3.979	4.01
Fe^{3+}	0				0.01	
Fe^{2+}	0					
Mg	0				0.022	
Σ_{oct}	4				4.011	4.01

1, Ideal composition based on the structural formula, $Al_2Si_2O_5(OH)_4$; 2, kaolinite, Zettlitz, Czechoslovakia; 3, kaolinite, St. Austell, England; 4, kaolinite, Macon, Georgia; 5, dickite, Barkly East, Cape Province, South Africa; 6, nacrite, Eureka Tunnel, St. Peter's Dome, Colorado.

Analyses 1, 5, and 6 are taken from Newman (1987) in which original data for analyses 5 and 6 (recalculated after subtracting impurity) were referred from Schmidt and Heckroodt (1959) and Blount et al. (1969), respectively. Analyses 2, 3, and 4 are taken from Grim (1968) in which original data for the analysis 2 is referred from Ross and Kerr (1931) and analyses 3 and 4 from Kerr et al. (1950). For these references, see Grim (1968).

TABLE 21.8 Halloysite Mineral Analyses

	1	2	3
SiO_2	40.09	46.2	44.7
Al_2O_3	35.38	39.84	28.1
TiO_2	nr	0.02	
Fe_2O_3	tr	0.17	12.8
FeO	nr	nr	nr
MgO	tr	0.02	0.1
CaO	0.77	0.34	tr
K_2O	tr	0.02	tr
Na_2O	0.1	0.01	1.7
H_2O total			
H_2O^+	15	14	13.7
H_2O^-	8.61		
Total	100.51	100.62	100.7
Numbers of cations on basis of $O_{10}(OH)_8$			
Si	3.907	3.957	4.029
Al	0.093	0.043	
Fe^{3+}			
Σ_{tet}	4	4	4.029
Al	3.971	3.978	2.985
Ti		0.001	
Fe^{3+}		0.011	0.868
Mg		0.003	0.013
Σ_{oct}	3.971	3.993	3.866
Ca	0.08	0.031	0
K	0	0.002	0
Na	0.019	0.002	0.297

1, Hydrated halloysite (halloysite—1 nm), Wagon Wheel Gap, Colorado (Larsen and Wherry, 1917); 2, halloysite (halloysite—0.7 nm), Djebel Debar, Morocco (Garrett and Walker, 1959); 3, iron-rich hydrated halloysite, Hokkaido, Japan (Wada and Mizota, 1982).

See Newman (1987) for original data.

three-layer trigonal (3T), six-layer hexagonal (6H), and six-layer rhombohedral (6R) are reported (Bailey, 1969, 1980, 1988).

For the dioctahedral subgroup, kaolin minerals include kaolinite (one-layer triclinic, 1Tc), dickite (two-layer monoclinic, 2M), and nacrite (two-layer monoclinic, 2M), which are in polytypic relation. Examples of the chemical composition of kaolin minerals are given in Table 21.7. Kaolinite has triclinic symmetry. Oxygen atoms and hydroxyl ions between the layers are paired with hydrogen bonding. Because of this weak bonding, random displacements between the layers are quite common and result in a poor symmetry rather than distinctive triclinic kaolinite. Dickite and nacrite are polytypic varieties of kaolinite. Both of them consist of a double 1:1 layer and have monoclinic symmetry but distinguish themselves by different stacking sequences of the two 1:1 silicate layers.

21.2.3.3 Halloysite—Hydrated Halloysite Group

Halloysite has a composition close to that of kaolinite (Table 21.8) and is characterized by its tubular nature (see Figure 21.27) in contrast to the platy nature of kaolinite particles. Although tubular forms are the most common, other morphological varieties are also known: prismatic, rolled, pseudospherical (Figure 21.16), and platy forms. The structure of halloysite is believed to be

FIGURE 21.16 Electron micrograph of halloysite (with spherical structure), altered from glassy tuff. (From Sudo, T. 1959. *Mineralogical study on clays of Japan*. Maruzen Co. Ltd., Tokyo, Japan.)

similar to that of kaolinite, but no precise structure has been revealed yet. Halloysite has a hydrated form with composition of $Al_2Si_2O_5(OH)_4 \cdot 2H_2O$. This hydrated form irreversibly changes to a dehydrated variety at relatively low temperatures (60°C) or upon being exposed to conditions of low relative humidity. The dehydrated form has a basal spacing about the thickness of a kaolinite layer (approximately 0.72 nm), and the hydrated form has a basal spacing of about 1.01 nm, the difference of 0.29 nm is approximately the thickness of one molecular layer of water. Consequently, in the hydrated form of halloysite, its silicate layers are separated by monomolecular water layers that are lost during dehydration.

21.2.4 2:1 Layer Type (T–O–T Type)

21.2.4.1 General

As seen in Figure 21.8b, the unit silicate layer of T–O–T type comprises one octahedral sheet sandwiched by two tetrahedral sheets, which are oriented in opposite directions and may be termed a 2:1 layer type. A majority of phyllosilicates hold a backbone of this type for their structures. This type includes talc, pyrophyllite, micas, brittle micas, vermiculite, and smectite.

21.2.4.2 Talc–Pyrophyllite Group

Minerals of this group have the simplest form of 2:1 layer with its unit thickness of approximately 0.92–0.96 nm—that is, the structure consists of an octahedral sheet sandwiched by two tetrahedral sheets. Talc and pyrophyllite represent the trioctahedral and dioctahedral members, respectively, of the group. In the ideal case, the structural formula of talc is expressed by $Mg_3Si_4O_{10}(OH)_2$ and pyrophyllite by $Al_2Si_4O_{10}(OH)_2$. Selected analyses are given in Table 21.9. Figure 21.17 shows these two structures as an extended crystal model, where the staggering between tetrahedral sheets above and below and the difference in octahedral sheets between talc and pyrophyllite is clearly seen. Each 2:1 layer of these minerals is electrostatically neutral; therefore, van der Waals forces hold the silicate layers together. One-layer triclinic and two-layer monoclinic forms are known for polytypes of pyrophyllite and talc. The ferric iron analog of pyrophyllite is called ferripyrophyllite.

21.2.4.3 True Mica Group

Mica has a basic structural unit of 2:1 layer type, like pyrophyllite or talc, but some of the silicon atoms (ideally one-fourth) are always replaced by aluminum atoms. This results in a charge deficiency that is balanced by potassium ions between the unit layers (Figure 21.18). The sheet thickness (basal spacing or dimension along the direction normal to the basal plane) is fixed at about 1 nm. Typical examples are muscovite, $KAl_2(Si_3Al)O_{10}(OH)_2$ for dioctahedral species and phlogopite, $KMg_3(Si_3Al)O_{10}(OH)_2$, and biotite, $K(Mg,Fe)_3(Si_3Al)O_{10}(OH)_2$, for trioctahedral species. Formulas rendered may vary slightly due to possible substitution within certain structural sites. The interlayer cation can be sodium instead of potassium, notably paragonite

TABLE 21.9 Talc and Pyrophyllite Analyses

	1	2	3	4	5	6
SiO_2	63.37	62.67	63.9	63.57	66.04	66.7
Al_2O_3		tr	0.03	29.25	28.15	28.3
TiO_2		nr	0.1	0.04	nr	
Fe_2O_3		nr	0.21	0.1	0.64	
FeO		2.46	nr	0.12	nr	
MnO		0.01	nr	None	nr	
MgO	31.88	30.22	31.49	0.37	0.04	
CaO		nr	0.08	0.38	0.01	
K_2O		nr	0.01	0.02	nr	
Na_2O		nr	0.02	tr	0.04	
H_2O^+	4.75	4.72	4.86	5.66	5.27	5
H_2O^-				0.66		
Total	100	100.02	100.7	100.17	100.19	100
Numbers of cations on basis of $O_{10}(OH)_2$						
Si	4	3.99	4.01	3.89	3.975	4
Al	0			0.11	0.025	
Fe^{3+}	0					
Σ_{tet}	4	3.99	4.01	4	4	4
Al	0			1.99	1.975	2
Fe^{3+}	0		0.01	0.005	0.03	
Fe^{2+}	0	0.13		0.005		
Mg	3	2.87	2.94	0.035	0.005	
Σ_{oct}	3	3	2.95	2.036	2.01	2
Ca	0		0.005	0.025		
Na	0				0.005	
K	0					

1, Ideal composition, $Mg_3Si_4O_{10}(OH)_2$; 2, talc, Muruhatten, Northern Sweden (Du Rietz, 1935); 3, talc, Manchuria (Brindley et al., 1977); 4, pyrophyllite, pale blue, Honami mine, Nagano Prefecture, Japan (Iwao and Udagawa, 1969); 5, pyrophyllite, Ibitiara Bahia, Brazil (Lee and Guggenheim, 1981); 6, ideal composition, $Al_2Si_4O_{10}(OH)_2$.

Analysis 2 is referred from Deer et al. (1962). Analyses 3, 4, and 5 are taken from Newman (1987). See Newman (1987) for original source of each analysis except analyses 1 and 2.

in this case, which has an ideal formula of $NaAl_2(Si_3Al)O_{10}(OH)_2$. Various polytypes of the micas are known to occur. Among them one-layer monoclinic ($1M$), two-layer monoclinic ($2M$ including $2M_1$ and $2M_2$), and three-layer trigonal ($3T$) polytypes are the most common. The majority of clay-size micas are dioctahedral aluminous species; those similar to muscovite are called illite (Bailey et al., 1984). The illites are different from muscovite in that the amount of substitution of aluminum for silicon is less, sometimes only one-sixth of the silicon ions are replaced. This reduces a net unbalanced charge deficiency from 1 to about 0.65 per chemical formula unit. As a result, the illites have a higher silica-to-alumina molecular ratio and lower potassium content than the muscovites (Graf von Reichenbach and Rich, 1975; Fanning et al., 1989). To some extent, octahedral aluminum ions are replaced by magnesium (Mg^{2+}) and iron ions (Fe^{2+}, Fe^{3+}). In the illites, stacking disorders of the layers are common, but their polytypes are often unidentifiable. Celadonite and glauconite are ferric ion-rich species of dioctahedral micas. The ideal

(a)   (b)
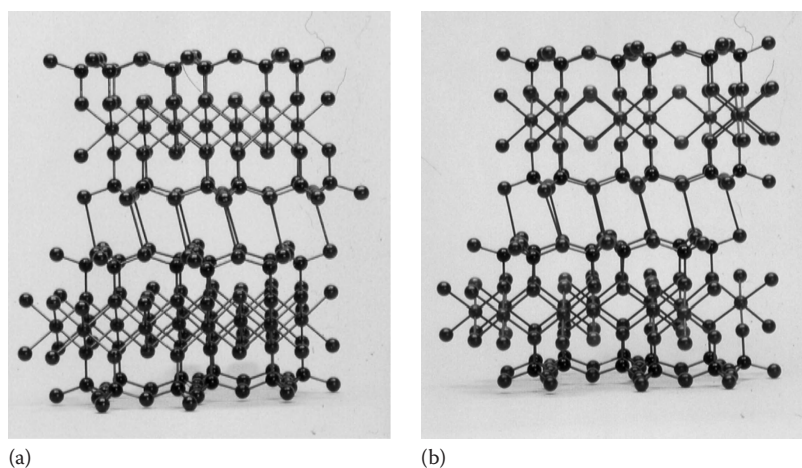
FIGURE 21.17   Extended crystal structure models of (a) talc and (b) pyrophyllite.

| Oxygens | (OH) Hydroxyls | ● Aluminum | Potassium |
| And ● | Silicons (one-fourth replaced by aluminums) | | |
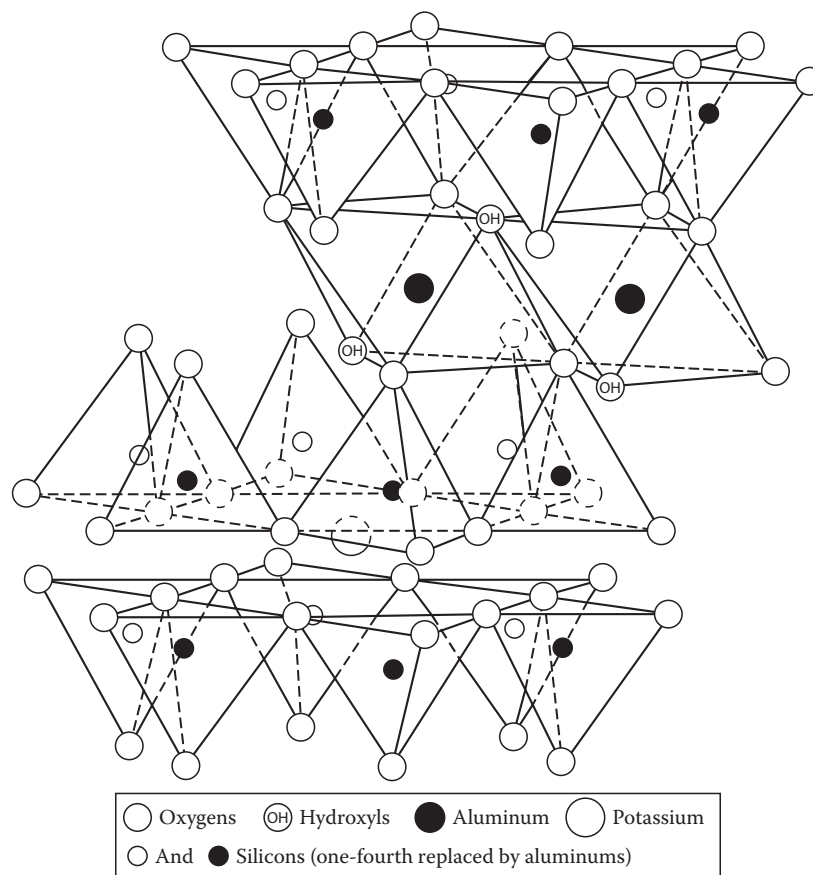
FIGURE 21.18   The structure of muscovite. (From Grim, R.E. 1968. Clay mineralogy. McGraw-Hill Book Co. Inc., New York. With permission.)

composition of celadonite may be expressed by $K(Mg,Fe^{3+})$ $(Si_{4-x}Al_x)O_{10}(OH)_2$, where $x = 0–0.2$. Glauconite is a dioctahedral mica species with tetrahedral Al substitution greater than 0.2 and octahedral Fe^{3+} or R^{3+} (total trivalent cations) greater than 1.2. Unlike illite, a layer charge deficiency of celadonite and glauconite arises largely from the unbalanced charge due to ionic substitution in the octahedral sheets. Some chemical analyses of dioctahedral micas (Tables 21.10 and 21.11) and of trioctahedral

micas (Table 21.12) are given for reference. Comprehensive reviews on true micas are available (Bailey, 1984; Fleet, 2003).

21.2.4.4 Brittle Mica Group

Brittle micas are distinguished from true micas in having a layer charge of ~2.0 per formula unit, which arises entirely within the tetrahedral sheet or partly within the tetrahedral sheet and partly octahedral sheet. The tetrahedral composition varies

TABLE 21.10 Dioctahedral Mica Analyses (1)

	1	2	3	4	5	6	7	8
SiO_2	45.26	45.87	45.24	42.63	44.94	48.42	51.19	55.61
Al_2O_3	38.4	38.69	36.85	20.87	27.56	27.16	9.23	0.79
TiO_2			0.01	0.45	0.98	0.87	0.13	
Cr_2O_3				0.24	4.75			
V_2O_3				13.72				
Fe_2O_3			0.09			6.57	18.15	17.19
FeO		0	0.02	3.76	1.07	0.81	1.78	4.02
MnO		0	0.12	0.23	0		0.01	0.09
MgO		0.1	0.08	2.34	2.34	0	3.34	7.26
CaO			0	0.01	0	0	0.58	0.21
BaO				0.95	6.42	0.07		
K_2O	11.82	10.08	10.08	10.16	7.44	11.23	7.98	10.03
Na_2O		0.64	0.64	0.06	0.27	0.35	0.02	0.19
Cs_2O			0.2					
Rb_2O			0.93			0.19		
Li_2O			0.49					
F			0.91					
Cl								
H_2O^+	4.52	4.67	4.12			4.31	5.21	4.88
H_2O^-		0.7	0.09			0.19	1.95	
O≡F,Cl	100	100.05	100.24 −0.38	95.21	95.77	100.17	99.67	100.27
Total	100	100.05	99.86	95.21	95.77	100.17	99.67	100.27
Numbers of cations on basis of $O_{10}(OH)_2$								
Si	3	3.009	3.024	3.061	3.136	3.297	3.71	3.995
Al	1	0.991	0.976	0.939	0.864	0.703	0.29	0.005
Fe^{3+}	0							
Σ_{tet}	4	4	4	4	4	4	4	4
Al	2	2.001	1.927	0.827	1.403	1.477	0.505	0.065
Cr					0.262	0.045		
Fe^{3+}	0		0.005	0.005		0.337	0.99	0.93
Fe^{2+}	0	0	0.001	0.226	0.062	0.046		0.24
Mg	0	0.01	0.008	0.25	0.243		0.36	0.775
Li			0.132					1.06
Other M			0.008		0.051			
Σ_{oct}	2	2.011	2.081	2.132	2.021	1.905	1.965	2.015
Ca	0		0.005	0.001			0.045	0.015
Ba				0.027	0.176	0.002		
Na	0	0.081		0.008	0.037	0.046	0.005	0.025
Rb						0.008		
K	1	0.844		0.931	0.662	0.976	0.735	0.92
Σ_{Int}		0.925			0.875	1.032	0.785	0.965

1, Ideal muscovite composition, $KAl_2(Si_3Al)O_{10}(OH)_2$; 2, muscovite, Methuen Township, Ontario (Hurlbut, 1956); 3, muscovite (rose), Rociada, New Mexico (Heinrich and Levinson, 1953); 4, muscovite (vanadian), Hemlo gold deposit, Ontario (Pan and Fleet, 1992); 5, muscovite (barian–chromian), Hemlo-Heron Bay (Pan and Fleet, 1991); 6, phengite (high-silica muscovite), Morar, Inverness-shire (Lambert, 1959); 7, glauconite, Praha-Prosek (Cimbalnikova, 1971); 8, celadonite, Reno, Nevada (Hendricks and Ross, 1941).

Analyses 2–6 are taken from Fleet (2003). Analyses 7 and 8 are from Newman 1987 and recalculated. See Fleet (2003) or Newman (1987) for original source of each analysis.

TABLE 21.11 Dioctahedral Mica Analyses (2)

	1	2	3
SiO$_2$	48.34	41.79	38.54
Al$_2$O$_3$	37.87	18.98	25.61
TiO$_2$	0.3	0.21	0
Cr$_2$O$_3$		0.29	
V$_2$O$_3$		17.55	
Fe$_2$O$_3$	1.02		
FeO		1.15	2.44
MnO		0.08	0.07
MgO	0.11	2.18	1.78
CaO	0	0	0.01
BaO		3.29	
K$_2$O	3.25	9.76	0.53
Na$_2$O	0	0.16	
Cs$_2$O			25.29
Rb$_2$O			0.25
Li$_2$O			0.44
NH$_4$	3.85		
F			1
H$_2$O$^+$	4.96		3.27
H$_2$O$^-$	0.31		
OΞF,Cl	100.01	95.44	99.23
	0	0	−0.42
Total	100.01	95.44	98.81
Numbers of cations on basis of O$_{10}$(OH)$_2$			
Si	3.078	3.047	3.161
Al	0.922	0.953	0.839
Fe^{3+}			
Σ$_{tet}$	4	4	4
Al	1.92	0.679	1.637
Ti	0.014	0.012	
Cr		0.017	
V		1.026	
Fe^{3+}	0.049		
Fe^{2+}		0.07	0.167
Mn		0.005	0.005
Mg	0.01	0.237	0.218
Li			0.145
Other *M*			
Σ$_{oct}$	1.993	2.046	2.172
Ca			0.001
Na		0.081	
K	0.264	0.844	0.055
Rb			0.013
Cs			0.885
NH$_4$	0.566		
Σ$_{Int}$	0.83	0.925	0.954

1, Tobelite, Horo deposit, Toyosaka Hiroshima Prefecture, Japan (Higashi, 1982); 2, roscoelite, Hemlo gold deposit, Ontario, Canada (Pan and Fleet, 1992); 3, nanpingite, Nanping pegmatite field, Fujian, China (Yang et al., 1988).

Analyses 1, 2, and 3 are taken from Fleet (2003). See Fleet (2003) for original source of data.

from SiAl$_3$ to nearly Si$_3$Al and this is supplemented by octahedral sheet charges that range from +1.0 to 0 to nearly −1.0. As the layer charge is higher than that of true micas, divalent interlayer cations are required for compensation, instead of monovalent.

Clintonite is a trioctahedral species of ideal composition, Ca(Mg$_2$Al)(SiAl$_3$)O$_{10}$(OH)$_2$. Because high Al content in the tetrahedral sheet and Al atomic radius size is larger than Si, the hexagonal symmetry of the tetrahedral sheet is modified to ditrigonal (Figure 21.4b) by a tetrahedral rotation of 23° to match the lateral dimensions of the octahedral sheet. The reported polytypes are 1*M*, 2*M*$_1$, and 3*T*. Margarite is a known sole dioctahedral species in the brittle mica group, having ideal composition CaAl$_2$(Si$_2$Al$_2$)O$_{10}$(OH)$_2$, and with a 2*M*$_1$ structure. Other species belonging to the group include kinoshitalite (1*M* and 2*M*$_1$), with ideal formula BaMg$_3$(Si$_2$Al$_2$)O$_{10}$(OH)$_2$; anandite (2*Or*), with Ba(Fe,Mg)$_3$(Si$_3$Fe^{3+})O$_{10}$(OH)S; and bityite (2*M*$_1$), with BaLiAl$_2$(Si$_2$AlBe)O$_{10}$(OH)$_2$. Polytypes in brackets are only of high abundance. Analyses of representative brittle micas are given in Table 21.13.

21.2.4.5 Vermiculite Group

The vermiculite unit structure consists of talc-like 2:1 silicate layers separated by two molecular layers of water (approximately 0.48 nm thick). Substitutions of aluminum (Al^{3+}) for silicon (Si^{4+}) in tetrahedral sheets constitute the chief charge imbalance, but the net-charge deficiency may be partially balanced by other substitutions within the 2:1 silicate layer; but a residual net-charge deficiency always exists, commonly in the range 0.6–0.8 per O$_{10}$(OH)$_2$. This charge deficiency is satisfied with interlayer cations, which are closely associated with the water molecules between the silicate layers (Figure 21.19); Mathieson and Walker, 1954. As the net-charge deficiency is higher than smectites, in vermiculites interlayer water molecules are better coordinated with interlayer cations. In macroscopic vermiculites of hydrothermal origin, unlike those in soils, the balancing cation is mostly magnesium (Mg^{2+}), sometimes associated with a small amount of calcium (Ca^{2+}). The interlayer cation, however, is readily replaced by other inorganic and organic cations. A number of water molecules are related to the hydration state of cations located at the interlayer sites. Therefore, the basal spacing of vermiculite changes from about 1.05 to 1.57 nm, depending upon its crystal size, relative humidity, and the kind of interlayer cation. If potassium or ammonium ions are not present in the interlayer sites, in some extreme cases, heating vermiculite to temperatures as high as 500°C drives the water out from between the silicate layers, but the mineral quickly rehydrates at room temperature to maintain its normal basal spacing of approximately 1.4–1.5 nm. Special care would, therefore, be required to examine the dehydration–rehydration temperature of a vermiculite mineral in question.

If complete and irreversible dehydration occurs, the basal spacing becomes 0.902 nm. Figure 21.20 illustrates various hydration states of vermiculite. Note, however, that these states are affected by crystal size of the sample. Some dioctahedral analogs of vermiculite have also been reported to occur in soils (Brown, 1953).

TABLE 21.12   Trioctahedral Mica Analyses

	1	2	3	4	5	6	7	8
SiO_2	43.21	42.98	39.17	36.05	33.96	35.21	50.31	45.51
Al_2O_3	12.2	12.9	11.24	14.2	13.1	9.96	19.95	20.59
TiO_2		0.33	2.23	4.38	3.55	0.87	0.22	0.24
Fe_2O_3		0.91	1.86	6.24	3.06	6.57	0.49	1.51
FeO		2.7	16.58	13.35	32.04	42.11	2.55	9.2
MnO		0.08	0.89	0.23	0.65		2.63	0.53
MgO	28.99	25.93	13.51	12.23	0.97		0.02	0.3
CaO		0.06	0.2		0.42		0	0.27
BaO			0.05					
K_2O	11.29	10.63	9.29	9.36	8.47	9.2	10.14	9.68
Na_2O		0.25	0.62	0.14	0.27		0.49	0.6
Cs_2O							0.06	
Rb_2O							0.97	
Li_2O			0.18	0.02			5.39	3.66
F		6.04	3.46		0.82		7.65	7.45
Cl				0.41	0.34			
H_2O^+	4.31	1.7	1.64	3.21	2.5	3.52	0.88	1.9
H_2O^-		0.7	0.09				0.66	1.21
O≡F,Cl	100	105.21	101.28	100	100.37	100	102.41	102.65
	0	−2.54	−1.48	−0.09	−0.43		−3.22	−3.14
Total	100	100.67	99.8	99.91	99.95	100	99.19	99.51
Numbers of cations on basis of $O_{10}(OH)_2$								
Si	3	2.985	3.024	2.742	2.812	3.297	3.71	3.285
Al	1	1.015	0.976	1.258	1.188	0.703	0.29	0.715
Fe^{3+}	0							
Σ_{tet}	4	4	4	4	4	4	4	4
Al	2	0.04	1.927	0.015	0.091	1.477	0.505	1.035
Cr						0.045		
Fe^{3+}	0	0.05	0.005	0.357	0.191	0.337	0.99	0.08
Fe^{2+}	0	0.155	0.001	0.861	2.219	0.046		0.555
Mg	0	2.685	0.008	1.387	0.12		0.36	0.03
Li			0.132	0.006				1.06
Other M		0.02	0.008		0.051			
Σ_{oct}	2	2.95	2.081	2.892	2.888	1.905	1.965	2.81
Ca	0	0.005	0.005		0.037		0.045	0.02
Ba				0.027		0.002		
Na	0	0.035		0.021	0.079	0.046	0.005	0.085
Rb						0.008		
K	1	0.94		0.908	0.895	0.976	0.735	0.89
Σ_{Int}		0.98		0.929	1.011	1.032	0.785	0.995

1, Ideal phlogopite composition, $KMg_3(Si_3Al)O_{10}(OH)_2$; 2, phlogopite, Ontario (Newman, 1987); 3, biotite, North Burgers, Ontario (Rausell Colom et al., 1965): Total includes 0.12% Cl and 0.05% Rb_2O; 4, biotite, Valle Vercelli, Northwestern Italy (Brigatti and Davoli, 1990); 5, annite, Pikes Peak granite, Colorado (Hazen and Burnham, 1973); 6, ideal annite composition $KFe_3^{2+}(Si_3Al)O_{10}(OH)_2$; 7, lepidolite, Wakefield, Quebec, Canada (Steven, 1938; Backhaus, 1983); 8, zinnwaldite, Cinovec, Czechoslovakia (Rieder, 1970).

Analyses 2, 3, and 8 are taken from Newman 1987. Analyses 4, 5, and 7 are from Fleet (2003). See Newman (1987) or Fleet (2003) for original source of each analysis.

TABLE 21.13   Brittle Mica Analyses

	1	2	3	4	5	6
SiO_2	15.84	24.58	30.34	29.9	24.59	33.37
Al_2O_3	44	20.06	49.8	25.9	0.8	36.51
TiO_2	nr	0.16	0.12		0.08	
V_2O_4				5.4		
V_2O_3				18.3		
Fe_2O_3	2.04	0.71	0.35	0.4	21.15	
FeO	nr	0.04	0.51		22.7	
BeO						7.3
Mn_2O_3		3.24				
MnO	nr	7.38	0.01		0.92	
MgO	20.33	16.6	0.59	1.6	2.89	0.04
CaO	13.01	0.05	11.43		0.01	14.42
BaO		17.85		9.35	23.05	
Li_2O			0.28			2.39
K_2O	0.19	3.3	0.27	0.7	0.24	0.04
Na_2O	0.26	0.68	1.95	1.32	0.06	0.29
F		0.21				
H_2O^+	4.56	2.9	4.65	6	5.27	5.72
H_2O^-		0.2		1.1		
$O\Xi F$		−0.09				
Total	100.23	99.87	100.3	99.97	99.38	99.98

Numbers of cations on basis of $O_{10}(OH)_2$

	1	2	3	4	5	6
Si	1.108	2.052	2.012	2.192	2.507	
Al	2.892	1.948	1.988	1.808	0.096	
Fe^{3+}	0				1.397	
Σ_{tet}	4	4	4	4	4	
Al	0.734	0.222	1.903	0.431		
Ti		0.01	0.006		0.006	
V^{4+}				0.287		
V^{3+}				1.076		
Fe^{3+}	0.107	0.045	0.017	0.022	0.225	
Fe^{2+}	0	0.003	0.028		1.935	
Mn^{3+}		0.206			0.037	
Mn^{2+}		0.522	0.001		0.042	
Mg	2.119	2.066	0.058	0.175	0.439	
Li			0.075			
Σ_{oct}	2.96	3.074	2.088	1.991	2.684	
Ca	0.975	0.004	0.812			
Ba		0.584		0.269	0.921	
Na	0.035	0.11	0.251	0.188	0.012	
K	0.017	0.351	0.023	0.065	0.031	
Σ_{int}	1.027	1.029	1.086	0.522	0.964	

1, Clintonite, Zlatoust, Shishim Mts., Ural, Russia (Forman et al., 1967); 2, kinoshitalite, Nodatamagawa mine, Japan (Yoshii et al., 1973); 3, margarite, Chester, Massachusetts (Storre and Nitsch, 1974); 4, chernykhite, Karatau Mts., Kazakhstan (Ankinovich et al., 1972); 5, anandite, Wilagedera, Sri Lanka (Filut et al., 1985); 6, bityite, Londonderry, Australia (Anthony et al., 1995).

Bityite: ideally $CaLiAl_2(AlBeSi_2)O_{10}(OH)_2$. Analyses 1–5 are taken from Fleet (2003). See Fleet (2003) for original source of each analysis.
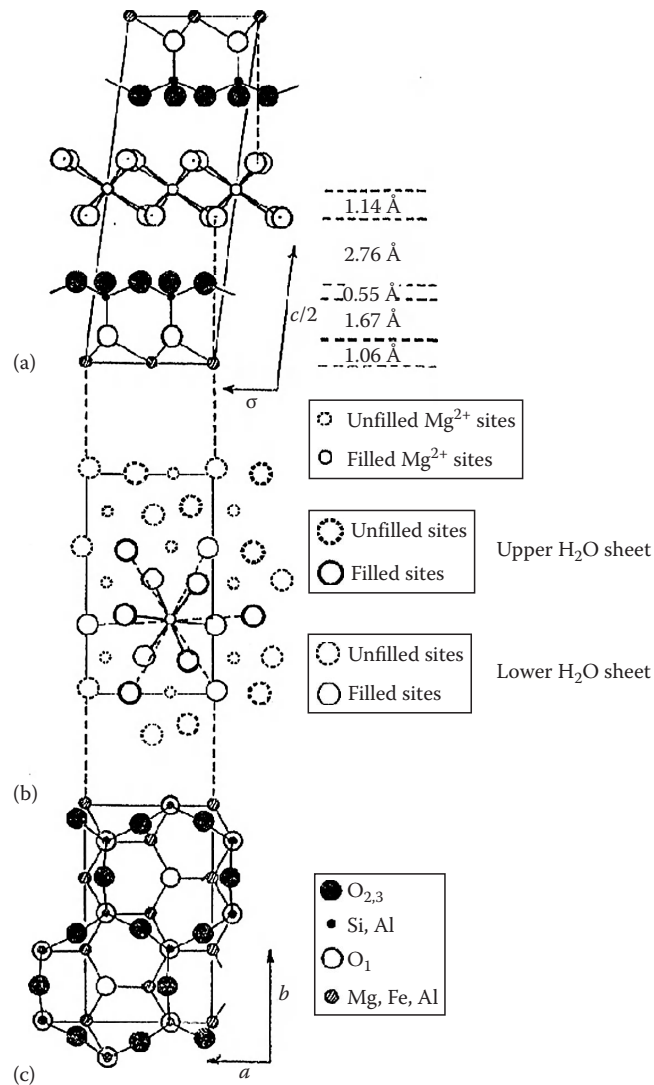
FIGURE 21.19 The crystal structure of Mg-vermiculite; (a) projection normal to *ac* plane, (b) projection normal to *ab* plane, showing interlayer region, and (c) projection normal to *ab* plane, showing one-half of a silicate layer (*z* = 0 to *c*/8). (From Brown, G. (ed.). 1961. The X-ray identification and crystal structures of clay minerals. Mineralogical Society, London, U.K. Reproduced with kind permission of The Mineralogical Society of Great Britain and Ireland.)

21.2.4.6 Smectite Group

The structural units of smectite (Figure 21.21) can be derived from the structures of pyrophyllite and talc. Unlike pyrophyllite and talc, the 2:1 silicate layers of smectite have a slight negative charge owing to ionic substitutions in the octahedral and tetrahedral sheets. The net-charge deficiency is normally smaller than that of vermiculite from ~0.2 to 0.6 per $O_{10}(OH)_2$ and balanced by the interlayer cations as in vermiculite (Table 21.14). This weak bond offers an excellent cleavage between the layers. The distinguishing feature of the smectite structure from the other types is that water and other
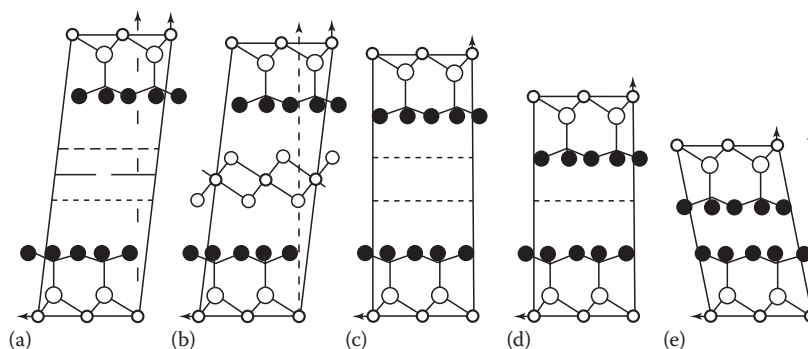
FIGURE 21.20 Various hydration states of Mg-vermiculite. Projections normal to the *ac* plane in the mineral at the various stages of hydration, showing the silicate layer relationships: (a) 1.481 nm phase, (b) 1.436 nm phase, (c) 1.382 nm phase, (d) 1.159 nm phase, and (e) 0.902 nm phase (key as for Figure 21.19). (From Brown, G. (ed.). 1961. The X-ray identification and crystal structures of clay minerals. Mineralogical Society, London, U.K. Reproduced with kind permission of The Mineralogical Society of Great Britain and Ireland.)

Exchangeable cation
nH$_2$O

○ Oxygen	(OH) Hydroxyl
● Aluminum, iron, magnesium	
○ ● Silicon, occasionally aluminum	
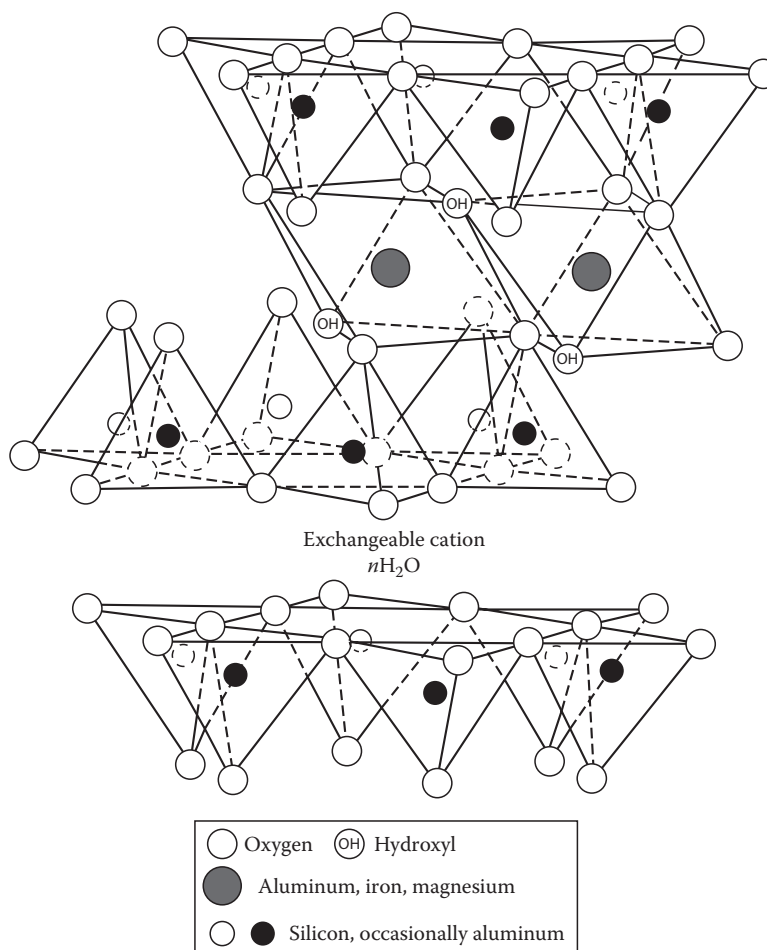
FIGURE 21.21 The structure of smectite. (From Grim, R.E. 1968. Clay mineralogy. McGraw-Hill Book Co. Inc., New York. With permission.)

polar molecules (in the form of certain organic substances) can, by entering between the unit layers, cause the structure to expand in the direction normal to the basal plane. Thus, this dimension may vary from about 0.96 nm, when no polar molecules are between the unit layers, to nearly complete separation of the individual layers.

The structural formula of dioctahedral aluminous smectites may be represented by $M_{x+y}^+ (Al_{2-y}Mg_y^{2+})(Si_{4-x}Al_x)O_{10}$ $(OH)_2 \cdot nH_2O$, with $0.2 \leq x + y \leq 0.6$, where M^+ is the interlayer exchangeable cation expressed as a monovalent cation and x and y are the amounts of tetrahedral and octahedral substitutions, respectively. The smectites with $y > x$ are called montmorillonite and those with $x > y$ are known as beidellite. In the latter type of smectites, those in which ferric iron is a dominant cation in the octahedral sheet instead of aluminum and magnesium are called

TABLE 21.14 Smectite and Vermiculite Analyses

	1	2	3	4	5	6	7	8
SiO_2	59.6	66.32	59.3	42.4	53.88	51.4	45.56	43.07
Al_2O_3	22.17	22.42	36.11	5.6	4.47	9	15.82	16.92
TiO_2	0.09				0.25		0.38	0.68
Fe_2O_3	4.32	3.3	0.5	32.53	0.6	5.4	1.4	11.75
FeO						4.8		0.95
MnO	Tr						0.13	0.36
MgO	2.73	4.07	0.1	0.32	31.61	26.1	29.66	22.63
CaO	0.14	0.03	0.02	0.38	0.01	3.2	6.86	4.66
K_2O	0.03	0.06	0.11	5.14	0.05	0.12	0.17	0.03
Na_2O	3.18	3.68	3.98			0.01	0.04	0.06
H_2O^+	6.02				9.28			
Total H_2O				14.03				
Total	100.17	99.88	100.12	100.02	100.15	100	99.98	101.11
Numbers of cations on basis of $O_{20}(OH)_4$								
Si	7.68	7.86	6.97	6.91	7.23	6.6	5.79	5.57
Al	0.32	0.14	1.03	0.77	0.71	1.36	2.21	2.43
Fe^{3+}	0			0.32	0.06	0.04		
Σ_{tet}	8	8	8	8	8	8	8	8
Al	3.05	2.99	3.98	0.29			0.16	0.14
Ti							0.04	0.07
Fe^{3+}	0.42		0.04	3.67	0.03	0.48	0.13	1.14
Fe^{2+}	0	0.29				0.52		0.1
Mn							0.01	0.04
Mg	0.52	0.72	0.02	0.04	5.97	5	5.62	4.36
Σ_{oct}	4	4	4.04	4	6	6	5.96	5.85
Ca	0.02			0.025		0.44	0.93	0.65
Mg					0.35			
Na	0.8	0.85	0.91		0	0.01		
K	0	0.01	0.02	1.13	0.01	0.02	0.03	0.02
$\Sigma_{Inter.\ charge}$	0.84	0.86	0.93	1.13	0.71	0.91	1.89	1.32

1, Montmorillonite, Clay Spur, Wyoming: Total includes 0.02% P_2O_5; 2, montmorillonite, Camp Berteau, France, ignited weight basis (Weir, 1965); 3, beidellite, Black Jack Mine, Owyhee County, Idaho (Weir, 1965); 4, nontronite, Garfield, Washington (Besson et al., 1983); 5, saponite, Krugersdorp, Transvaal, South Africa (Schmidt and Heystek, 1953); 6, saponite, Kozakov, Czechoslovakia (Suquet et al., 1975); 7, vermiculite, Llano County, Texas; 8, vermiculite, Ypung River, Western Australia.

See Newman (1987) for original source of each analysis.

nontronite or ferruginous smectite. Although less frequent, chromium (Cr^{3+}) and vanadium (V^{3+}) also are found as dominant cations in the octahedral sheets of the beidellite structure; in the case of chromium, the mineral is named volkonskoite.

The ideal structural formula of trioctahedral ferromagnesian smectites, the series saponite through iron-saponite, is given by $M_x^+(Mg,Fe^{2+})_3(Si_{4-x}Al_x)O_{10}(OH)_2 \cdot nH_2O$. Tetrahedral substitution is responsible for the net-charge deficiency in the smectite minerals of this series. Besides magnesium and ferrous iron, zinc, cobalt, and manganese are known to be dominant cations in the octahedral sheet. Zinc-dominant species are called sauconite. In other types of trioctahedral smectites, the net-charge deficiency arises largely from ionic substitution or a small number of cation vacancies in the octahedral sheets, or from both. Ideally *x* is 0, but most often it is less than 0.15. Thus, the octahedral composition varies to maintain

similar amounts of net-charge deficiency as those of other smectites. Typical examples are $(Mg_{3-y}\square_y)$ for stevensite and $(Mg_{3-y}Li_y)$ for hectorite, respectively, where $\square$ denotes a vacancy site in the structure. In stevensite, therefore, *y* sites out of three are vacant.

The structure of smectites described above is based on their ideal model. Their actual structures are more or less distorted and show different physical appearances due to various circumstances. Electron optical diffraction examinations are most effective for their characterization. According to the accounts of Méring and his colleagues (Méring and Brindley, 1967; Méring and Oberlin, 1967, 1971; Méring, 1975), three types of layer-stacking arrangement have been observed and classified. Particles of smectites are composed of a stacking of elementary layers, in which on the average 10–20 elementary layers are stacked together, and their arrangements are ordered, semiordered, or turbostratic.

The ordered arrangement should give a clear hexagonal net (spots) pattern by the electron microdiffraction diagram. In semiordered arrangement, the mutual orientation of the elementary layers is defined by arbitrary multiples of 60°. Therefore, the stacking is no longer a triperiodic crystal but maintains "hexagonal" symmetry. If disorientation is defined by rotations about the c^*-axis (normal to the basal plane of smectites), layer stacking is called turbostratic. With the disorientation, diffraction spots of hk become diffuse and elongated, and the extent of those effects are related to the degree of disorientation. Turbostratic arrangements are unique to smectite minerals and no other phyllosilicates with such structure are known. Typically montmorillonite, hectorite, and nontronite show turbostratic structure. Beidellite and saponite are, on the other hand, known to possess relatively ordered structures.

21.2.5 2:1:1 Layer Type (T–O–T–O Type)

21.2.5.1 Chlorite Group

The structure of the chlorite minerals consists of alternate mica-like layers and brucite-like hydroxide sheets about 1.4 nm thick (Figure 21.22). Structural formulas of most trioctahedral chlorites may be expressed by four end-member compositions:

$$(Mg_5Al)(Si_3Al)O_{10}(OH)_8 \quad \text{(clinochlore)}$$
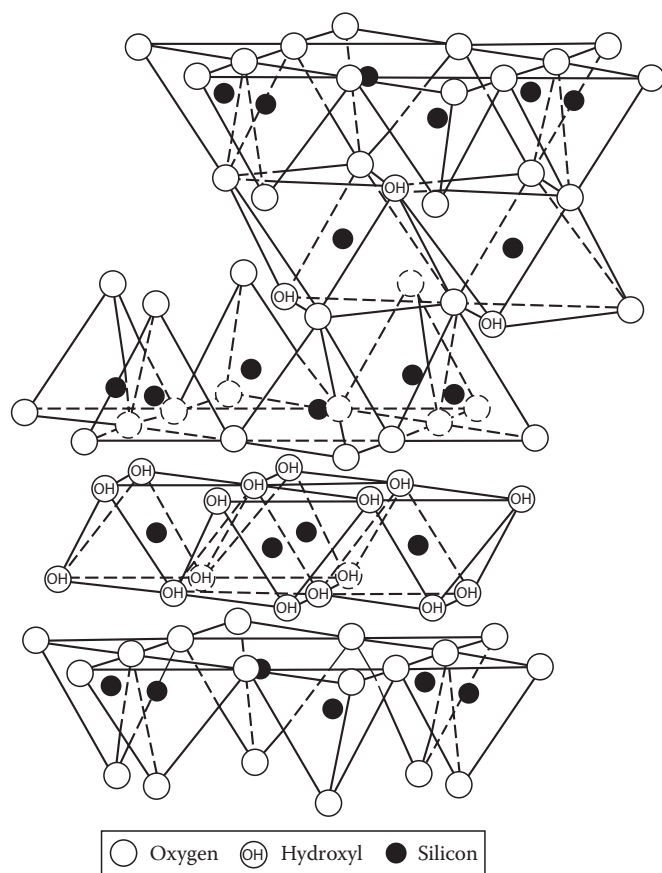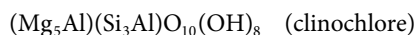
FIGURE 21.22 The structure of chlorite. (From Grim, R.E. 1968. Clay mineralogy. McGraw-Hill Book Co. Inc., New York. With permission.)

$$(Fe_5^{2+}Al)(Si_3Al)O_{10}(OH)_8 \quad \text{(chamosite)}$$

$$(Mn_5Al)(Si_3Al)O_{10}(OH)_8 \quad \text{(pennantite)}$$

$$(Ni_5Al)(Si_3Al)O_{10}(OH)_8 \quad \text{(nimite)}$$
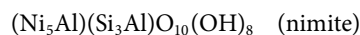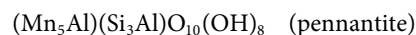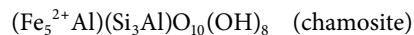
The unbalanced charge of the silicate layer is compensated by an excess charge on the hydroxide sheet caused by the substitution of trivalent cations (Al^{3+}, Fe^{3+}, etc.) for divalent cations (Mg^{2+}, Fe^{2+}, etc.). Chlorites, with an aluminous dioctahedral silicate layer and an aluminum hydroxide sheet are called donbassite and have the ideal formula of $Al_{4.33}(Si_3Al)O_{10}(OH)_8$ as an end-member for the dioctahedral chlorite. However, in many cases, the octahedral Al ions are partially replaced by magnesium, as in magnesium-rich aluminum dioctahedral chlorites, which are called sudoite. Cookeite is another type of dioctahedral chlorite, in which lithium (Li) substitutes for aluminum in the octahedral sheets. Selected chemical analyses are listed in Table 21.15. The symmetry of chlorite structures is variable due to the kind and amount of substitutions within the hydroxide sheet and within the tetrahedral and octahedral sites of the 2:1 silicate layer, the orientation of successive octahedral and tetrahedral sheets, and the way of the stacking of successive chlorite units. Specifically, as the chlorite unit contains hydroxide layers, the way of stacking between the basal oxygen atoms of the tetrahedral sheet and the hydroxyls in the first atomic plane of the hydroxide sheet is unique for chlorite. According to Shirozu and Bailey (1965), among six possible polytypes four are known to naturally occur and they are denoted as Ia, Ib ($\beta = 97°$), Ib ($\beta = 90°$), and IIb, respectively. Type I and Type II are distinguished by a difference in the shifting direction of the silicate layer with respect to that of the successive hydroxide sheet. Designation a indicates the case in which two-thirds of cations in the hydroxide sheet locate above tetrahedral cations and the remaining octahedral cations occupy at the center of the hexagonal rings, whereas designation b is the case in which cations of the hydroxide sheet distribute symmetrically upon projecting them onto the hexagonal ring. An overwhelmingly common polytype in nature is the single-layer monoclinic structure designated as IIb. A majority of trioctahedral chlorites including sudoite takes this polytype, while Ib ($\beta = 97°$) and Ib ($\beta = 90°$) polytypes are abundant in iron-rich chlorites. The Ia polytype is often realized in cookeite and donbassite. Chlorite structures are relatively thermally stable compared to kaolinite, vermiculite, and smectite minerals. Due to this, the presence of the XRD peak at 1.4 nm after heat treatment at 500°C–700°C is widely used for the identification of chlorite minerals.

21.2.6 Interstratified Layer Silicate Group (Mixed-Layer Type)

As seen from the foregoing discussion, structures of phyllosilicates are strikingly similar except for sheet sequence such as T–O and T–O–T and layer thickness. Because of similarity as such, nature finds that a layer silicate of one kind interstratifies with that of another kind to make up a new structure that differs from either component layer.

TABLE 21.15   Chlorite Analyses

	1	2	3	4	5	6	7	8
SiO_2	33.83	26.65	21.29	27.27	33	33.31	35.36	34.4
Al_2O_3	12.95	16.14	19.07	15.21	35.69	44.16	47.35	51.85
TiO_2						tr		
Fe_2O_3	2.25	6.69	6.67		2.74	1.72	0.52	
FeO	3.02	34.43		2.78	0.24	1.69	0.07	
MnO	0	0.07	39.82	0.06	0.28	0.18		
ZnO			0.97					
NiO				29.49				
CoO				0.38				
MgO	34.94	4.47	41.06	10.13	14.07	0.37	0.67	
Li_2O						3.4		
CaO				0.38		0	0.45	
K_2O					0	0.67		
Na_2O						0.7		
H_2O^+	13.11	11.42	12.11	10.48	13.83	13.19	14.28	13.75
H_2O^-	0	0.08		0.27		0.46		
Total	100.1	99.88	100.12	100.8	100.03	99.85	99.1	100
Numbers of cations on basis of $O_{10}(OH)_8$								
Si	3.21	3.05	2.59	3.01	3.01	2.975	3.155	3
Al	0.79	0.95	1.41	0.99	0.99	1.025	0.845	1
Fe^{3+}								
Σ_{tet}	4	4	4	4	4	4	4	4
Al	0.65	1.23	1.31	0.98	2.84	3.625	4.13	4.33
Fe^{3+}	0.16		0.6	0.36	0.19	0.115	0.035	
Fe^{2+}	0.24	3.87		0.26	0.02	0.125		
Mn				4	0.02	0.015		
Ni				2.62				
Co				0.04				
Zn			0.09					
Mg	4.94	0.76		1.68	1.91	0.05	0.09	
Li						1.22		
Σ_{oct}	5.99	5.86	6	5.96	4.98	5.15	4.26	4.33

1, Clinochlore, Zillertal, Austria; 2, chamosite, Schmiedefeld, Germany: Total Fe as Fe^{2+} in numbers of cation; 3, pennantite, Ushkatyn deposit, Kazakhstan; 4, nimite, Barberton, South Africa: Total octahedral cation number includes 0.04 Ca; 5, sudoite, (di, trioctahedral chlorite) Venn-Stavelot Massif, Ardennes, Belgium (Fransolet and Bourguignon, 1978); 6, cookeite, Northwestern USSR (Cerny, 1970); 7, donbassite, Novaya Zemlya, Russia (Drits and Lazarenko, 1967); 8, ideal donbassite composition, $Al_{4.33}(Si_3Al)O_{10}(OH)_8$.

Analyses 1–4 and 8 are taken from Anthony et al. (1995). Analyses 5–7 are taken from Newman (1987) and modified to fit with the table. For original data source, see Anthony et al. (1995) or Newman (1987).

The most striking examples of interstratified structures are those having a regular *AB AB* … type structure, where *A* and *B* represent two different component layers. The interstratified minerals of this type show a long periodicity along the *c**-axis (normal to the basal plane) of the structure, which is the sum of the two individual periodicities of the component layers *A* and *B*. Several minerals are known to have structures of this type, that is, rectorite (dioctahedral mica/montmorillonite), tosudite (dioctahedral chlorite/smectite), corrensite (trioctahedral vermiculite/chlorite) in two types with high- and low-layer charge vermiculite,

hydrobiotite (trioctahedral mica/vermiculite), aliettite (talc/saponite), and kulkeite (talc/chlorite). Analyses of selected samples of those regular interstratifications are shown in Table 21.16.

Other than the *AB AB* … type with equal numbers of the two-component layers in a structure, a variety of modes of the layer-stacking sequence is possible from nearly regular to completely random. The following interstratifications of two components are reported in these modes in addition to those given above: Illite–smectite, glauconite–smectite, dioctahedral mica–chlorite, dioctahedral mica–vermiculite, and kaolinite–smectite. As the mixing

TABLE 21.16 Interstratified Layer Silicates (Two Components with 1:1 Ratio and Regular Sequence)

	1	2	3	4	5	6	7
SiO_2	46.21	53.15	36.77	35.2	36.96	41.25	54.11
Al_2O_3	14.44	3.48	11.6	14	32.09	36.48	40.38
TiO_2				tr	0.34	0.07	0.01
Fe_2O_3		3.48	8.19	3.48	1.57	0.67	0.15
FeO			0.98	2.9	tr		
MnO		0.03	0.08	tr			0
MgO	37.85	27.4	20.04	28.5	8.2	1.27	0.78
CaO	0.07	1.1	1.94	0.93	2.21	3.97	0.52
K_2O	0.06		3.84	tr	0.23		3.87
Na_2O		1.18	0.12	tr	0.16		0.29
H_2O total		10.18					
H_2O^+			6.69	10.6	12.71	7.19	
H_2O^-			7.8	4.28	6.12	6.99	
Total	100.1	100	99.93	99.89	100.59	99.72	100.24
Anion basis	A	B	C	D	E	F	G
Si	13.14	14.66	5.77	6.062	12.5	5.85	12.84
Al	2.86	1.13	2.14	1.938	3.5	2.15	3.16
Fe^{3+}		0.21	0.09				
Σ_{tet}	16	16	8	8	16	8	16
Al	1.98			0.897	9.29	3.92	8.12
Ti					0.09	0.01	0
Fe^{3+}		0.51	0.88	0.45	0.4	0.07	0.03
Fe^{2+}			0.13	0.416			
Mn		0.01	0.01				0
Mg	16.05	11.26	4.68	7.362	4.13	0.15	0.09
Σ_{oct}	18.03	11.78	5.82	9.125	13.91	4.15	8.24
Total Ca	0.02	0.33	0.34[a]	0.172	0.8	0.59	0.14
Total K	0.02		0.77		0.1	0.15	0.09
Total Na	0.76	0.63	0.04		0.11	0.27	1.78
Σ	0.8	0.96	1.15	0.172	1.01		2.01
Exch. Ca							0.29
Exch. K							0.01
Exch. Na							0.13
Exch. Mg						0.16	
Σ_{charge}						0.32	0.72

Anion basis: A = E $[O_{40}(OH)_{20}]$; B = G $[O_{40}(OH)_8]$; C = F $[O_{20}(OH)_4]$; and D $[O_{20}(OH)_{10}]$.

1, Kulkeite, chlorite-talc, Derrag, Tell Atlas, Algeria (Schreyer et al., 1982); 2, aliettite, talc-saponite, Nure Valley, Piacenza Province, Italy (Alietti and Mejsner, 1980); 3, hydrobiotite, biotite–vermiculite, Rainy Creek, near Libby, Montana (Boettcher, 1966): Total includes 0.27% Cr_2O_3, 0.01% NiO, 0.01% SrO, 0.19% BaO, 0.0158% Rb_2O, 0.0005% Cs_2O and 0.06% P_2O_5; 4, corrensite, trioctahedral chlorite–vermiculite (or smectite) Sharbot Lake, Ontario, Canada (De Kimpe et al., 1987); 5, tosudite, chlorite–smectite, Niida, Akita Prefecture, Japan (Sudo and Shimoda, 1978); 6, rectorite, dioctahedral mica (Ca)–smectite, Sano mine, Nagano Prefecture, Japan (Matsuda et al., 1997); 7, rectorite, dioctahedral mica (Na)–smectite, Fort Sandeman, Baluchistan, Pakistan (Kodama, 1966): Total includes exchangeable Ca, Mg, and Sr.

See Newman (1987) for original data source of each analysis.

[a] Ba is included.

ratio (proportion of the component layers) of the two-component layers varies, the number of possible layer-stacking modes increases greatly. Theoretically, any number of the kinds of component layers can form interstratified structures. In practice, however, a maximum of three-component layers have been positively identified. Interstratified structures consisting of illite–chlorite–smectite and illite–vermiculite–smectite have been reported so far. Because certain interstratified structures are known to be stable under relatively limited conditions, their occurrence may be used as a geothermometer or other geoindicator. On the other hand, Nadeau (1985) and Nadeau et al. (1984a, 1984b) proposed the fundamental particle theory, which argues that interstratified structures may be considered as artifacts of sample preparation. Extra caution is required to describe interstratified layer silicates correctly.

21.2.7 Modulated Layer Silicate Group

21.2.7.1 General

Guggenheim and Eggleton (1988) defined modulated layer silicates as those minerals that contain a periodic perturbation to the basic layer silicate structure. The basic structure involves 1:1 or 2:1 layer configurations. Thus, as the severity of the modulations increases, 2D continuous layer-like qualities diminish. The octahedral sheet of a modulated layer silicate is invariably a brucite-like (trioctahedral) sheet. Palygorskite, sepiolite, greenalite, minnesotaite, caryopilite, zussmanite, stilpnomelane, ganophyllite, bementite, etc. are known to be mineral species belonging to this group. Strictly speaking, antigorite may also be treated as one of the modulated 1:1 layer silicates. In this chapter the "phyllosilicate" antigorite is thought to be represented in comparison with its other counterparts.

21.2.7.2 Sepiolite–Palygorskite Group

Sepiolite and palygorskite are papyraceous or fibrous hydrated magnesium silicate minerals (Figure 21.23) and are included in

FIGURE 21.23 Electron micrograph of palygorskite from Nawton, near Thirsk, Yorkshire. (From Gard, J.A. (ed.). 1971. The electron-optical investigation of clays. Monograph No 3. Mineralogical Society, London, U.K. Reproduced with kind permission of The Mineralogical Society of Great Britain and Ireland.)
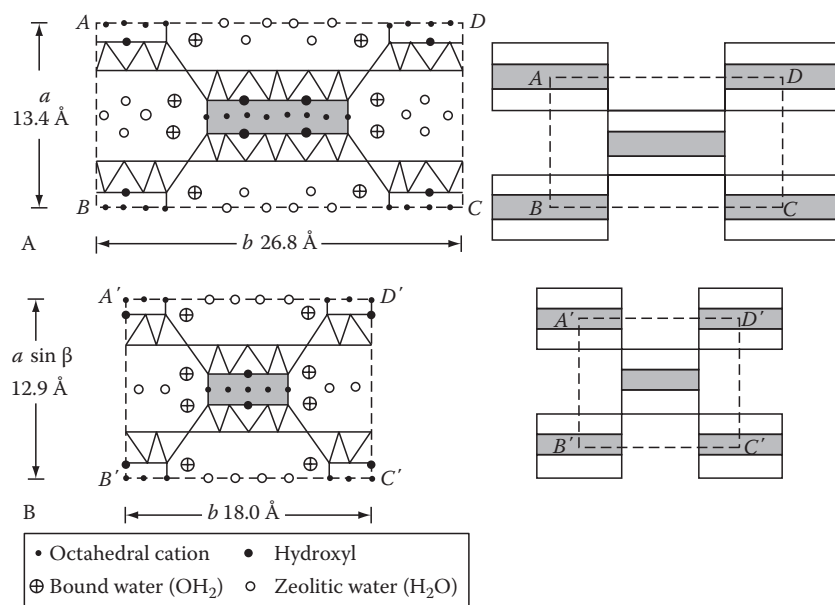
FIGURE 21.24 Structural sketches of palygorskite and sepiolite. (Reprinted from Sudo, T., S. Shimoda, H. Yotsumoto, and S. Aita. 1981. Electron micrographs of clay minerals. Developments in sedimentology. Vol. 31. Elsevier Scientific Publishing Co., New York. With permission from Elsevier.)

the phyllosilicates because they contain a continuous 2D tetrahedral sheet of composition Si_2O_5. They differ, however, from the other layer silicates because they lack continuous octahedral sheets. The structures of sepiolite and palygorskite are alike and can be regarded as consisting of narrow strips or ribbons of 2:1 layers that are linked stepwise at the corners (Figure 21.24). One ribbon is linked to the next by inversion of the direction of the apical oxygens of SiO_4 tetrahedra, in other words, an elongated rectangular box consisting of a continuous 2:1 layer is attached by the nearest boxes at their elongated corner edges. Therefore, channels or tunnels due to the absence of the silicate layers occur on the elongated sides of the boxes. The elongation of the structural element is related to the fibrous morphology of the minerals and parallel to the *a*-axis. Since the octahedral sheet is discontinuous, some octahedral magnesium ions are exposed at the edges and hold bound water molecules (OH_2). In addition to the bound water, variable amounts of zeolitic (i.e., free) water (H_2O) are contained in the rectangular channels. The major difference between the structures of sepiolite and palygorskite is the width of the ribbons, which is greater in sepiolite than in palygorskite. The width determines the number of octahedral cation positions per formula unit. Thus, sepiolite and palygorskite have the ideal compositions $Mg_8Si_{12}O_{30}(OH)_4(OH_2)_4(H_2O)_8$ and $(Mg,Al,\square)_5Si_8O_{20}(OH)_2(OH_2)_4(H_2O)_4$, respectively. Some of Mg in sepiolite may be replaced by Al, Fe^{3+}, or Fe^{2+}. Selected analyses of this mineral group are given in Table 21.17.

21.2.7.3 Greenalite, Minnesotaite, and Caryopilite

Greenalite was once expected to be a near iron end-member of serpentine. Guggenheim et al. (1982) showed that its structure is composed of coherent intergrowths of a trigonal polytype and a monoclinic polytype, with volume wise the former being larger than the latter and having a domain size of about 2 nm. Minnesotaite has a continuous octahedral sheet (Guggenheim and Bailey, 1982). Adjacent tetrahedral are present on either side of this sheet to form an approximate 2:1 layer. However, in contrast to talc, strips of linked hexagonal rings of tetrahedra are formed only parallel to *y* (Figure 21.25; Guggenheim and Eggleton, 1986). Some analyses of greenalite and minnesotaite are listed in Table 21.18, in which an analysis of caryopilite, as a mangan species of serpentine-like mineral, is included for comparison.

21.3 Occurrence of Phyllosilicates

21.3.1 As Rock-Forming Minerals

Known established species of phyllosilicates are no more than 100 (see Appendix). Although this is a rather small portion of the total mineral kingdom of some 3700 species, the occurrence of phyllosilicates is spread to a wide variety of geological environments. Among those described in the previous sections, perhaps biotite is the most; it is found in igneous rocks varying from granitic pegmatites, to granites, to syenites, to diorites, and to gabbros and peridotites. It also occurs in felsic volcanic rocks such as dacites, rhyolites, trachyte, and phonolites. In metamorphic rocks, including contact-metamorphosed rocks, biotite is formed under a wide range of temperature and pressure conditions. Muscovite is also a widespread and common rock-forming mineral, especially in pegmatites and granitic rocks. It is often found in metamorphic rocks as mica schists and as green schists with chlorite. Phlogopite occurs in metamorphosed magnesium limestones, dolomites, and ultrabasic rocks. Lepidolite is an uncommon mineral, found in lithium-rich pegmatites. Some brittle micas occur in regional contact-metamorphosed rocks.

TABLE 21.17 Palygorskite and Sepiolite Analyses

	1	2	3	4
SiO_2	55.03	53.75	52.5	52.36
Al_2O_3	10.24	10.23	0.6	0.23
Fe_2O_3	3.53	1.83	2.99	0.4
FeO		0.26	0.7	
MnO				0.01
MgO	10.49	9.39	21.31	22.6
CaO		2.29	0.47	0.83
K_2O	0.47	0.02		
Na_2O		tr		
H_2O^+	10.13	12.04	9.21	9.34
H_2O^-	9.73	10.16	12.06	13.85
Total	99.62	99.97	99.84	99.62
Numbers of cations on basis of				
	A	B		
Si	7.8	7.82	11.81	11.95
Al	0.2	0.18	0.16	0.05
Fe^{3+}			0.03	
Σ_{tet}	8	8	12	12
Al	1.51	1.57		0.01
Fe^{3+}	0.38	0.2	0.47	0.07
Fe^{2+}		0.03	0.13	
Mg	2.22	2.04	7.14	7.69
Ni				
Σ_{oct}	4.11	3.84	7.74	7.77
Ca		0.36	0.11	0.2
K	0.09			
Na				
Σ charge	0.09	0.72	0.22	0.4

A: $O_{20}(OH)_4(OH_2)_2$ for anion basis of palygorskite; B: $O_{30}(OH)_4(OH_2)_4$ for anion basis of sepiolite.

1, Palygorskite, Attapulgus, Georgia (Bradley, 1940); 2, palygorskite, Kuzu District, Tochigi Prefecture, Japan (Imai et al., 1969); 3, sepiolite, Ampandrandava, Maagascar (Caillère and Hénin, 1961); 4, sepiolite, Kuzu District, Tochigi Prefecture, Japan (Takahashi, 1966).

Analyses are taken from Newman (1987).

21.3.2 Hydrothermal Origin and the Term "Clay Minerals"

In hydrothermal deposits or wall-rock alteration zones, many of phyllosilicates except palygorskite and sepiolite have been found as alteration products associated with hot springs and geysers and as aureoles around metalliferous deposits. Frequently a zonal arrangement of phyllosilicates, normally as very fine-grained clayey materials, is observed around the source of the alteration. The zonal arrangement varies with the type of parent rocks and the nature of hydrothermal solution. Extensive clay zones formed in close association with "kuroko" deposits contain mica (illite), chlorite, tosudite, smectite, and mica–smectite interstratifications. Pottery stones that consist of kaolinite, illite, and pyrophyllite occur as alteration products of acidic volcanic
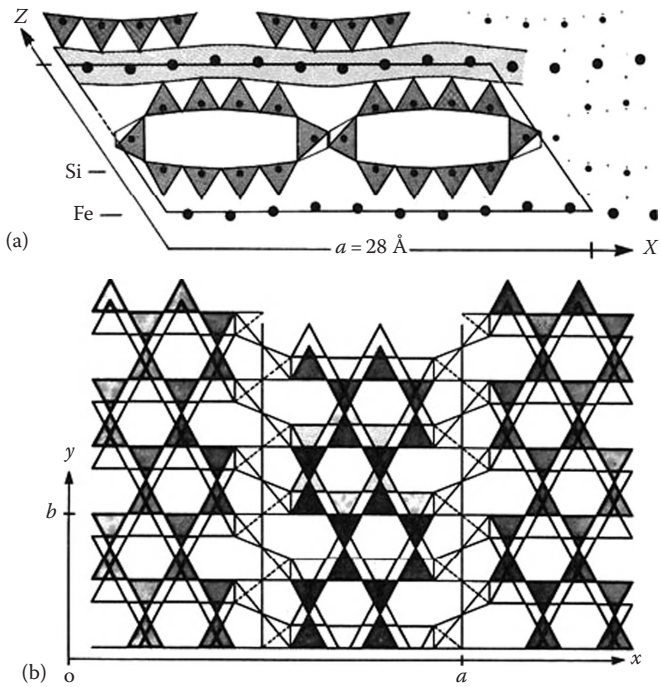
FIGURE 21.25 Structural sketches of minnesotaite: (a) Perspective view of the idealized structure in which the interlayer tetrahedra share corners to form a chain parallel to Y and (b) in the (001) plane drawing. (From Guggenheim, S., and S.W. Bailey. 1982. The superlattice of minnesotaite. Can. Mineral. 20:579–584.)

rocks, shales, and mudstone, while talc is formed from magnesium silicates, such as olivines, pyroxenes, and amphiboles. Generally speaking, near-neutral hydrothermal solutions generate rock alteration, including the formation of illite, chlorite, and smectite, whereas acid hydrothermal solutions generate the formation of kaolinite.

Most phyllosilicates of hydrothermal origin hold properties defined by the term "clay." The term "clay" is generally applied to the following: (1) a natural material with plastic properties, (2) an essential composition of particles found in a very fine-size fraction, which is customarily defined as particles smaller than 2 μm, and (3) a composition of very fine mineral fragments or particles that are mostly hydrous layer silicates of aluminum or occasionally containing magnesium and iron (Grim and Kodama, 1997). Although, in a broader sense, clay minerals can include virtually any minerals in the clay-size fraction, the definition adopted here is restricted to the hydrous layer silicates, which are conveniently used to distinguish themselves from macroscopic phyllosilicates. In other words, the term "clay minerals" is a synonym of microcrystalline phyllosilicates.

21.3.3 Diagenesis and Low-Grade Metamorphism

Analyses of many ancient sediments in many parts of the world indicate that very old argillaceous (clay-rich) sediments (physilites) are largely composed of illite and chlorite. These clay minerals are also found in carbonate rocks (calcretes). Smectite is a

TABLE 21.18 Modulated Layer Silicate Analyses

	1	2	3	4
SiO_2	33.58	36.5	35.4	51.29
Al_2O_3		0.16	2.38	0.61
TiO_2				0.04
Fe_2O_3	11.16	nr	nr	2
FeO	45.19			33.66
Total Fe as FeO[a]		46.7	0.95	nr
MnO		1.5	49.2	0.12
MgO		4.53	1.23	6.26
CaO		0.02	0.16	0
K_2O			0.03	0.03
Na_2O			0.16	0.08
H_2O total	10.07			
H_2O^+				5.54
H_2O^-				0.24
Total	[100]	89.41	89.4	99.87
Anion basis	$O_5(OH)_{3.28}$	$O_{10}(OH)_8$	$O_{10}(OH)_8$	
Si	2.18	4	4	3.95
Al				0.05
Fe^{3+}				
Σ_{tet}	2.18	4	4	4
Al		0.02	0.32	
Fe^{3+}	0.45			
Fe^{2+}	2.34	4.28	0.09	2.28
Mn		0.14	4.7	0.01
Mg		0.74	0.21	0.72
Σ_{oct}	2.79	5.18	5.32	3.01

1: Greenalite, Mesabi Range, Minnesota. Recalculated to 100% after reduction of SiO_2 and other impurities. Anion basis: $O_5(OH)_{3.28}$; 2: greenalite, Sokoman iron formation, Knob Lake, Labrador, Canada; 3: caryopilite, Hurricane Claim, Olympic Peninsula, Washington; 4: Minnesotaite, Mesabi district, Minnesota.

Analyses 1 and 4 are taken from Anthony et al. (1995). Analyses 2 and 3 are taken from Newman (1987).

For original data source of each analysis, see Anthony et al. (1995) or Newman (1987).

common component of many shales of Mesozoic and younger ages. As temperature and pressure increase with the progression of diagenesis, clay minerals in sediments under these circumstances change to those stable under given conditions. Therefore, certain sensitive minerals may serve as indicators for various stages of diagenesis. Typical examples are the crystallinity of illite (Kubler, 1966), illite and chlorite polytypes, and the conversion of smectite to illite. The data reported indicate that smectite was transformed into illite through interstratified illite–smectite mineral phases as diagenetic processes advanced. Much detailed work has been devoted to the conversion of smectite to illite in the lower Cenozoic–Mesozoic sediments because such a conversion appears to be closely related to oil-producing processes (Burst, 1969). In the very low-grade metamorphic zone, which is considered to be an intermediate zone between diagenetic and metamorphic zones, the occurrence of rectorite, kulkeite, pyrophyllite, and talc has been reported. Well-crystallized micas and

chlorites are often major components in metamorphic rocks. Smectites are known to occur in sediments of pyroclastic materials as the result of devitrification of volcanic ash in situ. Much discussion appears to favor marine origin for glauconite and nonmarine origin for celadonite (Dunoyer de Sagonzac, 1970; Weaver, 1989; Vielde, 1992).

21.3.4 In Soils, Sediments, and Weathering Products

The formation of the clay minerals by weathering processes is determined by the nature of the parent rock, climate, topography, vegetation, and the time period during which these factors have operated. Climate, topography, and vegetation influence weathering processes by their control of the character and direction of movement of water through the weathering zone. When the dominant movement of water is downward through the alteration zone, any alkaline or alkaline-earth elements tend to be leached and primary minerals containing these components are first degraded and then broken down. If the leaching is intense, then after the removal of the alkalis and alkaline earths, the aluminum or silica may be removed from the alteration zone. This will depend on the pH of the downward-seeping waters. The pH of such water is determined, in turn, by the climate and cover of vegetation. Under warm and humid conditions, with long wet and dry periods, the surface organic material tends to be completely oxidized. The downward-seeping waters, therefore, are neutral or perhaps slightly alkaline and silica will be removed, whereas aluminum and iron will be left behind and concentrated. The result is a lateritic type of soil. Under more temperate conditions, the surface organic material is not completely oxidized and the downward-seeping water contains organic acids. In this case, aluminum and iron oxides are leached and the silicon is left behind; podzolic types of soils will develop. Under these conditions, mica and chlorite tend to be transformed into expandable clay minerals such as vermiculite and smectite or their intergrades as interstratifications. Allophane and imogolite may be present as newly formed clay minerals. In andosols, which are the soils developed on volcanic ash, allophane and imogolite as well as hydrated halloysite and halloysite are dominant components. (Wada and Greenland, 1970) Smectite is usually the sole dominant component in vertisols, which are clayey soils. Smectite and illite, with occasional small amounts of kaolinite, occur in mollisols and prairie chernozemic soils. Illite, vermiculite, smectite, chlorite, and interstratified clay minerals occur in podzolic soils. The alkaline and alkaline-earth elements remain close to the surface and the dry grassland soils (chernozem) containing illite, chlorite, and smectite, will develop. In dry areas, the dominant movement of water is not downward and leaching does not take place. In extreme dry desert areas (desert soils, some aridsols), where the concentration of magnesium is particularly high, the formation of palygorskite–sepiolite minerals has been reported. Kaolinite is the dominant component in laterite soils (oxisols). Clay minerals other than those mentioned above usually occur in various soils as minor components inherited from the parent materials of those soils (cf. Kittrick, 1985).

In general, the process of transportation and sedimentation little affects the weathering products. However, in lagoon–estuary areas, some mineral transformations are possible because environments change from nonmarine to marine and vice versa. As burial sedimentation and compaction continues, temperature and pressure increase. Weathering products in sediments change accordingly. The weathering products are transformed into other clay minerals or decomposed completely to provide constituents required for the neoformation of clay minerals, depending on the chemical environments of the sedimentation and the types of clay minerals present in sediments prior to diagenesis.

In soils, we observe the alteration of phyllosilicates by weathering action, as in a notable case like the transformation of micas into vermiculite, smectite, or randomly interstratified layer silicates (mica–vermiculite, mica–smectite, mica–vermiculite–smectite). On the other hand, we also notice the formation of chloritic minerals from vermiculite or montmorillonite by precipitating hydroxide sheets between the silicate layers, is also noticed as a results of accretion. Other observations have supported some systematic weathering patterns: Kaolinite is formed from aluminum silicates, particularly feldspar. Chlorite is found in igneous rocks as an alteration product from Mg–Fe silicates such as olivines, pyroxenes, amphiboles, and biotite micas, and serpentine is also found in the manner similar to chlorite. In meteorites, the occurrence of iron-rich serpentines in carbonaceous chondrites is reported.

Weathering and subsequent transportation result in the accumulation of sediments at the bottom of lakes, rivers, and oceans. In the Mississippi River system, for example, smectite, illite, and kaolinite are the major components in the upper Mississippi and Arkansas Rivers, whereas chlorite, kaolinite, and illite are the major components in the Ohio and Tennessee Rivers. This is the case for nonmarine conditions. Under marine conditions, at the Gulf of Mexico, for example, smectite, illite, and kaolinite are found to be the major clay mineral composition and their compositions vary from place to place. In some limited regions, it is noteworthy that these compositions are significantly altered by other factors such as airborne effects. The high-kaolinite concentration off the west coast of Africa near the equator reflects this effect.

In a supergene enrichment process, the formation of nickel deposits at New Caledonia is well known, where Ni–serpentine minerals occur in serpentine and peridotite rocks.

21.4 Phyllosilicates in Soil Environments

21.4.1 Introduction

In the foregoing section in which phyllosilicates of hydrothermal origin were discussed, the term "clay minerals" was introduced and the term can be the synonym of phyllosilicates of clay size. The number of phyllosilicate species that occur either in clay size or in a wide range of particle sizes from macroscopic to microscopic is rather limited. Although the numbers of species are small, their quantities are huge. Hence, clay components have great impact on soil environments. The phyllosilicates commonly found in

soils and which belong to the category above include hydrated halloysite, halloysite, kaolinite, micas, chlorites, vermiculites, smectites, and their interstratified minerals. Allophane and imogolite are often associated. The occurrence of clay minerals other than these in soils is either uncommon or rare. Because of fine particles, until relatively recently no appropriate analytical techniques were available by which to determine the precise nature, composition, and structure of clays. Therefore, clays had long been believed to be noncrystalline "colloidal" substances before XRD techniques developed in the 1920s, followed a few years later by improved microscopic and thermal procedures, which established that clays are composed of a few groups of crystalline minerals (Bradley and Grim, 1948). The introduction of electron microscopic methods was very useful to determine characteristic shape and size of clay minerals. Relatively modern analytical techniques such as infrared absorption, neutron diffraction, Mössbauer spectroscopy, and nuclear magnetic resonance have been applied to advance our knowledge of crystal chemistry of clay minerals (Greenland and Hayes, 1978; Wilson, 1992).

21.4.2 Physical Properties

21.4.2.1 Optical Index, Specific Density, and Particle Shape and Size

Clay mineral particles are commonly too small for the measurement of optical properties. Oriented aggregates that are large enough for optical measurements can, however, be prepared by allowing the flake-shaped clay mineral particles to settle from a clay water suspension on a horizontal surface. The particles settle with one flake on top of another so that their basal plane surfaces are essentially parallel. Refractive indices of clay minerals generally fall within a relatively narrow range from 1.47 to 1.68. This range may be subdivided into several groups represented by major mineral species: 1.47–1.52 (allophane, hydrated halloysite); 1.53–1.58 (kaolinite minerals, vermiculite); 1.58–1.65 (nontronite, glauconite, celadonite); and 1.57–1.68 (trioctahedral chlorites). All remaining clay minerals have refractive indices ranging from 1.54 to 1.63, except that smectite minerals have a range from 1.48 to 1.61. In general, iron-rich species show high refractive indices, whereas water-rich porous mineral species have lower refractive indices. Specific densities are as follows: 2.5–2.7 $g \cdot cm^{-3}$ (kaolinite-group minerals); 2.5–2.9 (pyrophyllite, talc, clay micas); 2.6–3.3 (chlorite minerals); 2.6–3.0 (vermiculite); 2.5–2.8 (smectite); ~2 (palygorskite–sepiolite group); 1.7–2.4 (bulk); and 2.6–2.8 (particle) for imogolite and allophane. Table 21.19 lists density, crystallographic units, and optical data of major clay minerals.

The size and shape of clay minerals have been determined by electron micrographs. Well-crystallized kaolinite occurs as well-formed six-sided flakes (Figure 21.26), frequently with a prominent elongation in one direction. Particles with maximum surface dimensions of 0.3 to about 4 μm and thickness of 0.05–2 μm are common. The flakes of disordered kaolinite have poorly developed hexagonal outlines. Halloysite commonly occurs as tubular units with an outside diameter ranging from 0.04 to 0.15 μm (Figure 21.27). Electron micrographs of smectite frequently show

TABLE 21.19 Crystallographic, Specific Gravity, and Optical Data of Some Selected Phyllosilicates

	D (g mL⁻³)	Crystallographic Data									Optical Data			
		a	*b*	*c*	α	β	γ	Z			α	β	γ	2V
Halloysite	2.55–2.265	0.514	0.89	1.49		101.9°		[4]						nd
Kaolinite	2.61–2.68	0.515	0.895	0.739	91.8°	104.5°–105°	90°	[2]	Biaxial (−)		1.553–1.566	1.559–1.569	1.560–1.570	23°–50°
Dickite	2.6	0.515	0.894	1.4424		96°44′		4	Biaxial (+)		1.560–1.561	1.561–1.563	1.566–1.567	52°–80°
Nacrite	2.5–2.7	0.8909	0.5146	1.5697		113°42′		4	Biaxial (−)		1.557–1.560	1.562–1.563	1.563–1.566	40°–90°
Lizardite	2.55	0.5325		0.7259				2	Uniaxial (−)		1.538–1.554	1.546–1.560	1.546–1.560	Small
Chrysotile	≈2.55	0.535	0.925	0.733		94.2°					1.532–1.544		1.545–1.553	Small
Antigorite	2.65	4.353	0.9259	0.7263		91°8.4′		16	Biaxial (−)		1.558–1.567	1.565	1.562–1.574	37°–61°
Talc	2.58–2.83	0.5287	0.9158	1.895		99.30°		4	Biaxial (−)		1.539–1.55	1.589–1.594	1.589–1.60	0°–30°
Pyrophyllite	2.65–2.9	0.516	0.8966	0.9347	91.18°	100.46°	89.64°	2	Biaxial (−)		1.534–1.556	1.586–1.589	1.596–1.601	53°–62°
Muscovite	2.77–2.88	0.519	0.904	2.008		95°30′		4	Biaxial (−)		1.552–1.576	1.582–1.615	1.587–1.618	30°–47°
Glauconite	2.4–2.95	[0.527]	[0.914]	[1.009]		≈100°		2	Biaxial (−)		1.592–1.610	1.614–1.641	1.614–1.641	0°–20°
Celadonite	2.95–3.05	0.523	0.906	1.013		100°55′		2	Biaxial (−)		1.606–1.625	nd	1.579–1.661	5°–8°
Phlogopite	2.78–2.85	0.53078	0.91901	1.01547		100.08°		2	Biaxial (−)		1.530–1.590	1.557–1.637	1.558–1.637	0°–15°
Biotite	2.7–3.3	0.53	0.92	1.02		100°		2	Biaxial (−)		1.565–1.625	1.605–1.696	1.605–1.696	0°–25°
Clinochlore	2.60–3.02	0.535	0.9267	1.427		96.35°		2	Biaxial (±)		1.571–1.588	1.571–1.588	1.576–1.597	0°–50°
Chamosite	3.0–3.4	0.5373	0.9306	1.4222		97°53′		2	Biaxial (−)		1.595–1.671	1.599–1.684	1.599–1.685	0°–30°
Donbassite	2.63	0.5174	0.8956	1.426		97.83°		[2]	Biaxial (+)		1.728	1.729	1.735	52°
Vermiculite	2.2–2.6	0.524	0.917	2.86		94°36′		4	Biaxial (−)		1.520–1.564	1.530–1.583	1.530–1.583	0°–15°
Montmorillonite	2–3	0.517	0.894	0.995		nd		1	Biaxial (−)		1.492–1.503	1.513–1.534	1.513–1.534	10°–25°
Beidellite	2–3	0.5179	0.897	1.757		≈90°		nd	Biaxial (−)		1.494	1.536	1.536	
Nontronite	2.2–2.3	[0.525]	[0.91]	[1.53]		≈90°		nd	Biaxial (−)		1.567–1.600	1.604–1.632	1.605–1.643	25°–68°
Saponite	2.24–2.30	0.53	0.914	1.69		≈97°		nd	Biaxial (−)		1.48–1.54	1.50–1.58	1.50–1.58	0°–40°
Greenalite	2.85–3.15	0.554	0.955	0.744		104°20′		2						
Minnesotaite	3.01	0.5623	0.9419	0.9624	85.21°	95.64°	90.00°	4	Biaxial		1.580–1.592	nd	1.615–1.632	Small
Palygorskite	>1.0–2.6	1.278	1.786	0.524		95.78°		4	Biaxial (−)		1.522–1.528	1.530–1.546	1.553–1.548	30°–61°
Sepiolite	>2	0.521	2.673	1.35				4	Biaxial (−)		1.515–1.520	nd	1.525–1.529	0°–50°

Source: Adapted from Yong, R.N., and B.P. Warkentin. 1975. Soil properties and behaviour. Developments in geotechnical engineering 5. Elsevier Scientific Publishing Co., Amsterdam, the Netherlands.

Unit cell: *a*, *b*, and *c* are in nanometers. *Z* is the number of formula units per unit cell and [*Z*] is estimated one. Compilation by Kodama.
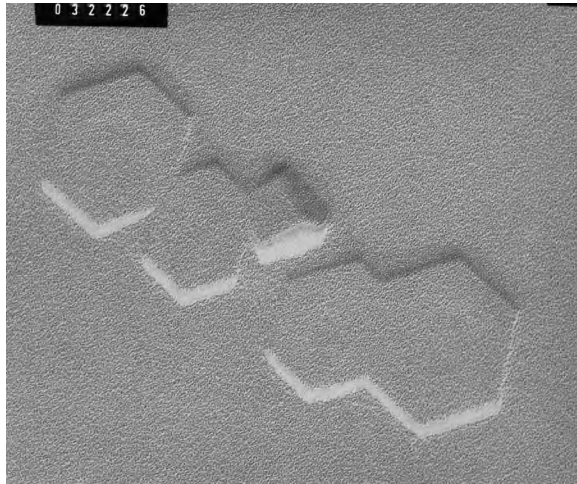
FIGURE 21.26 Electron micrograph of kaolinite, from Mont Megantic, Quebec. Quick freeze–dry and Pt shadowing. The longer side edge of the photograph equals about 2 μm.
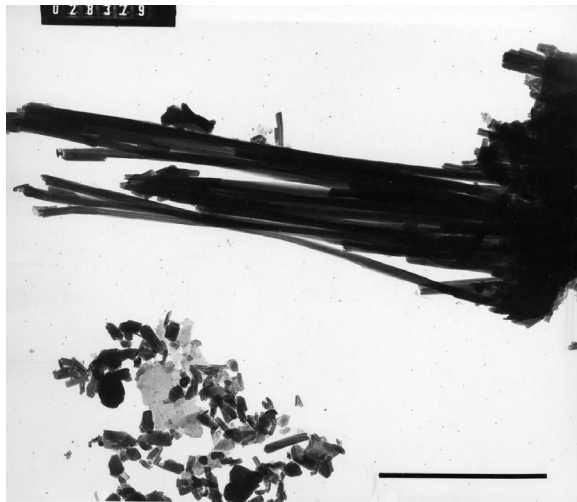
FIGURE 21.27 Electron micrograph of halloysite, in a paleosol from Gaspe, Quebec. Pt shadowing. The scale indicates 1 μm. (From Ross, G.J., H. Kodama, C. Wang, J.T. Gray, and L.B. Lafreniere. 1983. Halloysite from a strongly weathered soil at Mont Jacques Cartier, Quebec. Soil Sci. Soc. Am. J. 47:327–332.)

broad undulating mosaic sheets. In some cases, the flake-shaped units are discernible but frequently they are too small or thin to be seen individually without special attention (Oberlin and Méring, 1962). Illite occurs in poorly defined flakes commonly grouped together in irregular aggregates (Figure 21.28). Many of the flakes have a diameter 0.1–0.5 μm and the thinnest flakes are approximately 3 nm thick. Although sizes vary more widely, vermiculite, chlorite, pyrophyllite, talc, and serpentine minerals except for chrysotile are similar in character to the illites. As shown in Figure 21.11, chrysotile occurs in slender tube-shaped fibers with an outer diameter of 10–30 nm. Their lengths commonly reach several microns. Electron micrographs show that palygorskite occurs as elongated laths, singly or in bundles.
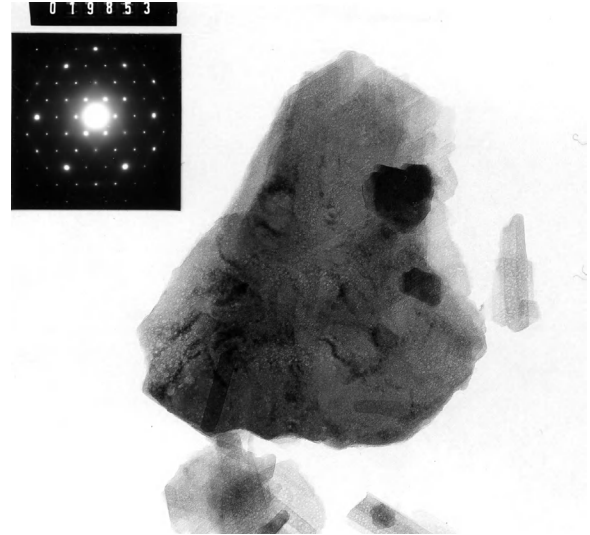
FIGURE 21.28 Electron micrograph of illite, from Eldorado, Saskatchewan. Although outline is irregular, it appears as a single crystal (about 1–1.5 μm across). (From Kodama, H., and R.S. Dean. 1980. Illite from Eldorado, Saskatchewan. Can. Mineral. 18:109–118.)

Frequently, the individual laths are many microns in length and 5–10 nm in width. Sepiolite occurs in similar lath-shaped units, but the laths are somewhat thicker and shorter than the palygorskite. As mentioned earlier in the structures, allophane occurs in very small spherical particles (3–5 nm in diameter), individually or in aggregated forms, whereas imogolite occurs in long thread-like tubes of several microns in length. One of the interesting features of layer silicate minerals is their modes of particle association. When a clay suspension of flaky particles flocculates, three different modes of particle association may occur: face-to-face, edge-to-face, and edge-to-edge. These modes are determined by the electrostatical interaction energy between particles, which is influenced by the kind of salt and salt concentration in the medium as well as the surface conditions of the particles.

21.4.2.2 Clay–Water Relations

Clay minerals contain water in several forms. The water may be held in pores and may be removed by drying under ambient conditions. Water also may be adsorbed on the surface of clay mineral structures and in smectites, vermiculites, hydrated halloysite, sepiolite, and palygorskite; this water may occur in interlayer positions or within structural channels. Finally, the clay mineral structures contain hydroxyls that are lost as water at elevated temperatures. The water adsorbed between layers or in structural channels may further be divided into zeolitic (free) and bound waters. The latter is bound to exchangeable cations or directly to the clay mineral surfaces. Zeolitic and bound waters may be removed by heating to temperatures on the order of 100°C–200°C and in most cases, except for hydrated halloysite, is regained readily at ordinary temperatures. Exceptions include those minerals like K-saturated vermiculite in which the interlayer K atoms do not associate with zeolitic water. The bound water is generally agreed to have a structure that is different from bulk water, being more ice-like in its

TABLE 21.20 Cation-Exchange Capacity and Specific Surface Area Data of Representative Clay Minerals, Including Imogolite and Allophane

	Cation-Exchange Capacity at pH 7 (cmol$_c$ kg^{-1})	Specific Surface Area (m^2 g^{-1})
Kaolinite	3–5	5–40
Hydrated halloysite	40–50	~1100
Illite	10–40	10–100
Chlorite	10–40	10–55
Vermiculite	100–150	~760
Smectite	80–120	40–800
Palygorskite–sepiolite	3–20	40–180
Imogolite	20–30	≈1540
Allophane	30–135	~2200

TABLE 21.21 Consistency Data of Representative Clay Minerals Including Allophane

	Plastic Limit	Liquid Limit	Plastic Index (PI)
Kaolinite–Na	26	52	26
Kaolinite–Ca	26	73	37
Illite–Na	34	61	27
Illite–Ca	40	90	50
Montmorillonite–Na	97	700	603
Montmorillonite–Ca	63	177	114
Palygorskite–sepiolite	145	171	26
Allophane (undried)	136	231	95
Allophane (dried)	78	85	7

Source: Adapted from Yong, R.N., and B.P. Warkentin. 1975. Soil properties and behaviour. Developments in geotechnical engineering 5. Elsevier Scientific Publishing Co., Amsterdam, the Netherlands.

PI = liquid limit − plastic limit. Numerical figures are expressed in dag kg^{-1}.

structure, even though still liquid. As the thickness of the adsorbed water increases outward from the surface, the nature of the bound liquid water changes abruptly or gradually from ice-like to that of bulk water. Ions and molecules adsorbed on the clay mineral surface exert a major influence on the thickness of the adsorbed water layers and on the nature of this water. The bound liquid water with ice-like characteristics may extend out from the clay mineral surfaces as much as 6–10 nm. Hydroxyl ions are driven off by heating clay minerals to temperatures of 400°C–700°C. The rate of loss of the hydroxyls and the energy required for their removal are specific properties characteristic of the various clay minerals. The reaction for dioctahedral minerals such as kaolinite is abrupt, whereas the loss takes place rather gradually in trioctahedral minerals. This dehydroxylation process results in the oxidation of Fe^{2+} to Fe^{3+} in ferrous-iron-bearing clay minerals. The amount of the adsorbed water is closely related to the surface area of clay minerals that is, of course, determined by their particle size and shape as well as the type of clay minerals. The range of measured surface areas of some clay minerals is given along with that of cation-exchange capacities in Table 21.20. The capacity of water retention of clay minerals is generally proportional to the extent of their surface area (e.g., Ross, 1978). As the water content increases, clays become plastic and then change to a near-liquid state. The amounts of water required for the two states are defined by the plastic and liquid limits. These limits vary with the kind of exchangeable cations and the salt concentration in the adsorbed water. The plastic and liquid limits for different clay minerals are given in Table 21.21. The plasticity index (PI), the difference between the two limits, gives a measure for the rheological (flow) properties of clays. A good example is a comparison of the PI of montmorillonite with that of allophane or palygorskite. The former is considerably greater than the latter, indicating that montmorillonite has a prominent plastic nature. Such rheological properties of clay minerals have great impacts on building foundations, highway construction, chemical engineering, and soil structure in agricultural practice.

21.4.2.3 High-Temperature Reactions

When heated at temperatures beyond dehydroxylation, the clay mineral structure may be destroyed or simply modified, depending on the composition and structure of the clay minerals. In the presence of fluxes, such as iron or potassium, fusion may follow dehydroxylation very quickly. In the absence of such components, particularly for aluminous dioctahedral minerals, a succession of new phases may be formed at increasing temperatures prior to fusion. Thus, in the case of kaolinite, the first high-temperature phase formed is a silica–alumina spinel or γ-alumina plus amorphous silica that is followed at a higher temperature by the development of mullite and cristobalite prior to fusion. In general terms, the first high-temperature phases are a consequence of the original structure of the clay mineral, whereas the later phases are more in accord with the chemical composition. Information on high-temperature-phase change of the clay minerals has been obtained by the use of an x-ray diffractometer to which a small high-temperature furnace is attached. This unit provides x-ray data, while the sample is at an elevated temperature. Information concerning high-temperature reactions is important for ceramic science and industry.

21.4.3 Chemical Properties

21.4.3.1 Layer Charge and Ion Exchange

Depending upon deficiency in the positive or negative charge balance (locally or overall) of mineral structures, clay minerals are able to adsorb certain cations and anions and retain them around the outside of the structural unit in an exchangeable state, generally without affecting its basic silicate structure. These adsorbed ions are easily exchanged by other ions. The exchange reaction differs from simple sorption because it has a quantitative relationship between reacting ions. The range of the cation exchange capacities of the clay minerals is given in Table 21.20 along with specific surface areas. Exchange capacities vary with particle size, perfection of crystallinity, and nature of the adsorbed ion; hence, a range of values exists for a given mineral rather than a single specific capacity. With certain clay minerals—such as imogolite, allophane, and to some extent kaolinite—that have hydroxyls at the surfaces of their structures, exchange capacities also vary

with the pH of the medium, which greatly affects dissociation of the hydroxyls. The ion-exchange charge created as such is called pH-dependent charge or variable charge to distinguish it from the permanent layer charge originating by the ionic substitution in the structure that is independent from pH. The rate of ion exchange varies with clay mineral type and the nature and concentration of the ions. In general, the reaction for kaolinite is most rapid, being almost instantaneous. It is slower for smectites and for sepiolite and palygorskite and requires even long time, perhaps hours or days, to reach completion for illites. Under a given set of conditions, the various cations are not equally replaceable and do not have the same replacing power. Calcium, for example, will replace sodium more easily than sodium will replace calcium. Sizes of potassium and ammonium ions are similar, and the ions are fitted in the hexagonal cavities of the silicate layer. Vermiculite and vermiculitic minerals preferably and irreversibly adsorb these cations and fix them between the layers. Such a preference of one exchangeable cation over another, so-called ion selectivity, may be a tool to exploit layer-charge characteristics of clay minerals in question. Heavy metal ions such as copper, zinc, and lead are strongly attracted to the negatively charged sites on the surfaces of 1:1 layer minerals, allophane and imogolite, which is caused by the dissociation of surface hydroxyls of these minerals.

The ion exchange properties of the clay minerals are extremely important because these properties determine their physical characteristics and economic use. The availability and retention of fertilizer in soils, the adsorption and release of toxic elements in soil and aquatic environments, plasticity, and other clay properties depend to a great extent on ion exchange in general and on the identity of the exchange cation.

21.4.3.2 Solubility

The solubility of the clay minerals in acids varies with the nature of the acid, the acid concentration, the acid to clay ratio, the temperature, the duration of treatment, and the chemical composition of the clay mineral to be attacked. It also varies as a function of heating of the clay minerals and the firing temperatures prior to the acid attack. In general, ferromagnesian clay minerals are more soluble in acids than the aluminum counterparts. Basset (1960) considered that hydroxyl orientation may be an important factor for mica alteration, because in trioctahedral micas hydrogen atoms are oriented directly toward the interlayer K atoms, but in dioctahedral micas, they are tilted away from the interlayer cation. Incongruent dissolutions may result from reactions in a low acid concentration medium where the acid first attacks the adsorbed or interlayer cations and then the components of the octahedral sheet of the clay mineral structure. Frequently with aluminous clay minerals in particular, the tetrahedral silica sheets are not attacked and the morphology of the clay minerals may be retained after solution of all components except silica. In the case of higher acid concentration, such stepwise reactions may not be recognizable and the dissolution appears to be congruent. One of the important factors controlling the rate of dissolution is the concentration in the aquatic medium of the elements extracted from the clay mineral—the greater the concentration of an element in the solution, the more hindered the extraction of the element. In alkaline solutions, a cation exchange reaction first takes place and then the silica part of the structure is attacked. The reaction depends upon the same variables as those stated for acid reactions.

21.4.3.3 Interactions with Inorganic and Organic Compounds

Smectites, vermiculites, and other expansible clay minerals can accommodate relatively large and multivalent inorganic cations between the layers. Because of this multivalency, the interlayer space is only partially occupied by such inorganic cations, which are distributed in the space-like islands. Hydroxy polymers of aluminum, iron, chromium, zinc, and titanium are known examples of such interlayering materials. Most of these are thermally stable and stand as pillars to allow a porous structure in the interlayer space. The resulting complexes, often called pillared clays, exhibit attractive properties as catalysts—namely, large surface area, high porosity, regulated pore size, high solid acidity. Cationic organic molecules, such as certain aliphatic and aromatic amines, pyridines, and methylene blue, may replace inorganic exchangeable cations present in the interlayer of expansible minerals. These organic compounds contain nitrogen atoms, which are able to be positively charged. Polar organic molecules may replace adsorbed water on external surfaces and interlayer positions. Ethylene glycol and glycerol are known to form stable specific complexes with smectites and vermiculites. The formation of such complexes is frequently utilized for the identification of these minerals. On the other hand, potassium acetate can penetrate between kaolin-type silicate layers, which are neutral and held together by hydrogen bonds, and the d_{001} of the mineral expands to 1.42 nm (Wada, 1959, 1961); if dimethyl sulfoxide, DMSO, is used, the expansion is to 1.12 nm (Camazano and Garcia, 1966); if formamide, to 1.0 nm (Weiss et al., 1963). These intercalations with organic compounds are often utilized for the characterization of kaolin minerals. However, experimental results are not always consistent. Caution is required for firm conclusions.

As organic molecules coat the clay mineral surfaces, the nature of clay particle surfaces changes from hydrophilic to hydrophobic, thereby losing its tendency to bind water. Consequently, the affinity of the material for oil increases, so that it can react with additional organic molecules. As a result, the surface of such clay minerals can accumulate organic materials. Some of the clay minerals can serve as catalysts, promoting reactions in which one organic substance is transformed to another on the mineral's surface. Some of these organic reactions develop particular colors, which may be of diagnostic value in identifying specific clay minerals. Organically clad clay minerals are used extensively in paints, inks, and plastics.

21.4.4 Uses of Clays

Clays are perhaps the oldest materials of which man has manufactured various artifacts. The making of fired bricks possibly started some 5000 years ago and was probably the second earliest industry of mankind next to agriculture. The use of clays (probably smectite) as soaps and absorbents was reported in *Natural History* by a Roman author, Pliny the Elder (AD 77). Clays composed of kaolinite are

required for the manufacture of porcelain, whiteware, and refractories. The absence of iron in this clay mineral gives it a white burning color, and the absence of alkalis and alkaline earths gives it a very high fusion temperature that makes it refractory. Whiteware bodies frequently contain talc, pyrophyllite, feldspar, and quartz in addition to the kaolinite clay, in order to develop desirable shrinkage and burning properties. If the kaolinite is poorly crystalline, the clay will have higher plastic and bonding properties.

Clays composed of a mixture of clay minerals, in which illite is most abundant, are used in the manufacture of brick, tile, stoneware, and glazed products. Small proportions of smectite in such clays provide good plastic properties, but in large amounts, smectite is undesirable because it causes too great a degree of shrinkage. Besides the ceramic industry, kaolinite is used as an extender in aqueous-based paints and as filler in natural and synthetic polymers. Smectitic clays (bentonite) are used primarily in the preparation of muds for drilling oil wells. This type of clay, which swells to several times its original volume in water, provides desirable colloidal and wall-building properties. Palygorskite and sepiolite clays also are used because of their resistance to flocculation (grouping or clustering of individual grains or flakes) under high salinity conditions. Certain clay minerals, notably palygorskite, sepiolite, and some smectites, possess substantial ability to remove color bodies from oil. These so-called Fuller's earths are used in processing many mineral and vegetable oils. Because of their large absorbing capacity, Fuller's earths are also commercially used for preparing animal litter trays and oil and grease absorbents. Acid treatment of some smectite clays increases their decolorizing ability. Much gasoline is manufactured by using catalysts prepared from either smectite-type or kaolinite–halloysite-type clay minerals. The preparation may be by acid treatment to modify the structure or by the use of kaolinite and halloysite as a source of alumina and silica for the synthesis of new zeolite-type structures. Large tonnages of kaolinite clays are used as a paper filler and a paper coating to give sheen and carry pigments. The coating clays are washed to free them from grit and then are processed by physical and chemical techniques to improve their whiteness and viscosity. In general, well-crystallized kaolinite that cleaves easily into thin flakes is desired. Palygorskite–sepiolite minerals and acid-treated smectites are used in the preparation of carbonless carbon paper because of the color they develop during reactions with certain colorless organic compounds. Large tonnages of bentonite are used as bonding agents in foundry sands for casting metals. Some poorly crystallized illites and kaolinites are also used for this purpose. Bentonite is combined with lime and coke to pelletize finely ground iron ore, which renders it suitable for use in blast furnaces. Because many clay minerals have aluminum oxide contents on the order of 30%–40%, they are potential ores for aluminum. A variety of processes have been developed to extract this element from clays but clays are not yet competitive with bauxite as a source of aluminum. Extremely large tonnages of reasonably pure clays, preferably of the kaolinite type, would be required for this purpose. Clays have a tremendous number of miscellaneous uses and for each use a particular type of clay with particular properties is important. For example, palygorskite–sepiolite and smectite clays are used as carriers for insecticides and herbicides. Smectite clays are used as plasticizing agents and kaolinite clays are used as extenders and fillers in a large number of organic and inorganic bodies. Kaolinite and smectite clays also are used in a variety of pharmaceutical and medical preparations. For cosmetic products, kaolinite, talc, mica, and smectite clays are used. Smectite and kaolinite clays can be coated with organic molecules for use in many organic systems. Often such organic-clad clays are tailor-made to fit a particular organic system, and thus become an integral part of the system rather than simple diluents. Typical examples are their application to printing inks, oil paints, and lubrication greases. Table 21.22 gives a glimpse of industrial uses of kaolinite, smectite, and palygorskite–sepiolite minerals.

TABLE 21.22 Industrial Uses of Clay Minerals

Kaolinite	Smectite	Palygorskite–Sepiolite
Paper coating	Drilling mud	Drilling fluids
Paper filling	Foundry bond clay	Paint
Extender in paint	Pelletizing iron ores	Paper
Ceramic raw material	Sealants	Ceramics
Filler in rubber	Animal feed bonds	Asphalt emulsions
Filler in plastics	Bleaching clay	Cosmetics
Extender in ink	Industrial oil absorbents	Sealants
Cracking catalysts	Agricultural carriers	Adhesives
Fiberglass	Cat box absorbents	Pharmaceuticals
Foundries	Beer and wine clarification	Catalyst supports
Desiccants	Medical formulations	Animal feed
Cement	Polishing and cleaning agents	Petroleum refining
Pencil leads	Detergents	Anticracking agent
Adhesives	Aerosols	Reinforcing fillers
Tanning leather	Adhesives	Cat box absorbents
Pharmaceuticals	Pharmaceuticals	Suspension fertilizer
Enamels	Food additives	Agricultural carriers
Pastes and glues	Deinking of paper	Industrial floor absorbents
Insecticide carriers	Tape-joining compounds	Mineral and vegetable oil
Medicines	Emulsion stabilizer	Refining
Sizing	Crayons	Tape-joint compounds
Textiles	Cement	Environmental absorbents
Food additives	Desiccants	
Bleaching	Cosmetics	
Fertilizers	Paints	
Plaster	Paper	
Filter aids	Fillers	
Cosmetics	Ceramics	
Crayons	Catalysts	
Detergents	Pencil leads	
Roofing granules		
Linoleum		
Polishing compounds		

21.5 Identification of Soil Phyllosilicates

21.5.1 Introduction

Clay mineralogical analysis is a very important first step to elucidate soil behavior, because the clay fraction of soil is most active and influential to the physical, chemical, and biological reactions of soil. X-ray diffraction is a universally accepted method for identifying crystalline mineral components (Brindley, 1951; Brown, 1961; Brindley and Brown, 1980). Unlike the case where only pure crystalline substances are dealt with, soils are complex mixtures of various minerals with a wide range of crystallinity and compositions. Because of these complexities, we often have to compromise with identification and quantification of mineral group rather than mineral species, and the estimation of relative amounts of minerals may often be the best for a given circumstance. Nevertheless, such data provide useful information about mineral translocations and transformations that might have taken place in the soil profiles in question.

21.5.2 X-Ray Diffraction Methods

21.5.2.1 Basics

X-ray diffraction techniques provide a set of reflections from crystalline substances with various diffraction intensities as a function of diffraction angle. The simplest is, therefore, to take standard diffraction diagrams of the principal clay minerals and of the most commonly occurring impurities and to make direct comparison with the diffraction diagrams of unknown clays. For this purpose, *Mineral Powder Diffraction File Data Book* published by JCPDS (Joint Committee on Powder Diffraction Standards) is handy and useful. X-ray reflection takes place from lattice planes with indices, *hkl*, according to the Bragg law,

$$n\lambda = 2d_{hkl} \sin\theta \qquad (21.1)$$

where
  d_{hkl} is lattice spacing for planes *hkl*
  λ is the wavelength
  θ is the glancing angle of reflection
  n is the order of the reflection

The latter two are often called Bragg's angle and the order of Bragg's reflection, respectively. Thus, diffraction experiments measure the quantity from

$$\frac{d_{hkl}}{n} = \frac{\lambda}{2\sin\theta} \qquad (21.2)$$

and this is the lattice spacing for *n*th-order reflection, which is readily understood by introduction of the reciprocal space concept. The Bragg equation gives no intensity information. Diffraction intensities are generally expressed by the relation

$$I = \text{const.(Lp)} \, IFI^2G \qquad (21.3)$$

where
  the constant factor includes all the basic electronic terms, which arises from the fundamental x-ray scattering process
  Lp is the Lorentz-polarization factor
  F is the structure factor that is related to number, kind, and array fashion of the atoms on a lattice plane and the factor can be a positive or negative number
  G is the Laue function that determines diffraction peak profile (Kodama et al., 1971) and, therefore, provides information about the distribution of crystallite sizes and shapes and their crystal imperfections

Figure 21.29 illustrates XRD patterns from oriented samples of muscovite and phlogopite for comparison. They show a series of basal reflections, which can be traced down to at least fifth order. The two minerals have a similar structure except for their octahedral cations, aluminum in muscovite and magnesium in phlogopite. This difference affects the IFI^2 value and reflects on the diffraction pattern. Thus, the ratio $(I_{001}/I_{002})_{\text{phlogopite}}$: $(I_{001}/I_{002})_{\text{muscovite}}$ may be used as one of the criteria for identification. Similarly, this approach is also useful for the determination of chlorite species (Petruk, 1964; Bailey, 1972). If mineralogical data accumulation is the main purpose, more reflections should be observed. Figure 21.30 illustrates differences in diffraction patterns from random and oriented powder specimens of a muscovite. To determine polytypes, not only are the basal reflections required, but the *b* unit cell dimension from the d_{060} reflection and impurities must also be known.

21.5.2.2 Distinguishing among Soil Phyllosilicates

As mentioned earlier, the types of phyllosilicates occurring in soils (and particularly in its clay fraction) are limited to only several layer silicate structures (Figure 21.31), and interstratified layer silicates can also be considered as a combination of two or three basic structures. Obviously from Figure 21.31, the identification of those basic layer structures is due to determination of respective layer thickness (distance normal to the basal plane), which is obtained by XRD as d_{001}-spacing. By choosing effective and critical conditions under which basic layer units are clearly distinguished, criteria for the identification of clay minerals by XRD can be set up (Bradley, 1945). For example, the distinction between smectite and vermiculite is made as follows: Mg-saturation and treated with glycerol expand the d_{001}-spacing of smectite to 1.7–1.8 nm but do not change the d_{001}-spacing at 1.4 nm of vermiculite (Walker, 1949). With the same treatment, chlorite, mica, and kaolinite do not change their d_{001}-spacings at 1.4, 1.0, and 0.72 nm, respectively. On heating at 550°C, however, the structure of kaolinite is destroyed. The d_{001}-spacing at about 1.0 nm of hydrated halloysite contracts to 0.72 nm on heating at 100°C, therefore easily distinguishable from mica. The distinction between vermiculite and chlorite is based on the heat treatment at 550°C or saturation with K$^+$, either of which causes the reduction of the d_{001}-spacing at 1.4–1.0 nm in vermiculite but no change in chlorite. These diagnostic criteria are based on
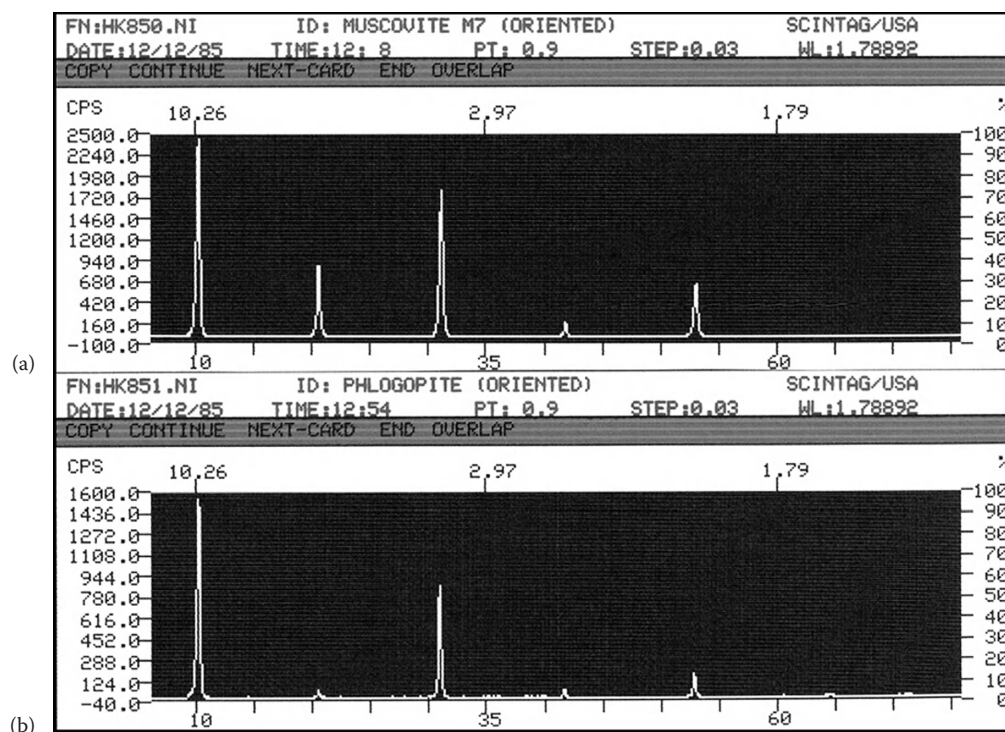
FIGURE 21.29 XRD patterns of preferred oriented samples: (a) muscovite and (b) phlogopite.
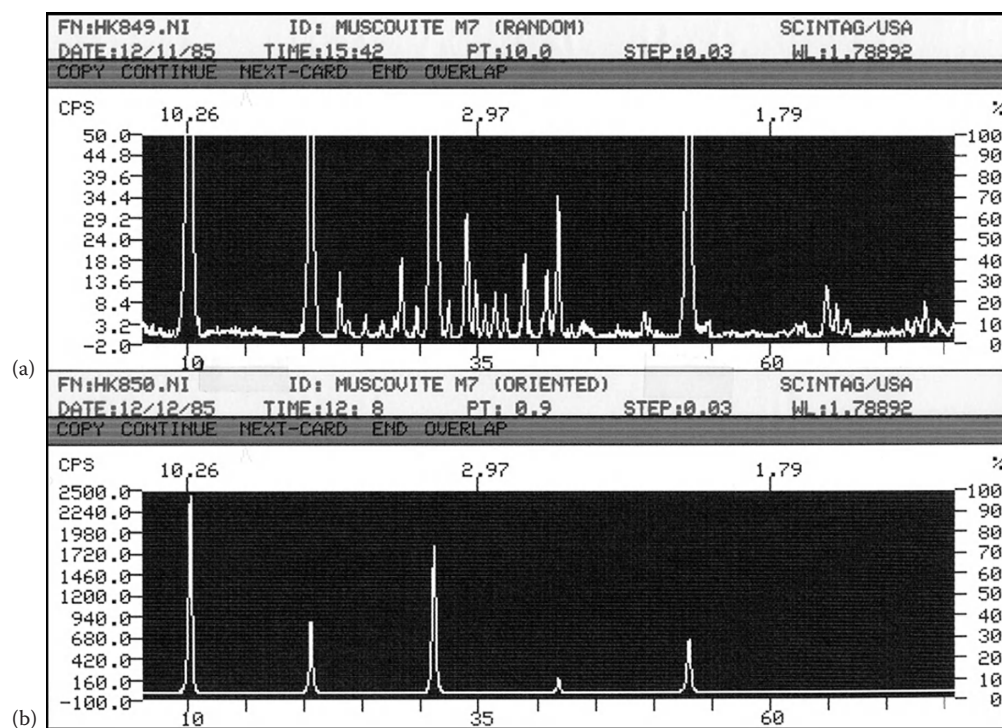
FIGURE 21.30 XRD patterns of muscovite: (a) random sample and (b) oriented sample.
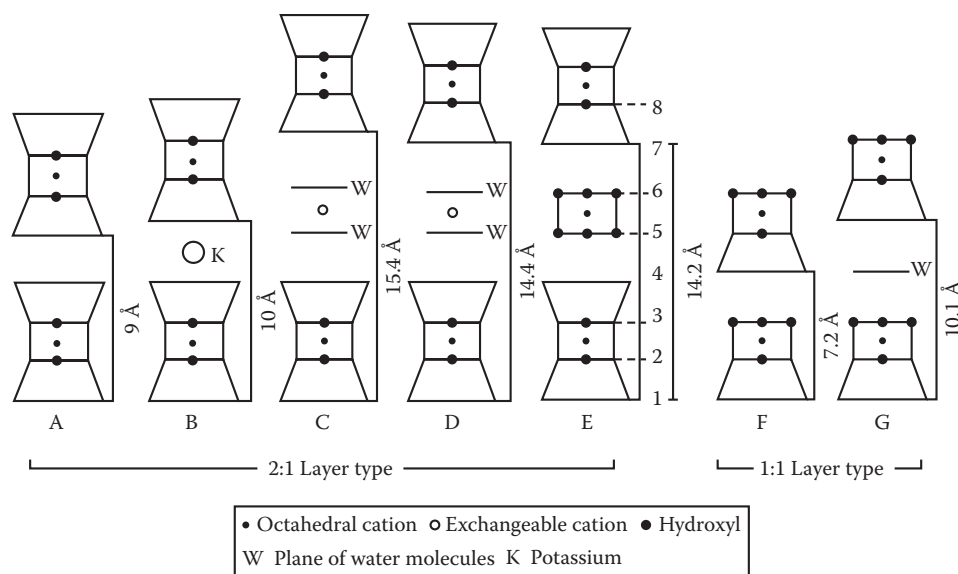
FIGURE 21.31 Diagrammatic representation of principal clay mineral structures, comparing the relative layer thickness of the unit structures. A, Talc–pyrophyllite; B, mica; C, smectite; D, vermiculite; E, chlorite; F, serpentine–kaolinite; and G, hydrated halloysite (halloysite—1.0 nm). In this presentation, tetrahedral and octahedral anions and tetrahedral cations are omitted. (From Sudo, T., S. Shimoda, H. Yotsumoto, and S. Aita. 1981. Electron micrographs of clay minerals. Developments in sedimentology. Vol. 31. Elsevier Scientific Publishing Co., New York.)

generally accepted specific properties of each mineral group for its definition, which may be categorized as follows:

- *Mica.* A 1.0 nm basal spacing and integral series of higher order reflections do not shift upon either hydration or glycerol salvation or heating at 550°C and 700°C.
- *Kaolinite.* Characterized by a series of basal reflections at 0.72, 0.36 nm, and so forth, which do not shift upon either hydration or solvation with glycerol but disappear upon heating at 550°C.
- *Chlorite.* An integral series of basal reflections associated with the largest spacing of 1.4 nm; does not expand by glycerol sorption and does not shift upon heating at 550°C. On heating at 700°C, the intensity of the 1.4 nm peak is enhanced in comparison with those of other peaks. One should bear in mind that the basal intensity ratio varies with chemical composition. The higher order reflections such as third- and fourth-order reflections at 0.473 and 0.355 nm, respectively, may also be used for the distinction of chlorite from kaolinite.
- *Vermiculite.* According to the conventional definition, its Mg form maintains the 1.44 nm basal spacing upon solvation with glycerol and hydration, but the spacing collapses to 1.0 nm upon heating to 550°C, whereas the K form gives the contracted spacing of 1.0 nm without heating. Although its Ca form also shows symptoms similar to those of the Mg form, in some cases, the symptoms become less definite, that is, some of the Ca forms expand beyond 1.44 nm on glycerol solvation depending upon their layer-charge characteristics.
- *Smectite.* Under intermediate humidity (about 50%), the Na form of smectite shows an expansion of its basal spacing to 1.24 nm, and the Ca (also Mg) form to 1.54 nm

(Mooney et al., 1952). These forms expand to 1.8 nm upon solvation with glycerol. Heating to 550°C causes a reduction of the basal spacing to 1.0 nm. It collapses further to 0.96 nm at 700°C, accompanied by its third-order peak at 0.32 nm. This behavior is also a characteristic of vermiculite, but nonexpandable clay minerals like mica retain their third-order peak at 0.333 nm.

After clay mineral components are distinguished at a level of the mineral group category, and if circumstances allow, the procedure of identification continues to determination at the level of mineral subgroup and finally mineral species. To reach the final stage of identification, most likely further tests are needed. For the distinction between montmorillonite and beidellite, for example, the so-called Greene-Kelly test (Greene-Kelly, 1953, 1955) based on the Hofmann–Klemen effect (Hofmann and Klemen, 1950) is often applied. For apparently chloritic minerals derived from vermiculite by acquiring hydroxide sheets between vermiculite silicate sheets, the interlayered vermiculite can be distinguished from chlorite by complexing with sodium citrate to remove the interlayer material. Further information may be obtained from XRD data if complexities due to the combination of clay mineral components themselves and associated impurity are relatively limited.

Ideally, any pretreatment would be avoided to obtain information about mineral components of soil as they are. However, cases are often unavoidable. An extreme example is given in Figure 21.32, where only a pretreatment of the clay fraction by Tiron improves XRD patterns because masking substances are mainly amorphous silica. Many soils cannot disperse without treatment for fractionation. Organic matter, carbonate, and iron oxide removal are often prerequisite. Another example
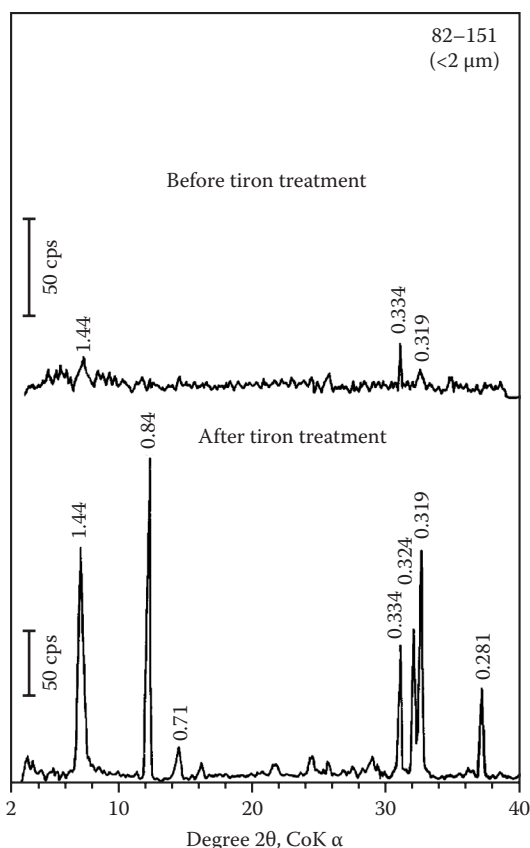
FIGURE 21.32 XRD patterns (in nm) of a soil clay before and after tiron treatment, exemplifying any suitable chemical pretreatments are often required to obtain decent XRD patterns of soil clays for identification. (From Kodama, H., and G.J. Ross. 1991. Tiron dissolution method to remove and characterize inorganic components in soils. Soil Sci. Soc. Am. J. 55:1180–1187.)

represents effects of various treatments on XRD patterns (Figure 21.33). After fractionation, subsamples (from a clay fraction of podzolic B horizon) were subjected to various treatments. After destruction of organic matter, the <2 mm size fraction was dispersed in water by shaking overnight on an end-over-end shaker, and the clay fraction was collected by sedimentation, saturated with Mg^{2+}, and freeze-dried. This illustrates the necessity of pretreatments for amorphous substances, which mask crystalline components. A quick reference to major chemical dissolution methods is given in Table 21.23. Details of pretreatments may be consulted with appropriate manuals (Bear, 1964; Jackson, 1967; Thorez, 1975; McKeague, 1976; Klute, 1986; Dixon and Weed, 1989; Carter, 1993).

21.5.2.3 Soil Phyllosilicates of Intermediate Nature

In soil environments, especially of the temperate climate region, a majority of the weathering actions on phyllosilicates includes hydration, ion exchange, selective depletion, and partial decomposition of the minerals. As a whole, the actions are considered to be gentle ones. Under these circumstances, micas partially loose their interlayer cations to become hydrated
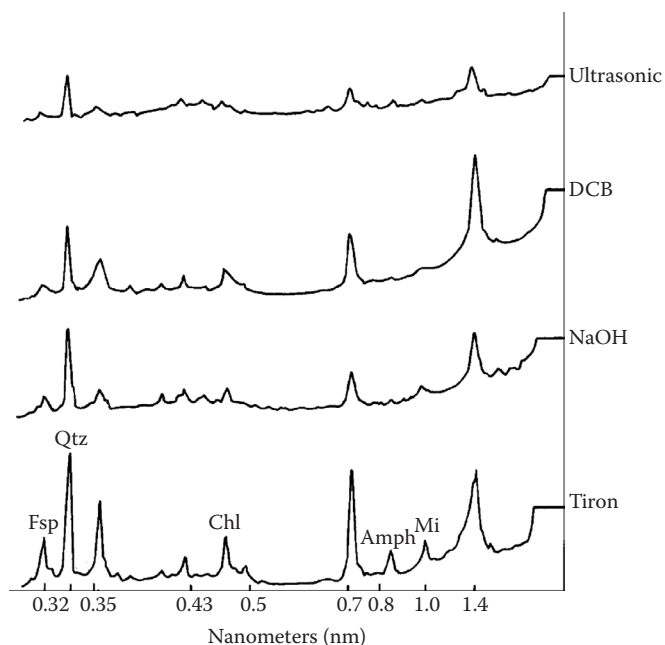
FIGURE 21.33 XRD patterns of a Mg-saturated and glycerol-solvated clay sample after four separate pretreatments. After organic matter removal, <2 mm fraction was collected from the Whatcom Bf horizon and dispersed by an end-to-end shaker and then the clay fraction was obtained by sedimentation. (From Wang, C., H. Kodama, and N.M. Miles. 1981. Effect of various pretreatments on X-ray diffraction patterns of clay fractions of podzolic B horizons. Can. J. Soil Sci. 61:311–316.)

and expandable. Chlorites gradually experience a mild attack on their hydroxide sheets to alter toward vermiculite-like minerals. Vermiculites may undergo further ion exchange and hydration with their interlayer cations to alter toward smectites. On the contrary, in a lower horizon of the soil profile, an accretion process such as acquiring interlayer substances may take place. By this process, the following transformation routes are also possible: smectites through vermiculites to chlorites, vermiculites to chlorites, micas to chlorites, smectites through vermiculites to micas, vermiculites to micas, and so on. As these actions constantly work on minerals in soil, encountering a phyllosilicate mineral that cannot be correctly identified with a definitive soil phyllosilicate is no surprise. For example, although a soil phyllosilicate indicates a nature similar to vermiculite or smectite, positive identity as either entity may be impossible. This is due to the fact that the soil phyllosilicates possess an intermediate nature. We have observed such cases in the series mica–vermiculite, vermiculite–chlorite, smectite–chlorite, etc. One of the notable accretion processes in soils is the reaction of hydroxyl-Al ions with 2:1 phyllosilicates. Aluminum ions solubilized from minerals hydrolyze readily and are polymerized to form precipitates between expandable silicate layers like vermiculite or smectite to form a chlorite-like mineral. They can be referred to as hydroxyl-Al interlayered vermiculite or hydroxyl-Al interlayered smectite, respectively. These names have been shortened to the acronyms HIV and HIS for which Barnhisel and Bertsch (1989) proposed to standardize the terminology.

TABLE 21.23 Major Chemical Dissolution Methods

Category	Chemical Reagents Used for Dissolution	Selected References
(a) Alkaline	NaOH	Foster (1953), Hashimoto and Jackson (1960)
	Na_2CO_3	Mitchell and Farmer (1962), Follett et al. (1965)
(b) Acid	HCl	Deb (1950)
(a) + (b) alkaline + acid	NaOH/HCl	Segalen (1968)
(c) Complexing	$H_4C_2O_4$, acetic acid method	Chester and Hughes (1967)
	$H_2C_2O_4$, oxalic acid method	Ball and Beaumont (1972)
	$(NH_4)_2C_2O_4/H_2C_2O_4$, acid NH_4–oxalate method	Tamm (1922), Schwertmann (1959)
	$Na_2C_2O_4/H_2C_2O_4$, acid Na–oxalate method	Henmi and Wada (1976)
	$C_6H_4Na_2O_8S_2$, alkaline tiron method	Biermans and Baert (1977)
	$Na_4P_2O_7$, Na–pyrophosphate method	McKeague (1967)
	$K_4P_2O_7$, K–pyrophosphate method	Ball and Beaumont (1972)
(d) Reducing	$NH_2OH \cdot HCl/HCl$, hydroxylamine method	Chao and Zhou (1983), Ross et al. (1985)
(c) + (d) Reducing + complexing	$Na_2C_6H_5O_7/NaHCO_3/Na_2S_2O_4$ sodium–dithionite method	Mehra and Jackson (1960)
	$K_2C_2O_4/H_2C_2O_4/Mg$, Jeffries' method	Jeffries (1946)
	$(NH_4)_2C_2O_4/H_2C_2O_4/Na_2S_2O_4$, combined oxalate method	Duchaufour and Souchier (1966)

21.5.2.4 Interstratified Layer Silicates

As mentioned earlier in the structure of interstratified layer silicates, due to the structural similarities of the clay minerals, interstratified minerals can exist in which individual crystallites are composed of elementary layers of two or more types. Such mixed structures, in some cases, do not show characteristic features of each of the component layers but possess their own identity. They show more or less intermediate nature of the component layers, they are thus called intergrades. Therefore, this intermediate nature may be useful to identify component layers. Table 21.24 gives an example of diagnostic criteria. Interstratified, or mixed layer, clay minerals represent a special case of intergrowths, in relation to a genetic relationship between components. Neglecting all questions regarding interlayer cations, ionic substitution, and layer charge, montmorillonites and vermiculites, for instance, are essentially hydrated varieties of mica having interlamellar water molecules. Instead of water, introduction of hydroxide sheets between silicate layers of these minerals results in chlorite. Therefore, interstratifications among those minerals are expected to be common, as in fact they are. Figure 21.34 gives a graphic presentation for each of the clay minerals as an elementary component, which is known to be composed of an interstratified structure with another elementary component.

The most striking interstratifications occurs when two-component layers A and B alternate regularly to make an interstratified structure, $AB\,AB \ldots$ type, having a long d_{001}-spacing = d_{001}-spacing (layer A) + d_{001}-spacing (layer B). Let P_{AB} be the probability that B succeeds A given that the first layer is A, perfect regularity must satisfy the conditions $P_{AB} = 1$ and $P_{AA} = 0$. As the evaluation of such conditions is not always possible, a practical measure is recommended by the Clay Minerals Society Nomenclature Committee (Bailey et al., 1982). This is based on a statistical test on a well-defined series of at least 10 00-summation spacings ($d_{AB} = d_A + d_B$), for which the suborders are integral. The coefficient of variation

TABLE 21.24 Diagnostic Criteria for Interstratified Minerals in Soils

Type of Interstratification[a]	Glycerol	Humid, >95% RH	Air-Dry, 45% RH	Very Dry, <3% RH	110°C	550°C
Mica–vermiculite	[b]	[b]	1.0–1.4	1.0–1.15	1.0–1.15	1
Mica–chlorite	[b]	[b]	1.0–1.4	[b]	[b]	1.0–1.4
Mica–smectite	0.9–1.0 or ≈1.8	Expansion	1.0–1.55	1.0–1.15	1.0–1.15	1
Vermiculite–chlorite	[b]	[b]	1.4	1.15–1.4	1.15–1.4	1.0–1.4
Vermiculite–smectite	1.4–1.8	Expansion	1.4–1.54	1.15	1.15	1
Chlorite–vermiculite	1.4–1.8	Expansion	1.4–1.54	1.15–1.4	1.15–1.4	1.0–1.4
Mica–vermiculite–chlorite	[b]	[b]	1.0–1.4	Contraction	Contraction	1.0–1.4
Mica–vermiculite–smectite	Expansion	Expansion	1.0–1.54	Contraction	Contraction	1
Vermiculite–chlorite–smectite	Expansion	Expansion	1.4–1.54	Contraction	Contraction	1.0–1.4

Source: Modified after Kodama, H., and J.E. Brydon. 1968a. A study of clay minerals in podzol soils in New Brunswick, eastern Canada. Clay Miner. 7:295–309.

Expansion and contraction mean shift of the original first-order reflection (air-dry) toward large and small spacing sides, respectively.

[a] As Ca-saturated specimens.

[b] Indicates that the basal spacing is unchanged from air-dry.

FIGURE 21.34 Diagrammatic representation of layer silicate type combination found in nature as interstratified clay minerals. ■ = frequent occurrence; □ = less-frequent occurrence. (Modified after Sudo, T., S. Shimoda, H. Yotsumoto, and S. Aita. 1981. Electron micrographs of clay minerals. Developments in sedimentology. Vol. 31. Elsevier Scientific Publishing Co., New York.)

(CV) of the d_{00}-values should be less than 0.75 to demonstrate adequate regularity of alternation. The CV is defined as CV = $100s/X$, where the standard deviation for a small sample is $s = [\Sigma(X_i − X)^2/(n − 1)]^{1/2}$, X_i is an individual observed $− x$ d_{00}-value, X is the mean of the X_i values, and n is the number of observed X_i values. In Table 21.25, an example is shown using the case of rectorite–Ca, which is a regularly interstratified mineral, smectite–Ca mica. The observed long d_{001}-spacing is 2.8 nm on glycolation. This agrees with the summation of 1.8 nm for smectite and 1.0 nm for mica when they are glycolated. The XRD data of the rectorite show observable basal reflections up to the 20th in that recording range. The $d \times \ell$ values in the table correspond to observed $− x$ d_{00}-values (X_i). Calculations for this rectorite result in 2.751 nm for the average (X) and 0.463 for the CV value that is much smaller than the recommended value of 0.75. The majority of interstratified layer silicates is not well ordered, so they do not show basal reflection of a long spacing ($d_{AB} = d_A + d_B$) nor a rational integral series of basal reflections.

TABLE 21.25 Example of the CV Test for Evaluating Regularity of Interstratifications

00ℓ	d (Å)	I	d (Å) $\times \ell$
	Rectorite–Ca, Sample # 1820		
1	28.09	847	28.09[a]
2	14.01	174	28.02[a]
3	9.267	84	27.8
4	6.923	40	27.69
5	5.495	7	27.47
6	4.588	31	27.53
7	3.933	23	27.53
8	3.442	33	27.54
9	3.065	42	27.58
10	2.723	4	27.23
11	2.503	7	27.53
12	—	—	—
13	—	—	—
14	1.963	6	27.48
15	1.832	5	27.48
16	—	—	—
17	1.615	3	27.46
18	1.527	3	27.49
19	1.443	2	27.42
20	1.372	3	27.4
Average			27.51
CV			0.463

Source: After Matsuda, T., H. Kodama, and A.-F. Yang. 1997. Ca-rectorite from Sano mine, Nagano prefecture, Japan. Clays Clay Miner. 45:773–780.

[a] Excluded from CV calculations.

Characterization requires:

1. Determination of component layers A and B (or, and C)
2. Evaluation of the mixing ratio for the component layers A and B, that is, A:B = m:$(1 − m)$, where $0 < m < 1$
3. Probability for layer-stacking sequence as expressed by P_{AA}, P_{AB}, etc

The determination of component layers can be done by choosing appropriate diagnostic tests for them. A relatively simple method is available to evaluate the mixing ratio of two components in an irregularly interstratified structure. The principle of the method is based on the peak migration due to the interference phenomenon of x-rays (Méring, 1949). According to Méring's idea, Figure 21.35 is constructed to explain the principle, using
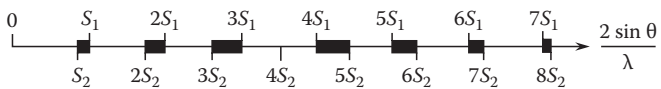
FIGURE 21.35 Demonstration of Méring's method, in which peak migration is used for evaluating the mixing ratio of two-component layers in an interstratified clay mineral. S_1: Na-saturated rectorite, S_2: Ca-saturated rectorite.

modified rectorite samples. Sodium (Na)-saturated rectorite (S_1) and Ca-saturated rectorite (S_2) were prepared separately. Because the hydration state of Na-saturated smectite differs from that of Ca-saturated smectite under relative humidity at 50%, S_1 gives d_{001} at 2.24 nm and S_2 at 2.54 nm. In this figure, the 00ℓ reciprocal lattice rows of reciprocal spacing $S_1 = 1/d_1$ and $S_2 = 1/d_2$ are given and the heavy lines, such as between S_2 and S_1, indicate the regions in reciprocal space where reflections are predicted to appear as two-component peaks merging together when the two are mixed coherently (in respect with x-rays) without physical segregation and do not constitute a regularly interstratified structure. The assumption is that the peaks move linearly from the position for one pure component toward the nearest position of the other component as the proportions of the components alter. This may not be rigorously correct, but information obtained this way is quite useful. This estimation should be performed by inspecting the peak migrations in as many regions as possible.

As the next step, it is desirable to determine the probability of layer-stacking sequence. A direct Fourier transform method developed by MacEwan and Wilson (1980) and MacEwan et al. (1961) has been applied widely. With mathematical approximation, MacEwan formulates the equation as

$$W_R = \left(\frac{a}{\pi}\right) \Sigma i(s)\cos(\mu_S R) \qquad (21.4)$$

with

$$i(s) = \frac{I_s}{\Theta_s IF_s I^2}$$

where

  W_R is the distribution function defined as the probability of finding another layer at the distance R measured perpendicularly from any layer

  μ_S, I_s, Θ_s, and $IF_s I^2$ are values at the position of the intensity maximum

  $\mu_S = 2\sin\theta/\lambda$

  I_s is the integrated intensity

  Θ_s is the operational function

  $IF_s I^2$ is the layer structure factor

Therefore, a Fourier transform on a series of basal reflections of unknown mixed-layer mineral gives the probability of finding layer B succeeding layer A (P_{AB}) and so on. An application to soil minerals is shown in Figure 21.36. This case happened to be a tertiary system, smectite–vermiculite–mica. The first three peaks (h_A, h_B, h_C) correspond to the probabilities of occurrence of the fundamental component layers, A (smectite), B (vermiculite), and C (mica), respectively. Obviously $P_A + P_B + P_C = 1$, therefore, $P_A = 0.53$, $P_B = 0.27$, and $P_C = 0.20$ were obtained from the observed heights for sample SMJ 1. Knowing the relation between peak heights and occurrence frequencies (for instance, $h_{AA} = P_A P_{AA}$, $h_{AB} = P_A P_{AB} + P_B P_{BA}$, etc.), calculating heights for all possible layer sequences and comparing the calculated with
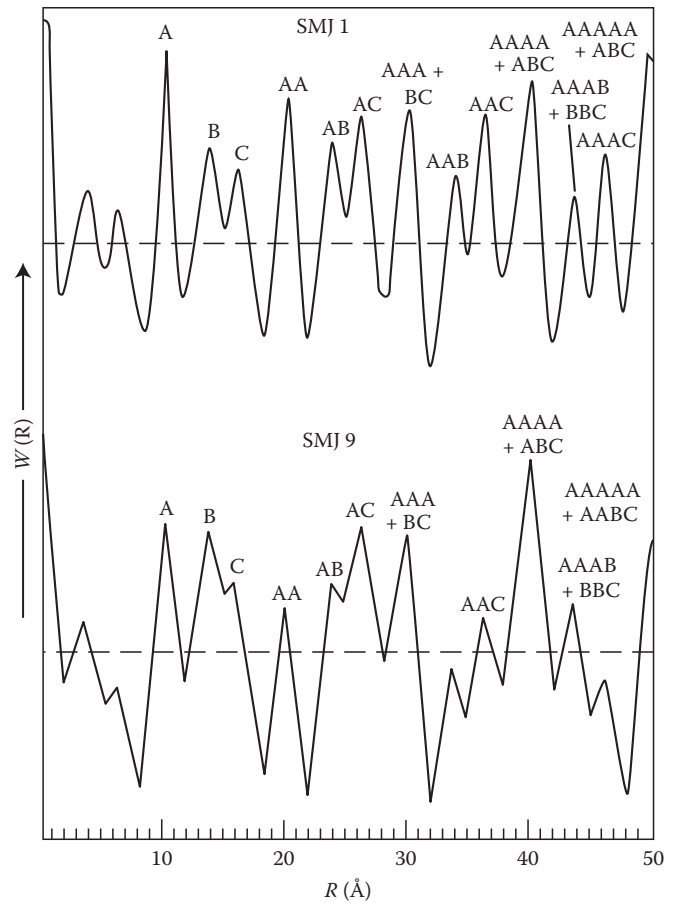
FIGURE 21.36 Example of direct Fourier transform method by MacEwan on soil clays. (From Kodama, H., and J.E. Brydon. 1968a. A study of clay minerals in podzol soils in New Brunswick, eastern Canada. Clay Miner. 7:295–309.)

the observed allow the probability coefficients to be estimated from the results of the Fourier transform. The direct method is applicable to any interstratifications but caution should be paid to the precision of the results obtained by the Fourier method, which may be affected by premature termination of basal reflection series and assessing the baseline for the obtained occurrence probability curve. Thus, preliminary information concerning the mixing ratio and layer-stacking sequence is acquired and diffraction patterns can be simulated by computer calculations. Several simulation approaches are available (Kakinoki and Komura, 1952, 1954; Sato, 1965, 1973; Reynolds, 1980; Moore and Reynolds, 1989; Yuan and Bish, 2010) for further refinement of the structure of interstratified layer silicates, if cases allow.

21.5.3 Other Instrumental Methods

21.5.3.1 X-Ray Diffraction Photograph Methods

Undoubtedly x-ray diffractometer methods are dominating to collect diffraction data nowadays. However, in the case where only a limited quantity of sample is available, classical photographic methods by the Debye–Scherrer camera is still useful

to obtain maximum information on a mineral in a fraction of milligram. This camera method can avoid preferred orientation to which phyllosilicates have a strong tendency. The determination of polytypes requires details of nonbasal reflections. Therefore, the use of camera techniques is beneficial. Gandolfi camera is one type of a modified Debye–Scherrer x-ray powder diffraction camera. In this camera (Figure 21.37a), the mounted sample rotates around two axes, that is, the sample stage arm rotating around the center of the camera and the sample on a goniometer head itself rotating around an axis tilted 45° to the central rotating axis. With this camera, a single-mineral crystal, or grain, or aggregate, in a fraction of millimeter size, is sufficient to provide adequate powder diffraction patterns without grinding. Examples are presented in Figure 21.38b. Wilson and Clark (1978) developed a method for identifying minerals on thin sections by a combination of optical observations and XRD analysis. This is an interesting area to be explored further for routine application.

(a) (b)
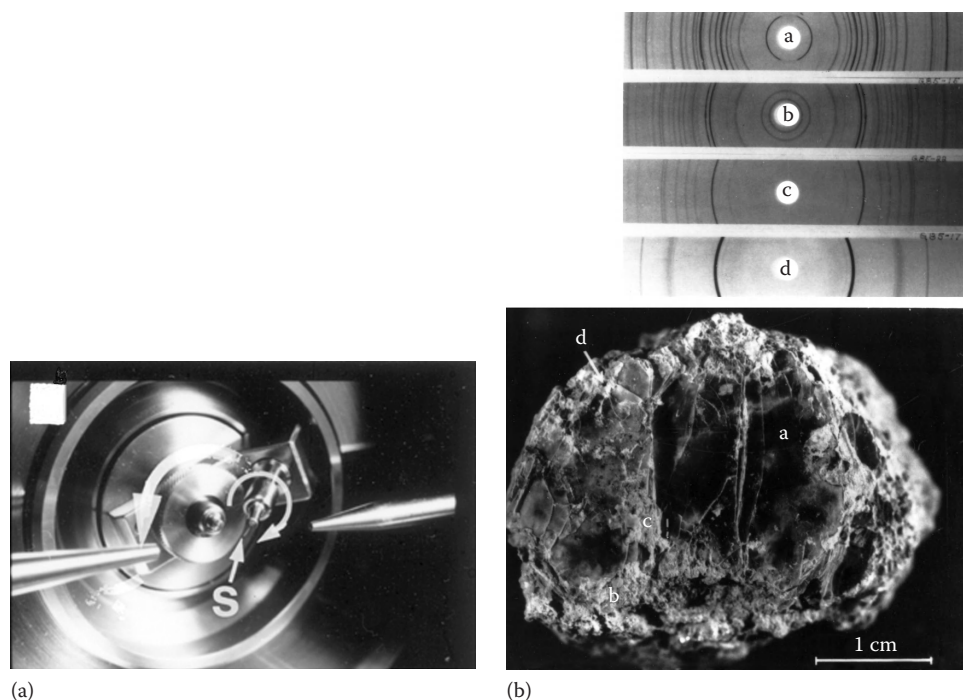
FIGURE 21.37 An application of Gandolfi diffraction camera method to a small grain of mineral for collecting data without pulverizing the sample. (a) Shows the mounted sample rotating around two axes. (b) XRD patterns obtained. A, Phlogopite; B, corrensite; C, calcite; and D, graphite. A, B, C, and D correspond to those locations in the weathered phlogopite crystal piece where samples were taken. (Modified after De Kimpe, C.R., N. Miles, H. Kodama, and J. Dejou. 1987. Alteration of phlogopite to corrensite at Sharbot Lake, Ontario. Clays Clay Miner. 35:150–158.)

(a) (b)
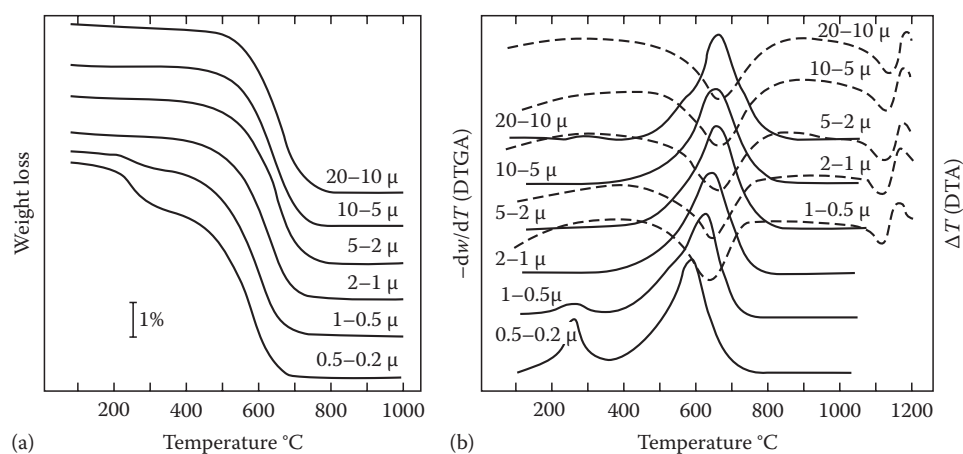
FIGURE 21.38 (a) TGA curves and (b) DTGA (solid lines) and DTA (dashed lines) curves of the fractionated samples of a naturally occurring micro-crystalline muscovite. (From Kodama, H., and J.E. Brydon. 1968b. Dehydroxylation of micro-crystalline muscovite. Trans. Faraday Soc. 64:3112–3119.)

21.5.3.2 Methods for Specific Characterization

As discussed in the section on crystal structures, the 1:1 type structures possess hydroxyls in two different octahedral environments, that is, one locates inside at the interface between the octahedral and tetrahedral sheets, and another is exposed to the outside of the octahedral sheet. This is one of the unique characteristics of the 1:1 layer silicates, which may be used for their identification. Infrared absorption spectroscopy can provide details of the OH-stretching bonds in the region of 4000–3000 cm^{-1} in terms of its orientation against the basal plane and its bond distance (O–H$\cdots$O) as affected by octahedral environments, including elemental compositions (Tuddenham and Lyon, 1959; Nakamoto, 1963). The orientation is known to depend on octahedral cation vacancies; therefore, a similar view can be extended to the difference in dioctahedral and trioctahedral 2:1 type structures as well (Serratosa and Bradley, 1958). A number of papers have been published along these lines (only selected ones: Vedder and McDonald, 1963; Vedder, 1964; Farmer, 1974; Robert and Kodama, 1988).

Dealing with iron-bearing phyllosilicates, the actual amount of iron incorporated structurally; the Fe^{2+}:Fe^{3+} ratio as oxidation state; structural configuration, like distribution of these cations between *cis-* and *trans*-octahedral sites; the amount of Fe^{3+} in tetrahedral sites; and occasionally their role in the magnetic nature of the minerals are sought. In the presence of organic matter, the determination of Fe^{2+} is impossible by wet-chemical methods. Application of the Mössbauer effect under these situations is very appropriate. The ^{57}Fe Mössbauer spectroscopy provides parameters such as isomer shift, IS (mm s^{-1}), quadrupole splitting, ΔE_Q (mm s^{-1}), and internal magnetic hyperfine field, H (in *kOe*), if magnetic ordering is present. Most key information is foiled in the quadrupole doublets. Spectra are recorded at temperatures ranging from 4 to 298 K (liquid He to room temperature) to identify the oxidation state, coordination environment, and magnetic exchange interactions occurring in the constituent minerals. The parameters allow the identification of Fe in various (oxy-hydr)oxide and phyllosilicate phases in mixed mineral systems. Analysis of Mössbauer spectra is carried out by computer using a least-squares minimization fitting routine for obtaining the best peak profile. Spectral parameters obtained after refinement may give answers to the questions above (Stucki et al., 1988; Coey, 1988; Murad, 1988; Murad and Cashion, 2004). Among minerals having a chlorite structure, the distinction between iron atoms in the octahedral sheet and iron atoms in the hydroxide sheet has yet to succeed. Murad and Wagner (1991, 1994) succeeded in obtaining detailed characterization of kaolinite minerals and illite by the application of Mössbauer spectroscopy.

The distribution of aluminum atoms that substitute for silicon atoms in the tetrahedral sheet is an interesting subject from a crystallographic point of view (Loewenstein, 1954). Investigations by x-ray methods were performed as done for muscovite by Gatineau (1964). Because the data needed for this purpose require tedious recording and collection of non-Bragg reflections, only specialized laboratories can afford such experiments. Besides the x-ray method, the application of ^{29}Si (and ^{27}Al) magic-angle spinning, solid-state nuclear magnetic resonance spectroscopy was introduced into the research in this line (Herrero et al., 1985). The spectra are obtained as chemical isomer (in this case ^{29}Si) shift (ppm), which is influenced by the chemical environment of the nucleus. The range where resonance spectra appear as a function of polymerization (if true mica, denoted as Q^3 for Si$_3$Al$_1$) and analysis of spectra provides information about Si and Al ordering in the tetrahedral sheet, including the number of next-neighbor Si in layer silicates. Conversely, using the ^{27}Al nucleus, spectra acquired supply information regarding Al and Si ordering and the number of next-neighbor Al in layer silicates (Sans and Serratosa, 1984). Data can be used to compliment each other.

The origin of the total layer charge in phyllosilicates is due to ionic substitutions by atoms of lower valence for those of higher valence like Al^{3+} → Si^{4+}, Fe^{2+} → Fe^{3+}, Mg^{2+} → Al^{3+}. The way in which layer charge is distributed in the structure of a specific phyllosilicate appears to closely relate to the course of weathering or alteration, which the mineral might take in nature. We often believe that phyllosilicates behave erratically, but this is because we normally treat them as if they were totally "structurally homogenous." For instance, when smectite is suspended in water containing Na$^+$ and Ca^{2+}, a demixing phenomenon is observed (Glaeser and Méring, 1954; Glaeser, 1958). Certain phyllosilicates behave as if the distribution of their layer charge is not symmetrical, in other words, the layer charge distribution in the structure has a polarity. In this connection, the *n*-alkylammonium exchange method developed by Lagaly and Wiess (1969) is useful for estimating interlayer charge and characterizing expandable layer silicates by making complexes with the surface of the expandable layer silicates, then the d_{001}-spacings of the complexes are measured. The magnitude of the spacing is related to the size and tilt angle (α) of the *n*-alkylammonium intercalated between the expandable silicate layers. The tilt angle (α), that is, the angle between the alkyl chain and the basal plane of the clay, is related to the layer charge. To evaluate layer charge irregularity, peak positions and profiles of higher order basal reflections may need to be inspected in detail (Lagaly, 1979, 1982; Ghabru et al., 1989) Vali and Koster (1986) developed a method that allows us to make the alkylammonium ion treatment of clay minerals in ultrathin sections for high-resolution transmission electron microscope (HRTEM) observations. The method proved promising and awaits further applications, especially studies on microenvironments.

The task of distinguishing and quantifying the various types of water, that is, physically adsorbed, zeolitic, interlayer, and structural, is not easy. Thermoanalytical methods including thermogravimetric analysis (TGA), differential TGA (DTGA), and differential thermal analysis (DTA) are often applied for this purpose. Although TGA is indispensable for obtaining weight loss, DTGA is helpful to evaluate a critical temperature for distinguishing between water of two different types. The features in a TGA curve are generally too broad to pinpoint precisely the temperatures over which a dehydration reaction occurs (see Figure 21.38).

Water retention is one of the key properties of clay mineral surfaces, as already mentioned. This is due in part to the small particle size and large surface area of the clay minerals, as illustrated by the fact that a handful of clay provides enough surface area to cover an area the size of a football field. Among clay minerals, expandable minerals possess greater surface area than nonexpandable minerals, because in the former surfaces both internal and external to the particles contribute to the total surface area. Surface area measurements by N_2 gas adsorption, notably using the Brunnauer, Emmett, and Teller (BET, method (Brunauer et al., 1938), are often applied for evaluating external surface area.

For total surface area (=external + internal) measurements, glycol retention method (Diamond and Kinter, 1958), ethylene glycol monoethyl ether (EGME) method (Carter et al., 1965), a modified EGME method (Eltantawy and Arnold, 1973), and methylene blue method (Hang and Brindley, 1970) are often applied.

Recent developments in analytical instrumentation enable us to explore objects in much smaller domains as they exist without separation, that is, almost in situ observations (e.g., Tessier, 1990). Perhaps, studies of clay minerals in soil aggregates and in the rhizosphere are certainly interesting. Ideas and concepts regarding mineral alteration mechanisms have long been speculated since Keller and Frederickson (1952) followed by Mortland et al. (1956). This appears recently to revive and improve with modern technologies (Robert and Berthelin, 1984; April and Keller, 1990; Hinsinger et al., 1992; Hinsinger and Jaillard, 1993). Approaches from different perspectives were attempted by Kodama et al. (1994) for the role of clay minerals in the rhizosphere and by Monreal and Kodama (1997) for the role of clay minerals as soil aggregates in microbial activity.

Interactions with microbes have received a great deal of attention. Huang and Schnitzer (1986) summarized an achievement of the workshop organized for interactions of soil minerals with natural organics and microbes. Reviews until 1997 can be seen in *Geomicrobiology* (Banfield and Nealson, 1997). Tazaki and Asada (2007) provided transmission electron micrographs showing intimate association where clay particles attached to a thick wall covering the bacteria. They explained that clay particles and the thick wall, a kind of cellular polymeric substance, are able to protect bacteria from toxic elements such as Hg. The research was performed by an appropriate application of energy-filtering transmission electron microscopy, scanning transmission microscopy equipped with energy dispersive x-ray spectroscopy, optical and fluorescence microscopy, XRD, and atomic absorption spectrometry.

21.5.4 Quantification

21.5.4.1 Basic Theory

Alexander and Klug (1948) showed that the XRD intensity (I_i) from the ith component in a multicomponent mixture can be related to the weight fraction of the component (x_i) by the equation,

$$I_i = \frac{K_i(x_i/\rho_i)}{(\Sigma \mu_i^* x_i)} \tag{21.5}$$

where

ρ_i and μ_i^* are the density and mass absorption coefficient of the component i, respectively

K_i depends upon the nature of component i and the geometry of the apparatus employed

By regarding the mixture of n components as if it consisted of just two components, component 1 to be analyzed and the sum of the other components designated as matrix, Equation 21.5 may be extended to a very useful form:

$$I_1 = \frac{K_1 x_1}{\{\rho_1[x_1(\mu_1^* - \mu_M^*) + \mu_M^*]\}} \tag{21.6}$$

where μ_M^* is the mass absorption coefficient of the matrix. Equation 21.6 is the basic relationship underlying quantitative analysis of multicomponent mixtures with the x-ray diffractometer (Alexander and Klug, 1948; Klug and Alexander, 1954). The diffraction intensity of component 1 is thus related to its concentration in the mixture and to relative mass absorption coefficients of the component and matrix. Because of this relationship, except for the case where $\mu_1^* = \mu_M^*$, that is, the case of a mixture consisting of only polymorphic substances, the intensity–concentration relation is generally not linear. Since mass absorption coefficients are functions of the chemical composition of the absorber and of the wavelength of x-rays irradiated, it is readily understood that the choice of radiation to be employed is important. Practical solutions to Equations 21.5 and 21.6 are discussed in the following section.

21.5.4.2 Intensity–Concentration Calibration Curve

For general analytical purposes, one possible solution of Equation 21.6 is to establish intensity–concentration working curves from multicomponent standard synthetic mixtures that simulate a natural system in question. This approach has been attempted for the quantitative estimation of clay minerals by Talvenheimo and White (1952) among others. However, numerous combinations of mineral associations occur in nature, so to cover all possible cases by preparing synthetic standard mixtures is not practical. Thus, this approach is necessarily restricted to simple systems. One of the more practical solutions to the analysis of complicated multicomponent systems may be to compare intensities of two components in a multicomponent mixture and to relate them to their weight proportions. By introducing the average mass absorption coefficient, μ^*, for the mixture, we can simplify Equation 21.5 to

$$I_i = K_i' \left[\frac{x_i}{(\rho_i \mu^*)} \right] \tag{21.7}$$

Then a comparison between intensities of the ith and jth components produces the relation independent of μ^* (Brindley, 1961) as follows:

$$\frac{I_i}{I_j} = \frac{K_i' x_i \rho_j}{(K_j' x_j \rho_i)} = K'' \left(\frac{x_i}{x_j} \right) \tag{21.8}$$

where K'' includes all constants, K_i', K_j', ρ_i, and ρ_j, for two mineral components i and j. Equation 21.8 suggests a linear relationship between the intensity ratio ($I_i{:}I_j$) and the ratio of weight fractions ($x_i{:}x_j$) for the two mineral components. Therefore, if the K'' is evaluated experimentally by synthetic mixtures, the relative weight fractions in unknowns can be determined from measurements of intensity ratios. In a multicomponent system, the weight fraction of each mineral component can be estimated by integration of the results on mineral pairs, which cover all mineral species in the sample. This method was applied to clay minerals by Theisen and Bellis (1964).

21.5.4.3 Internal Standard Method

The use of an internal standard also permits quantitative analysis without any direct knowledge of the mass absorption coefficients of components (μ_i^*), since the method is based on a relation similar to Equation 21.8. Suppose that a known and constant amount of an internal standard, which will be designated by the subscript "s" is added to an unknown mixture. Then the weight fraction of component 1 (x_1) in the original mixture is changed to x_1', which is expressed by the relation, $x_1 = x_1'/(1 - x_S)$. This relation can be derived as follows: since $w_S/x_S = (w + w_S)$, where w and w_S are the total sample weight before the addition of the standard and the weight of the standard added, respectively; thus, $1 - x_S = w/(w + w_S)$. On the other hand, $x_1' = w_1/(w + w_S)$ and $x_1 = w_1/w$, therefore $x_1 = x_1'/(1 - x_S)$. Taking the intensity ratio of I_1 to I_S in a manner similar to Equation 21.8, we obtain the relation

$$\frac{I_1}{I_S} = \frac{K_1 x_1' \rho_S}{K_S x_S \rho_1} = \frac{[K_1' x_1(1 - x_S)\rho_S]}{K_S x_S \rho_1} \qquad (21.9)$$

All components but x_1 in the right-hand term of Equation 21.9 are regarded as constant; Equation 21.9 can, therefore, be arranged to

$$x_1 = k\left(\frac{I_1}{I_S}\right) \qquad (21.10)$$

Thus, the concentration of component 1 is proportional to the intensity ratio $I_1{:}I_S$ alone and independent of mass absorption coefficients of a specific mineral and matrix. This method has been applied to clays and related minerals using internal standards such as fluorite (CaF_2) (Alexander and Klug, 1948), boehmite ($AlOOH$) (Brindley and Kurtossy, 1961), $Zn(OH)_2$ (Mossman et al., 1967), and pyrophyllite (Mossman et al., 1967).

21.5.4.4 Standard Additional Method (Spike Method)

As described by Brindley and Udagawa (1960), this method involves an addition of a known amount of a pure mineral component to a sample in which the mineral component is to be analyzed. The method is useful if the characteristics of the mineral component in the sample closely resemble those of the pure mineral component to be added. If the change in the average

μ^* of Equation 21.7 by this addition can be neglected, then the ratio ($I_1'{:}I_1$) of intensities of component 1 being analyzed before (I_1) and after (I_1') should be directly proportional to the ratio of weight fractions as follows (Brindley and Udagawa, 1960):

$$\frac{I_1'}{I_1} = \frac{(x_1 + x_1')}{[x_1(1 + x_1')]} \qquad (21.11)$$

where the right-hand term is the weight fraction of the component after the addition of the component to the sample. It should be borne in mind that the optimum amount of the pure mineral added depends upon the initial amount of the mineral component in the sample. As a general guide, 25%–100% of the initial amount may be recommended.

21.5.4.5 External Standard Method (Direct Method)

This method requires mass absorption coefficient measurements and has been applied by Williams (1959), Norrish and Taylor (1962), and others. Such measurements are, however, rather time-consuming and the method has only limited application to complicated mineral mixtures in soils.

21.5.4.6 Estimation Based on Intensity Factor Ratios

The diffraction intensity, I_{hkl}, of the hkl reflection is proportional to the square of the structure factor, F_{hkl}, thus, $I_{hkl} \sim KIF_{hkl}I^2$, where K depends upon experimental conditions and the value of F is related to the numbers and kinds of atoms involved in the reflection plane hkl of a mineral structure. Consider a mixture consisting of two minerals A and B with equal amounts and let I_A and I_B be the diffraction intensities from any reflection planes of minerals A and B, respectively. The ratio $I_B{:}I_A$ should be constant regardless of the absolute amounts of both minerals, and this constant factor, k, may be expressed by IF_BI^2/IF_AI^2, where F_A and F_B are the structure factors for the corresponding reflections of minerals A and B, respectively. The I_A is, therefore, equal to I_B/k. If the amounts of minerals A and B are not the same, they would be proportional to $I_A/1$ and I_B/k, respectively. This is the basic concept for the quantification method based on the intensity factor ratios. The concept can be extended easily to multi-component systems.

The preferred orientation method is widely applied to identify clay minerals. It is, therefore, very convenient and practical that the basal reflections are used for quantification. The basal intensity factor ratios may be obtained (a) directly from structure factor calculations or (b) by graphical approximation from squared moduli of layer structure factors prepared by Bradley (1953), MacEwan et al. (1961), and Cole and Lancucki (1966) for various types. By these approaches, Johns et al. (1954) first attempted the quantification of clay minerals and many investigators followed (e.g., Weaver, 1958; Schultz, 1960, 1964; Dell, 1973). Another way to evaluate the factors is by an experimental determination based upon synthetic standard mixtures. The advantage of the latter is that the factors evaluated can include all necessary correction factors due to experimental conditions

as well as the intensity factor. Therefore, the factors can be applied immediately to estimate quantities of clay minerals, as long as the experimental conditions are the same as those for the synthetic standard mixtures from which the factors were evaluated. In general practice for the quantification of clay minerals, intensity factors are determined using a series of binary mineral systems, like minerals *A* and *B,* with equal amounts at several different levels. Similar experiments should be extended to other binary systems like minerals *A* and *C*, or *A* and *D*, and so on. Obviously, the intensity factors of mineral *A* are expected for normalization for the whole. The mineral illite is often chosen for this purpose due to various merits. This approach was chosen by Oinuma et al. (1961). Most clay minerals vary in crystallinity and composition, which greatly affect their basal reflection intensities. Because of this, quantitative estimations made from these intensity factor methods are crude in comparison with those done by methods described in previous sections. In view of difficulties due to variations in clay minerals, however, such an approach is believed to be the best compromise.

21.5.4.7 General Outline for Soil Mineral Quantification

Since soils are complex mixtures of substances of varying degrees of crystallinity ranging from amorphous to strongly crystalline, achieving quantitative analysis with a comparable precision for all mineral components at the same time is almost impossible. Although a major emphasis should be put on clay minerals that usually constitute a large portion of crystalline substances in the clay-size fraction of soils, depending upon the type of minerals involved, a more precise estimation may be made by choosing a suitable method for a specific case. Thus, the general outline adapted in our laboratory for the soil mineral quantification is a combination of methods as follows (Kodama et al., 1977):

1. The method using intensity factor ratios for clay minerals (phyllosilicates) in general
2. The intensity–concentration calibration method for quartz
3. The internal standard method for hematite and goethite
4. A combination of the intensity–concentration and internal standard methods for feldspars and amphiboles
5. The standard addition method for minerals such as calcite, dolomite, gypsum, serpentine minerals, talc, pyrophyllite, and jarosite, which occur less frequently in soils and whose crystallinities do not vary much so that a standard representing the mineral to be analyzed in a sample is easily available

Addendum

After the preparation of this chapter, the author received a copy of the following manuscript: The Mica Group, *In Kluwer Encyclopaedia of Mineralogy* by A.E. Lalonde, 90 pp with nine figures. This appears to be available online.

Appendix

Glossary of Phyllosilicates

Notes:

1. Every valid mineral species names are listed in bold face.
2. Mineral group names are in italic.
3. Majority of interstratified minerals listed here are given in a form of binary system expressed by mineral group names (mica–smectite, vermiculite–smectite, etc.).

A

Aliettite 1:1 regular interlayering of talc–trioctahedral smectite-group mineral, *Interstratified Layer Silicate Group*

Allevardite = **Rectorite**

Alurgite = magnesian manganiferous **Muscovite**

Amesite $Mg_2Al(Si,Al)O_5(OH)_4$, *Serpentine–Kaolin Group*
> Amesite-2*H*
> Amesite-6*R*

Anandite $Ba,K(Fe^{2+},Mg)_3(Si,Al,Fe^{3+})_4O_{10}(O,OH)_2$, *Brittle Mica Group*
> Anandite-2*M*₁
> Anandite-2*Or*

Anauxite = mixture of kaolinite and amorphous silica

Annite $K(Fe^{2+})_3(AlSi_3)O_{10}(OH,F)_2$, *True Mica Group*

Antigorite $(Mg,Fe)_3Si_2O_5(OH)_4$, *Serpentine–Kaolin Group*

Aphrosiderite = ferroan **Clinochlore**

Aspidolite = sodium **Phlogopite**

Attapulgite = **Palygorskite**

B

Baileychlore $(Zn,Fe^{2+},Al,Mg)_6(Si,Al)_4O_{10}(OH)_8$, *Chlorite Group*

Bannisterite $KCa(Fe,Mn,Zn,Mg)_{20}(Si,Al)_{32}O_{76}(OH)_{16} \cdot 4–12H_2O$, *Modulated Layer Silicate Group*

Baumite = zirconian Caryopilite, zirconian **Greenalite**

Bementite $Mn_7Si_6O_{15}(OH)_8$, *Modulated Layer Silicate Group*

Beidellite $(Na,Ca)_{0.3}(Al)_2(Si,Al)O_{10}(OH)_2 \cdot nH_2O$, *Smectite Group*

Bentonite = (a) impure **Montmorillonite**, (b) **Beidellite**

Berthierine $(Fe^{2+},Fe^{3+},Mg)_{2–3}(Si,Al)_2O_5(OH)_4$, *Serpentine–Kaolin Group*
> Berthierine-1*M*

Biotite $K(Mg,Fe^{2+})_3(Al,Fe^{3+})Si_3O_{10}(OH,F)_2$, *True Mica Group*

Bityite $CaLiAl_2(AlBeSi_2)O_{10}(OH)_2$, *Brittle Mica Group*

Bowingite = **Saponite**

Bramallite = sodian **Muscovite (Illite)**

Brandisite = **Clintonite**

Brindleyite $(Ni,Mg,Fe^{2+})_2Al(Si,Al)O_5(OH)_4$, *Serpentine–Kaolin Group*

Brunsvigite = **Chamosite**

C

Cardenite = ferrian **Saponite**

Carlosturanite $(Mg,Fe^{2+},Ti)_{21}(Si,Al)_{12}O_{28}(OH)_{34}$, *Modulated Layer Silicate Group*

Caryopilite $(Mn^{2+},Mg)_3Si_2O_5(OH)_4$, *Modulated Layer Silicate Group*

Celadonite $K(Mg,Fe^{2+})_3(Fe^{3+},Al)Si_4O_{10}(OH)_2$, *True Mica Group*

Cerolite = (a) talc; (b) mixture of *serpentine* and **Stevensite**

Chamosite $(Fe^{2+},Mg,Fe^{3+},Al)_5Al(Si_3Al)O_{10}(OH,O)_8$, *Chlorite Group*
 Chamosite-7Å

Chernykhite $(Ba,Na)(V^{3+},Al)_2(Si,Al)_4O_{10}(OH)_2$, *True Mica Group*

Chlorite   A mineral group, with general structural formula, $(Mg,Fe^{2+},Fe^{3+},Li,Al,Mn,Ni,Zn)_{4-6}(Si,Al,B,Fe^{3+})_4O_{10}(OH,O)_8$

Chlorite–mica   See Mica–chlorite

Chlorite–smectite   Interlayering of chlorite and smectite minerals with various mixing ratio and layer-stacking sequence of the two components, *Interstratified Layer Silicate Group*

Chlorite–talc   See Talc–chlorite

Chlorite–vermiculite   Interlayering of chlorite and vermiculite minerals with various mixing ratio and layer-stacking sequence of the two components, *Interstratified Layer Silicate Group*

Chloropal = **Nontronite**

Chrysocolla $(Cu^{2+},Al)_2H_2Si_2O_5(OH)_4 \cdot nH_2O$, *Serpentine–Kaolin Group*

Chrysolite $Mg_3Si_2O_5(OH)_4$, *Serpentine–Kaolin Group*
 Chrysolite-1M_{cl}
 Chrysolite-2M_{cl}
 Chrysolite-2Or_{cl}

Clinochlore $(Mg,Fe^{2+})_5Al(Si_3Al)O_{10}(OH,O)_8$, forms a series with Chamosite, *Chlorite Group*

Clinochrysotile = **Chrysolite**-1M_{cl}

Clintonite $Ca(Mg,Al)_3(Al_3Si)O_{10}(OH)_2$, *Brittle Mica Group*
 Clintonite-1M
 Clintonite-2M

Cookeite $Li,Al_4(Si_3Al)O_{10}(OH)_8$, *Chlorite Group*

Corrensite   1:1 Regular interlayering of trioctahedral chlorite–smectite or vermiculite-group mineral, with approximate structural formula $(Mg,Fe,Al)_9(Si,Al)_8O_{20}(OH,O)_{10} \cdot nH_2O$, *Interstratified Layer Silicate Group*

Corundphilite = ferroan **Clinochlore**

Cronstedtite $Fe_3^{2+}Fe^{3+}(Si,Fe^{3+})_2O_5(OH)_4$, *Serpentine–Kaolin Group*
 Cronstedtite-1H

D

Daphnite = manganoan **Chamosite**

Delessite = magnesian **Chamosite** or ferrous **Clinochlore**

Deweylite = mixture of **Chrysotile** or **Lizardite** and **Stevensite**

Diabantite = ferroan **Clinochlore**, **Chamosite**

Dickite $Al_2Si_2O_5(OH)_4$, polymorphic relationship with kaolinite and nacrite,

Donbassite $Al_{4-5}(Si_3Al)O_{10}(OH)_8$, *Chlorite Group*

E

Eastonite = (a) **Biotite**, (b) **Vermiculite**, (c) **Phlogopite**

Endellite = (a) Hydrated halloysite, (b) **Halloysite**

Ephesite $NaLiAl_2(Al_2Si_2)O_{10}(OH)_2$, *Brittle Mica Group*

F

Falcondoite $(Ni,Mg)_4Si_6O_{15}(OH)_2 \cdot 6H_2O$, a nickeloan variety of **Sepiolite**, *Sepiolite–Palygorskite Group*

Ferriannite (Ferri-annite) $K(Fe^{2+},Mg)_3(Fe^{3+},Al)Si_3O_{10}(OH,F)_2$, *True Mica Group*

Ferriphlogopite = ferrian **Phlogopite**

Ferripyrophyllite $Fe_2^{3+}Si_4O_{10}(OH)_2$, compare **Pyrophyllite**

Fraipontite $(Zn,Al)_3(Si,Al)_2O_5(OH)_4$, *Serpentine–Kaolin Group*

G

Ganophyllite $(K,Na)_2(Mn,Al,Mg)_8(Si,Al)_{12}O_{29}(OH)_7 \cdot 8-9H_2O$, *Modulated Layer Silicate Group*

Garnierites = general term for hydrous nickel silicates, mainly **Népouite**, **Falcondoite**

Ghassoulite = **Hectorite**

Glauconite $(K,Na)(Fe^{3+},Al,Mg)_2(Si,Al)_4O_{10}(OH)_2$, *True Mica Group*

Gonyerite $(Mn^{2+},Mg)_5Fe^{3+}(Si_3Fe^{3+})O_{10}(OH)_8$, *Modulated Layer Silicate Group*

Greenalite $(Fe^{2+},Fe^{3+})_{2-3}Si_2O_5(OH)_4$, *Modulated Layer Silicate Group*

Griffithite = ferroan **Saponite**

Grovesite = **Pennantite**

H

Halloysite $Al_2Si_2O_5(OH)_4$, *Serpentine–Kaolin Group*
 Halloysite-7 Å $Al_2Si_2O_5(OH)_4$
 Halloysite-10 Å $Al_2Si_2O_5(OH)_4 \cdot 2H_2O$

Halloysite—hydrated halloysite   Interlayering of halloysite-7 Å and halloysite-10 Å, with various mixing ratio and stacking sequence of the two components, *Interstratified Layer Silicate Group*

Hectorite $Na_{0.3}(Mg,Li)_3Si_4O_{10}(F,OH)_2$, *Smectite Group*

Hendricksite $K(Zn,Mg,Mn^{2+})_3(Si_3Al)O_{10}(OH)_2$, *True Mica Group*

Hydrated halloysite = Halloysite-10 Å

Hydrobiotite   1:1 Regular interlayering of biotite and vermiculite, *Interstratified Layer Silicate Group*

Hydromicas   K-poor and H_2O-rich **Muscovite** (illite); altered **Muscovite**

Hydromuscovite = Hydromicas

I

Illite $(K,H_3O)(Al,Mg,Fe)_2(Si,Al)_4O_{10}[(OH)_2,H_2O]$, *True Mica Group*

Illite–chlorite   Compare Mica–chlorite, *Interstratified Layer Silicate Group*

Illite–smectite   Compare Mica–smectite, *Interstratified Layer Silicate Group*

Illite–vermiculite   Compare Mica–vermiculite, *Interstratified Layer Silicate Group*

K

Kämmererite = chromian **Clinochlore**

Kaolin A general term for kaolin minerals, with Serpentine forms *Serpentine–Kaolin Group*

Kaolin–smectite Interlayering of kaolin and smectite minerals with various mixing ratio and various stacking sequence of the two components, *Interstratified Layer Silicate Group*

Kaolinite $Al_2Si_2O_5(OH)_4$, *Serpentine–Kaolin Group*

Kellyite $(Mn^{2+},Mg,Al)_3(Si,Al)_2O_5(OH)_4$, *Serpentine–Kaolin Group*

Kerolite (a) Mixture of serpentine and stevensite; (b) talc

Kinoshitalite $(Ba,K)(Mg,Mn,Al)_3(Si,Al)_4O_{10}(OH,F)_2$, *Brittle Mica Group*

Kotschubeite = ferran **Clinochlore**

Kulkeite $(Na_{0.4}Mg_8Al(Al,Si_7)O_{20}(OH)_{10}$, a 1:1 interlayering of talc and trioctahedral chlorite, *Interstratified Layer Silicate Group*

L

Lepidolite $K(Li,Al)_3(Si,Al)_4O_{10}(F,OH)_2$, *True Mica Group*

Lepidolite-1*M*

Lepidolite-2*M*₁

Lepidolite-2*M*₂

Lepidolite-3*T*

Leptochlorite = iron-rich *Chlorite*, mainly **Chamosite**

Leuchtenbergite = iron-deficient **Clinochlore**

Lizardite $Mg_3Si_2O_5(OH)_4$, *Serpentine–Kaolin Group*

Lizardite-1*T*

Lizardite-2*H*

Lizardite-6*T*

M

Manganopyrosmalite *Modulated Layer Silicate Group*

Margarite $CaAl_2(Al_2Si_2)_4O_{10}(OH)_2$, *Brittle Mica Group*

Margarite-2*M*

Masutomilite $K(Li,Al,Mn^{2+})_3(Si,Al)_4O_{10}(F,OH)_2$, *True Mica Group*

Medomontite = mixture of **Chrysocolla** and mica or hydromica

Metahalloysite = Halloysite-7 Å, dehydrated halloysite

Mica A mineral group, with general structural formula, $(Ba,Ca,Cs,H_3O^+,K,Na,NH_4)(Al,Cr,Fe^{2+},Fe^{3+},Li,Mg,Mn^{2+},Mn^{3+},V,Zn)_{2-3}(Al,Be,Fe^{3+},Si)_4O_{10}(OH,F)_2$

Mica–chlorite Interlayering mica and chlorite minerals with various mixing ratio and layer-stacking sequence of the two components, *Interstratified Layer Silicate Group*

Mica–smectite Interlayering of mica and smectite minerals with various mixing ratio and layer-stacking sequence of the two components, *Interstratified Layer Silicate Group*

Mica–vermiculite Interlayering of mica and vermiculite minerals with various mixing ratio and layer-stacking sequence of the two components, *Interstratified Layer Silicate Group*

Minnesotaite $(Fe^{2+},Mg)_3Si_4O_{10}(OH)_2$, *Modulated Layer Silicate Group*

Montdorite = aluminum-deficient **Biotite**

Montmorillonite $(Na,Ca)_{0.3}(Al,Mg)_2Si_4O_{10}(OH)_2 \cdot nH_2O$, *Smectite Group*

Muscovite $KAl_2(Si_3Al)O_{10}(OH,F)_2$, *True Mica Group*

Muscovite-1*M*

Muscovite-2*M*₁

Muscovite-2*M*₂

Muscovite-3*T*

N

Nacrite $Al_2Si_2O_5(OH)_4$ polymorphic relationship with **Dickite**, **Halloysite**, and **Kaolinite**, *Serpentine–Kaolin Group*

Nanpiningite $Cs(Al,Mg,Fe^{2+},Li)_2(Si_3Al)O_{10}(OH,F)_2$, *True Mica Group*

Népouite $Ni_3Si_2O_5(OH)_4$, *Serpentine–Kaolin Group*

Nimite $(Ni,Mg,Fe^{2+})_5Al(Si_3Al)O_{10}(OH)_8$, *Chlorite Group*

Nontronite $Na_{0.3}Fe_2^{3+}(Si,Al)_4O_{10}(OH)_2 \cdot nH_2O$, *Smectite Group*

Norrishite $KLiMn_2^{3+}Si_4O_{12}$, *True Mica Group*

O

Odinite $(Fe^{3+},Mg,Al,Fe^{2+})_{2.5}(Si,Al)_2O_5(OH)_4$, *Serpentine–Kaolin Group*

Odinite-1*M*

P

Palygorskite $(Mg,Al)_2Si_4O_{10}(OH)$ $4H_2O$, *Sepiolite–Palygorskite Group*

Paragonite $NaAl_2(Si_3Al)O_{10}(OH)_2$, *True Mica Group*

Parsettensite $(K,Na,Ca)(Mn,Al)_7Si_8O_{20}(OH)_8 \cdot 2H_2O$, *Modulated Layer Silicate Group*

Pecoraite $Ni_3Si_2O_5(OH)_4$, dimorphic relationship with népouite, *Serpentine–Kaolin Group*

Pennantite $Mn_5^{2+}Al(Si_3Al)O_{10}(OH)_8$, *Chlorite Group*

Penninite = **Clinochlore**

Phengite $KAl_{1.5}(Mg,Fe^{2+})_{0.5}(Si_{3.5}Al_{0.5})O_{10}(OH,F)_2$, *True Mica Group*

Phlogopite $K(Mg,Fe^{2+})_3Si_3AlO_{10}(F,OH)_2$, *True Mica Group*

Polylithionite $KLi_2AlSi_4AlO_{10}(F,OH)_2$, *True Mica Group*

Preiswerkite $Na(Mg_2Al)(Si_2Al_2)O_{10}(OH)_2$, *True Mica Group*

Prochlorite = ferran Clinochlore

Pseudophite = **Clinochlore**

Pyrophyllite $Al_2Si_4O_{10}(OH)_2$, *Talc–Pyrophyllite Group*

Pyrophyllite-2*M*

Pyrophyllite–smectite Interlayering of pyrophyllite and smectite minerals with various mixing ratio and layer-stacking sequence of the two components, *Interstratified Layer Silicate Group*

R

Rectorite $(Na,Ca)Al_4(Si,Al)_8O_{20}(OH)_4 \cdot 2H_2O$, a 1:1 interlayering of dioctahedral mica and dioctahedral smectite and rectorite may further be distinguished into the following:

Ca–Rectorite = the mica-like layer whose composition is similar to that of margarite

Na–Rectorite = rectorite in which its nonexpandable layer is similar to paragonite

K–Rectorite = the mica-like layer whose composition is similar to that of muscovite

Ripidolite = ferran **Clinochlore**, magnesian **Chamosite**

Roscoelite $K(V^{3+},Al,Mg)_2(Si,Al)_4O_{10}(OH)_2$, *True Mica Group*

S

Saponite $(Ca,Na)_{0.3}(Mg,Fe^{2+})_3(Si,Al)_4O_{10}(OH)_2 \cdot 4H_2O$, _Smectite Group_

Iron saponite = saponite with Fe > Mg in its octahedral sheet

Sauconite $Na_{0.3}Zn_3(Si,Al)_4O_{10}(OH)_2 \cdot 4H_2O$, _Smectite Group_

Sepiolite $Mg_4Si_6O_{15}(OH)_2 \cdot 6H_2O$, _Sepiolite–Palygorskite Group_

Septechlorite = amesite, antigorite, berthierine, chrysotile, cronstedtite, lizardite, etc., trioctahedral kaolinite–serpentine minerals

Sericite = (a) **Muscovite (illite)**; (b) fine-grained mica-like minerals

Serpentine = a group name for minerals with the general formula, $(Mg,Fe,Ni)_3Si_2O_5(OH)_4$

Seybertite = **Clintonite**

Sheridanite = aluminum **Clinochlore**, ferroan **Clinochlore**

Shirozulite $KMn_3^{2+}(Si,Al)_4O_{10}(OH)_2$, _True Mica Group_

Siderophyllite $KFe_2^{2+}Al(Al_2Si_2)_4O_{10}(F,OH)_2$, _True Mica Group_

Smectite represent the smectite-group minerals, with general formula $X_{0.3}Y_{2-3}Z_4O_{10}(OH)_2 \cdot nH_2O$ where X = Ca, Li, Na; Y = Al, Cr, Cu, Fe^{2+}, Fe^{3+}, Li, Mg, Ni, Zn; Z = Al, Si, _Smectite Group_

Smectite–chlorite See Chlorite–smectite

Smectite–illite Compare Mica–smectite

Smectite–kaolin See Kaolin–smectite

Smectite–mica See Mica–smectite

Smectite–pyrophyllite See Pyrophyllite–smectite

Smectite–talc See Talc–smectite

Smectite–vermiculite See Vermiculite–smectite

Stevensite $Ca_{0.15}Mg_3Si_4O_{10}(OH)_2$, _Smectite Group_

Stilpnomelane $K(Fe^{2+},Mg,Fe^{3+})_8(Si,Al)_{12}O_{10}(O,OH)_{27}$, _Modulated Layer Silicate Group_

Sudoite $Mg_2(Al,Fe^{3+})_3Si_3AlO_{10}(OH)_8$, _Chlorite Group_

Swinefordite $(Ca,Na)_{0.3}(Li,Mg)_2(Si,Al)_4O_{10}(OH,F)_2 \cdot 2H_2O$, _Smectite Group_

T

Taeniolite $K(LiMg_2)Si_4O_{10}(F,OH)_2$

Talc $Mg_3Si_4O_{10}(OH)_2$, _Talc–Pyrophyllite Group_

Talc–chlorite Interlayering of talc and chlorite minerals with various mixing ratio and layer-stacking sequence of the two components, _Interstratified Layer Silicate Group_

Talc–saponite Compare Talc–smectite

Talc–smectite Interlayering of talc and smectite minerals with various mixing ratio and layer-stacking sequence of the two components, _Interstratified Layer Silicate Group_

Tarasovite A regular interstratified mineral of illite (I) and smectite (S) with nearly _IIIS_ sequence. The ratio = 0.75:0.25, _Interstratified Layer Silicate Group_

Thuringite = ferrian **Chamosite**

Tobelite $(NH_4,K)Al_2(Si_3Al)O_{10}(OH)_2$, _True Mica Group_

Tosudite A 1:1 interlayering of chlorite and a smectite-group phase, with composition, $Na_{0.5}(Al,Mg)_6(Si,Al)_8O_{18}(OH,F)_{12} \cdot 5H_2O$, _Interstratified Layer Silicate Group_

Trilithionite $K(Li_{1.5}Al_{1.5})(Si_3Al)_4O_{10}(F,OH)_2$, _True Mica Group_

V

Vermiculite A majority of this mineral are trioctahedral and with a sample composition, $(Mg,Ca)_{0.3}(Mg_{2.5}Fe_{0.5})(Al_{1.2}Si_{2.8})O_{10}(OH)_2 \cdot 4H_2O$, _Vermiculite Group_

Vermiculite A group name for minerals with the general formula, $(Mg,Ca)_{0.35}(Mg,Fe,Al)_3(Al,Si)_4O_{10}(OH)_2 \cdot 4H_2O$

Vermiculite–mica See Mica–vermiculite

Vermiculite–chlorite See Chlorite–vermiculite

Vermiculite–smectite Interlayering of vermiculite and smectite minerals with various mixing ratio and layer-stacking sequence of the two components, _Interstratified Layer Silicate Group_

Volkonskoite $Ca_{0.3}(Cr^{3+},Mg)_2(Si,Al)_4O_{10}(OH)_2 \cdot 4H_2O$, _Smectite Group_

W

Wonesite $(Na,K)(Mg,Fe,Al)_6(Si,Al)_8O_{20}(OH,F)_4$, _True Mica Group_

X

Xanthophyllite = **Clintonite**

Z

Zinnwaldite $KLiFe^{2+}Al(AlSi_3)O_{10}(F,OH)_2$, _True Mica Group_

Zussmanite $K(Fe^{2+},Mg,Mn^{2+})_{13}(Si,Al)_{18}O_{42}(OH)_{14}$, _Modulated Layer Silicate Group_

References

Alexander, L., and H.P. Klug. 1948. Basic aspects of X-ray absorption in quantitative diffraction analysis of powder mixtures. Anal. Chem. 20:886–889.

Anthony, J.W., R.A. Bideaux, K.W. Bladh, and M.C. Nichols. 1995. Handbook of mineralogy. Vol. II. Silica, silicates. Parts I and II. Mineral Data Publishing, Tucson, AZ.

April, R., and D. Keller. 1990. Mineralogy of the rhizosphere in forest soils of the eastern United States. Mineralogical studies of the rhizosphere. Biogeochemistry 9:1–18.

Bailey, S.W. 1969. Polytypism of trioctahedral 1:1 layer silicates. Clays Clay Miner. 17:355–371.

Bailey, S.W. 1972. Determination of chlorite compositions by X-ray spacings and intensities. Clays Clay Miner. 20:381–388.

Bailey, S.W. (ed.). 1984. Micas: Reviews in mineralogy. Vol. 13. Mineralogical Society of America, Chelsea, MI.

Bailey, S.W. (ed.). 1988. Hydrous phyllosilicates (exclusive of micas). Reviews in mineralogy. Vol. 19. Mineralogical Society of America, Chelsea, MI.

Bailey, S.W., G.W. Brindley, D.S. Fanning, H. Kodama, and R.T. Martin. 1984. Report of the CMS Nomenclature Committee for 1982 and 1983. Clays Clay Miner. 32:239.

Bailey, S.W., G.W. Brindley, H. Kodama, and R.T. Martin. 1982. Nomenclature for regular interstratifications. Report of the CMS Nomenclature Committee for 1980–1981. Clays Clay Miner. 30:76–78.

Ball, D.F., and P. Beaumont. 1972. Vertical distribution of extractable iron and aluminum in soil profiles from a brown earth–peaty podzol association. J. Soil Sci. 23:298–308.

Banfield, J.F., and K.H. Nealson (eds.). 1997. Geomicrobiology: Interactions between microbes and minerals. Reviews in mineralogy. Vol. 35. Mineralogical Society of America, Washington, DC.

Barnhisel, R.I., and P.M. Bertsch. 1989. Chlorites and hydroxyl-interlayered vermiculite and smectite, p. 729–788. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA Book Series No. 1. SSSA, Madison, WI.

Basset, W.A. 1960. Role of hydroxyl orientation in mica alteration. Bull. Geol. Soc. Am. 71:449–456.

Bear, F.E. (ed.). 1964. Chemistry of the soil. 2nd edn. Reinhold Publishing Corporation, New York.

Biermans, V., and L. Baert. 1977. Selective extraction of the amorphous Al, Fe and Si oxides using an alkaline tiron solution. Clay Miner. 12:127–135.

Blackburn, W.H., and W.H. Dennen. 1997. Encyclopedia of mineral names. The Canadian Mineralogist Special Publication 1. Mineralogical Association of Canada.

Bloss, F.D. 1971. Crystallography and crystal chemistry. Holt, Rinehart and Winston, Inc., New York.

Bradley, W.F. 1945. Diagnostic criteria for clay minerals. Am. Mineral. 30:704–713.

Bradley, W.F. 1953. Analysis of mixed-layer clay minerals. Anal. Chem. 25:727–730.

Bradley, W.F., and R.E. Grim. 1948. Colloid properties of layer silicates. J. Phys. Chem. 52:1404–1413.

Bragg, W.L. 1950. Atomic structure of minerals (2nd print). Cornell University Press, New York.

Brindley, G.W. (ed.). 1951. The X-ray identification and crystal structures of clay minerals. The Mineralogical Society (Clay Minerals Group), London, U.K.

Brindley, G.W. 1961. Quantitative analysis of clay mixtures, p. 489–516. *In* G. Brown (ed.) The X-ray identification and crystal structures of clay minerals. Mineralogical Society, London, U.K.

Brindley, G.W., and G. Brown (eds.). 1980. Crystal structures of clay minerals and their X-ray identification. Mineralogical Society, London, U.K.

Brindley, G.W., and S.S. Kurtossy. 1961. Quantitative determination of kaolinite by X-ray diffraction. Am. Mineral. 46:1205–1215.

Brindley, G.W., and S. Udagawa. 1960. High temperature reactions of clay minerals mixtures and their ceramic properties. I. J. Am. Ceram. Soc. 43:59–65.

Brown, G. 1953. Dioctahedral analogue of vermiculite. Clay Miner. Bull. 2:64–69.

Brown, G. (ed.). 1961. The X-ray identification and crystal structures of clay minerals. Mineralogical Society, London, U.K.

Brunauer, S., P.H. Emmett, and E. Teller. 1938. Adsorption of gases in multi-molecular layers. J. Am. Chem. Soc. 60:309–319.

Burst, J.F. 1969. Diagenesis of Gulf coast clayey sediments and its possible relationship to petroleum migration. AAPG Bull. 53:73–93.

Camazano, M.S., and S.G. Garcia. 1966. Complejos interlaminares de caolinita y halloisita con liquidos polares. An. Edafol. Agrobiol. 25:9–25.

Carter, M.R. (ed.). 1993. Soil sampling and methods of analysis. Canadian Society of Soil Science, Lewis Publishers, Boca Raton, FL.

Carter, D.L., M.D. Heilman, and C.L. Gonzalez. 1965. Ethylene glycol monoethyl ether for determination surface area of silicate minerals. Soil Sci. 100:356–360.

Chao, T.T., and L. Zhou. 1983. Extraction techniques for selective dissolution of amorphous iron oxides from soils and sediments. Soil Sci. Soc. Am. J. 47:225–232.

Chester, R., and M.J. Hughes. 1967. A chemical technique for the separation of ferro-manganese minerals, carbonate minerals and adsorbed race elements from pelagic sediments. Chem. Geol. 2:249–262.

Coey, J.M.D. 1988. Magnetic properties of iron in soil iron oxides and clay minerals, p. 397–466. *In* J.W. Stucki, B.A. Goodman, and U. Schwertmann (eds.) Iron in soils and clay minerals. D. Reidel, Dordrecht, the Netherlands.

Cole, W.F., and C.J. Lancucki. 1966. Tabular data of layer structure factors for clay minerals. Acta Crystallogr. 21:836–838.

De Kimpe, C.R., N. Miles, H. Kodama, and J. Dejou. 1987. Alteration of phlogopite to corrensite at Sharbot Lake, Ontario. Clays Clay Miner. 35:150–158.

Deb, B.C. 1950. The estimation of free iron oxides in soils and clays and their removal. J. Soil Sci. 1:212–220.

Deer, W.A., R.A. Howie, and J. Zussman. 1962. Rock-forming minerals. Vol. 3. Sheet silicates. John Wiley & Sons, New York.

Dell, C.I. 1973. A quantitative mineralogical examination of the clay-size fraction of Lake Superior sediments, p. 413–420. Proc. 16th Conf. Great Lakes Res, International Association of Great Lake Research.

Diamond, S., and E.B. Kinter. 1958. Surface areas of clay minerals as derived from measurements of glycerol retention. Clays Clay Miner. 5:334–347.

Dixon, J.M., and S.B. Weed (eds.). 1989. Minerals in soil environments. 2nd edn. SSSA Book Series No. 1. SSSA, Madison, WI.

Duchaufour, P., and B. Souchier. 1966. Note sur une méthod d'extraction combinée de l'aluminium et fer libre dans sols. Sci. Sol. 1:17–29.

Dunoyer de Segonzac, G. 1970. The transformation of clay minerals during diagenesis and low-grade metamorphism: A review. Sedimentology 15:281–346.

Eltantawy, I.M., and P.W. Arnold. 1973. Reappraisal of ethylene glycol mono-ether (EGME) method for surface area estimations of clays. J. Soil Sci. 24:232–238.

Fanning, D.S., V.Z. Keramidas, and M.A. El-Desoky. 1989. Micas, p. 551–634. *In* J.M. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Farmer, V.C. (ed.). 1974. The infrared spectra of minerals. Monograph No. 4. Mineralogical Society, London, U.K.

Fleet, M.E. 2003. Sheet silicates: Micas. *In* W.A. Deer, R.A. Howie, and J. Zussman (eds.) Rock-forming minerals. 2nd edn. The Geological Society, London, U.K.

Follett, E.A.C., W.J. McHardy, B.D. Michell, and B.F.L. Smith. 1965. Chemical dissolution techniques in the study of soil clays: Part I. Clay Miner. 6:23–34.

Foster, M.D. 1953. Geochemical study of clay minerals. III. The determination of free silica and free alumina in montmorillonites. Geochim. Cosmochim. Acta 3:143–154.

Gard, J.A. (ed.). 1971. The electron-optical investigation of clays. Monograph No 3. Mineralogical Society, London, U.K.

Gatineau, L. 1964. Structure reelle de la muscovite: Repartition des substitution isomorphes. Bull. Soc. Fr. Miner. Crystallogr. 87:321–355.

Ghabru, S.K., A.R. Mermut, and R.J. St. Arnaud. 1989. Layer-charge and cation-exchange characteristics of vermiculite (weathered biotite) isolated from gray luvisols in northeastern Saskatchewan. Clays Clay Miner. 37:164–172.

Glaeser, R. 1958. Détection de la démixion des cations Na, Ca dans une hectorite bi-Ioniques. C.R. Acad. Sci., Paris. 246:2909–2912.

Glaeser, R., and J. Méring. 1954. Isothermes d'hydration des montmorillonites bi-ioniques (Na, Ca). Clay Miner. Bull. 2:188–193.

Graf von Reichenbach, H., and C.I. Rich. 1975. Fine-grained micas in soils, p. 59–95. *In* J.E. Gieseking (ed.) Soil components. Springer-Verlag, Berlin, Germany.

Greene-Kelly, R. 1953. The identification of montmorillonoids in clays. J. Soil Sci. 4:233–237.

Greene-Kelly, R. 1955. Sorption of aromatic organic compounds by montmorillonite. Trans. Faraday Soc. 51:425–430.

Greenland, D.J., and M.H.B. Hayes (eds.) 1978. The chemistry of soil constituents. John Wiley & Sons, New York.

Grim, R.E. 1968. Clay mineralogy. McGraw-Hill Book Co. Inc., New York.

Grim, R.E., and H. Kodama. 1997. Clay minerals, p. 207–215. *In* Minerals and rocks. Encyclopaedia Britannica. Vol. 24.

Guggenheim, S., and S.W. Bailey. 1982. The superlattice of minnesotaite. Can. Mineral. 20:579–584.

Guggenheim, S., S.W. Bailey, R.A. Eggleton, and P. Wilkes. 1982. Structural aspects of greenalite and related minerals. Can. Mineral. 20:1–18.

Guggenheim, S., and R.A. Eggleton. 1988. Crystal chemistry, classification, and identification of modulated layer silicates, p. 675–725. *In* S.W. Bailey (ed.) Hydrous phyllosilicates (exclusive of micas). Reviews in mineralogy. Vol. 19. Mineralogical Society of America, Chelsea, MI.

Hang, P.T., and G.W. Brindley. 1970. Methylene blue absorption by clay minerals. Determination of surface areas and cation exchange capacities (clay-organic study XVIII). Clays Clay Miner. 18:203–212.

Hashimoto, I., and M.L. Jackson. 1960. Rapid dissolution of allophane and kaolinite–halloysite after dehydration. Clays Clay Miner. 7:102–113.

Herrero, C.P., J. Sanz, and J.M. Serratosa. 1985. Si, Al distribution in micas: Analysis by high-resolution ^{29}Si NMR spectroscopy. J. Phys. C: Solid State Phys. 53:151–154.

Higashi, S. 1982. Tobelite, a new ammonium dioctahedral mica. Mineral. J. 11:138–146.

Hinsinger, P., and B. Jaillard. 1993. Root-induced release of interlayer potassium and vermiculitization of phlogopite as related potassium depletion in the rhizosphere of ryegrass. J. Soil Sci. 44:525–534.

Hinsinger, P., B. Jaillard, and J.E. Dufey. 1992. Rapid weathering of a trioctahedral mica by the roots of ryegrass. Soil Sci. Soc. Am. J. 56:977–982.

Hofmann, U., and R. Klemen. 1950. Verlust der Austauschfahigkeit von Lithiumionen An Bentonit durch Erhizung. Z. Anorg. Allg. Chem. 262:95–99.

Huang, P.M., and M. Schnitzer (eds.). 1986. Interactions of soil minerals with natural organics and microbes. SSSA Special Publication No. 17. SSSA, Madison, WI.

Jackson, M.L. 1967. Soil chemical analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Jeffries, C.D. 1946. A rapid method for removal of free iron oxides in soils prior to petrographic analysis. Soil Sci. Soc. Am. Proc. 11:211–212.

Johns, W.D., R.E. Grim., and W.F. Bradley. 1954. Quantitative estimations of clay minerals by diffraction methods. J. Sediment. Petrol. 24:242–251.

Kakinoki, J., and Y. Komura. 1952. Intensity of X-ray diffraction by an one-dimensional disordered crystal. (1) General derivation in cases of the "Reichweite" s = 0 and 1. J. Phys. Soc. Jpn. 7:30–35.

Kakinoki, J., and Y. Komura. 1954. Intensity of X-ray diffraction by an one-dimensionally disordered crystal. II: General derivation in the case of the correlation range s ≥ 2. J. Phys. Soc. Jpn. 9:169–183.

Keller, W.D., and A.F. Frederickson. 1952. Role of plants and colloidal acids in the mechanism of weathering. Am. J. Sci. 250:594–608.

Kittrick, J.A. (ed.). 1985. Mineral classification of soils. SSSA Special Publication No. 16. SSSA, Madison, WI.

Klein, C., and C.S. Hurlbut, Jr., 1985. Manual of mineralogy. John Wiley & Sons, Inc., New York.

Klug, H.P., and L.E. Alexander. 1954. X-ray diffraction procedures for polycrystalline and amorphous materials. John Wiley & Sons, Inc., New York.

Klute, A. (ed.) 1986. Methods of soil analysis. Part I. Physical and mineralogical methods. 2nd edn. SSSA Book Series No. 5. ASA–SSSA, Madison, WI.

Kodama, H. 1990. Use of color-coded transparencies for visualizing layer silicate structures, p. 169–175. *In* V.C. Farmer and Y. Tardy (eds.) Proc. 9th Int. Clay Conf. 1989. Strasbourg. Sci. Géol., Mém. Strasbourg, France.

Kodama, H., and J.E. Brydon. 1968a. A study of clay minerals in podzol soils in New Brunswick, eastern Canada. Clay Miner. 7:295–309.

Kodama, H., and J.E. Brydon. 1968b. Dehydroxylation of microcrystalline muscovite. Trans. Faraday Soc. 64:3112–3119.

Kodama, H., and R.S. Dean. 1980. Illite from Eldorado, Saskatchewan. Can. Mineral. 18:109–118.

Kodama, H., L. Gatineau, and J. Méring. 1971. An analysis of X-ray diffraction line profiles of microcrystalline muscovites. Clays Clay Miner. 19:405–413.

Kodama, H., S. Nelson, A.-F. Yang, and N. Kohyama. 1994. Mineralogy of rhizospheric and non-rhizospheric soils in corn fields. Clays Clay Miner. 42:755–763.

Kodama, H., and G.J. Ross. 1991. Tiron dissolution method to remove and characterize inorganic components in soils. Soil Sci. Soc. Am. J. 55:1180–1187.

Kodama, H., G.C. Scott, and N.M. Miles. 1977. X-ray quantitative analysis of minerals in soils. Soil Research Institute. Tech. Bull. Agriculture Canada, Ottawa, Canada.

Kubler, B. 1966. La cristallinete de l'illite et les zones tout a fait superieures du metamorphisme, p. 105–122. Etages tectoniques Colloque de Neuchatel 1966. Institut de Geologie de l'Universite de Neuchatel.

Kunze, G. 1956. Die gewellte Struktur des Antigorits. I. Zeit. Krist. 108:82–107.

Lagaly, G. 1979. The layer charge of regular interstratified 2:1 clay minerals. Clays Clay Miner. 27:1–10.

Lagaly, G. 1982. Layer charge heterogeneity in vermiculites. Clays Clay Miner. 30:215–222.

Lagaly, G., and A. Weiss. 1969. Determination of the layer charge in mica-type layer silicates, p. 61–80. Proc. 3rd Int. Clay Conf. Vol. 1. Tokyo, Japan, September 5–9, 1969.

Liebau, F. 1985. Structural chemistry of silicates: Structure, bonding, and classification. Springer-Verlag, New York.

Loewenstein, W. 1954. The distribution of aluminum in the tetrahedral of silicates and aluminates. Am. Mineral. 39:92–96.

MacEwan, D.M.C., A. Ruiz Amil, and G. Brown. 1961. Interstratified clay minerals, p. 393–445. *In* G. Brown (ed.) The X-ray identification and crystal structures of clay minerals. Mineralogical Society (Clay Minerals Group), London, U.K.

MacEwan, D.M.C., and M.J. Wilson. 1980. Interlayer and intercalation complexes of clay minerals, p. 197–248. *In* G.W. Brindley and G. Brown (eds.) Crystal structures of clay minerals and their X-ray identification. Mineralogical Society, London, U.K.

Mathieson, A.McL., and G.F. Walker. 1954. Crystal structure of magnesium–vermiculite. Am. Mineral. 39:231–255.

Matsuda, T., H. Kodama, and A.-F. Yang. 1997. Ca-rectorite from Sano mine, Nagano prefecture, Japan. Clays Clay Miner. 45:773–780.

McKeague, J.A. 1967. An evaluation of 0.1 M pyrophosphate and pyrophosphate–dithionite in comparison with oxalate as extractants of the accumulation products in podzols and some other soils. Can. J. Soil Sci. 47:95–99.

McKeague, J.A. (ed.). 1976. Manual on soil sampling and methods of analysis. Soil Research Institute, Agriculture Canada, Ottawa, Canada.

Mehra, Q.P., and M.L. Jackson. 1960. Iron oxide removal from soils and clays by a dithionite–citrate system with sodium bicarbonate buffer. Clays Clay Miner. 7:317–327.

Méring, J. 1949. L'interférence des rayons X dans les systèmes a stratification désordonnée. Acta Crystallogr. 2:371–377.

Méring, J. 1975. Smectites, p. 98–120. *In* J.E. Gieseking (ed.) Soil components. Vol. II. Inorganic components. Springer-Verlag, New York.

Méring, J., and G.W. Brindley. 1967. X-ray diffraction band profiles of montmorillonite—Influence of hydration and exchangeable cations. Clays Clay Miner. 15:51–60.

Méring, J., and A. Oberlin. 1967. Electron-optical study of smectites. Clays Clay Miner. 15:3–34.

Méring, J., and A. Oberlin. 1971. The smectites, p. 193–229. *In* J.A. Gard (ed.) Electron-optical investigation of clays. Mineralogical Society, London, U.K.

Mitchell, B.D., and V.C. Farmer. 1962. Amorphous clay minerals in some Scottish soil profiles. Clay Miner. Bull. 5:128–144.

Monreal, C.M., and H. Kodama.1997. Influence of aggregate architecture and minerals on living habitats. Can J. Soil Sci. 77:367–377.

Mooney, R.W., A.G. Keenan, and L.A. Wood. 1952. Adsorption of water vapour by montmorillonite II. Effect of exchange ions and lattice swelling as measured by X-ray diffraction. J. Am. Chem. Soc. 74:1371–1374.

Moore, D.M., and R.C. Reynolds, Jr. 1989. X-ray diffraction and the identification and analysis of clay minerals. Oxford University Press, New York.

Mortland, M.M., K. Lawton, and G. Uehara. 1956. Alteration of biotite to vermiculite by plant growth. Soil Sci. 82:477–481.

Mossman, M.H., D.H. Freas, and S.W. Bailey. 1967. Orienting internal standard method for clay mineral X-ray analysis. Clays Clay Miner. 15:441–453.

Murad, E. 1988. Properties and behavior of iron oxides as determined by Mössbauer spectroscopy, p. 309–350. *In* J.W. Stucki, B.A. Goodman, and U. Schwertmann (eds.) Iron in soils and clay minerals. D. Reidel, Dordrecht, the Netherlands.

Murad, E., and J. Cashion. 2004. Mössbauer spectroscopy of environmental materials and their industrial utilization. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Murad, E., and U. Wagner. 1991. Mössbauer spectra of kaolinite, halloysite and the firing products of kaolinite: New results and a reappraisal of published work. Neues Jahrb. Mineral. Abh. 162:281–309.

Murad, E., and U. Wagner. 1994. The Mössbauer spectrum of illite. Clay Miner. 29:1–10.

Nadeau, P.H. 1985. The physical dimension of fundamental clay particles. Clay Miner. 20:499–514.

Nadeau, P.H., M.J. Wilson, W.J. McHardy, and J.M. Tait. 1984a. Interstratified clays as fundamental particles. Science 225:923–925.

Nadeau, P.H., M.J. Wilson, W.J. McHardy, and J.M. Tait. 1984b. Interparticle diffraction: A new concept for interstratified clays. Clay Miner. 19:757–769.

Nakamoto, K. 1963. Infrared spectra of inorganic and coordination compounds. John Wiley & Sons, Inc., New York.

Newman, A.C.D. (ed.). 1987. Chemistry of clays and clay minerals. Mineralogical Society, London, U.K.

Norrish, K., and R.M. Taylor. 1962. Quantitative analysis by X-ray diffraction. Clay Miner. Bull. 5:98–109.

Oberlin, A., and J. Méring. 1962. Observations en microscopie et microdiffraction electronique sur la montmorillonite Na. J. Microscopie. 1:107–120.

Oinuma, K., K. Kobayashi, and T. Sudo. 1961. Procedure of clay mineral analysis. Clay Sci. 3:179–193.

Pauling, L. 1929. The principles determining the structure of complex ionic crystals. J. Am. Chem. Soc. 51:1010–1026.

Petruk, W. 1964. Determination of the heavy atom content in chlorite by means of the X-ray diffractometer. Am. Mineral. 49:61–71.

Reynolds, R.C., Jr., 1980. Interstratified clay minerals, p. 249–303. *In* G.W. Brindley and G. Brown (eds.) Crystal structures of clay minerals and their X-ray identification. Mineralogical Society, London, U.K.

Robert, M., and J. Berthelin. 1984. Role of biological and biochemical factors in soil mineral weathering, p. 453–495. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and microbes. SSSA Special Publication No. 17. SSSA, Madison, WI.

Robert, J.-L., and H. Kodama. 1988. Generalization of the correlations between hydroxyl stretching wave numbers and composition of mica in the system $K_2O–MgO–Al_2O_3–SiO_2–H_2O$: A single model for trioctahedral and dioctahedral micas. Am. J. Sci. 288:196–212.

Ross, G.J. 1978. Relationships of specific surface area and clay content to shrink-swell potential of soils having different clay mineralogical compositions. Can J. Soil Sci. 58:159–166.

Ross, G.J., H. Kodama, C. Wang, J.T. Gray, and L.B. Lafreniere. 1983. Halloysite from a strongly weathered soil at Mont Jacques Cartier, Quebec. Soil Sci. Soc. Am. J. 47:327–332.

Ross, G.J., C. Wang, and P.A. Schuppli. 1985. Hydroxylamine and ammonium oxalate solution as extractants for iron and aluminum from soils. Soil Sci. Soc. Am. J. 49:783–785.

Sans, J., and J.M. Serratosa. 1984. ^{29}Si and ^{27}Al high-resolution MAS–NMR spectra of phyllosilicates. J. Am. Chem. Soc. 106:4790–4793.

Sato, M. 1965. Structure of interstratified (mixed-layer) minerals. Nature 208:70–71.

Sato, M. 1973. X-ray analysis of interstratified structure. J. Clay Sci. Soc. Jpn. 13:39–47.

Schultz, L.G. 1960. Quantitative X-ray determination of some aluminous clay minerals in rocks. Clays Clay Miner. 7: 216–224.

Schultz, L.G. 1964. Quantitative interpretation of mineralogical composition from X-ray and chemical data for the Pierre shale. U.S. Geological Survey Professional Paper 391-C.

Schwertmann, U. 1959. Die fraktionierte Extraktion der freien Eisenoxide in Boden, ihre Mineralogischen Formen und ihre Entstehungsweisen. Z. Pflanzenernahr. Dung. Bodenkd. 84:194–204.

Segalen, P. 1968. Note sur une méthod de détermination des produits minéraux amorphes dans certains sols a hydroxides tropicaux. Cah. ORSTOM Ser. Pedol. 6:105–126.

Serratosa, J.M., and W.F. Bradley. 1958. Determination of the orientation of the OH bond axes in layer silicates by infrared absorption. J. Phys. Chem. 62:1164–1167.

Shirozu, H., and S.W. Bailey. 1965. Chlorite polytypism: III. Crystal structure of an ortho-hexagonal iron chlorite. Am. Mineral. 50:868–885.

Smith, J.V., and H.S. Yoder. 1956. Experimental and theoretical studies of mica polymorphs. Mineral. Mag. 31:209–235.

Stucki, J.W., B.A. Goodman, and U. Schwertmann. 1988. Iron in soils and clay minerals. NATO ASI series. Series C: Mathematical and physical sciences. Vol. 217. D. Reidel, Dordrecht, the Netherlands.

Sudo, T. 1959. Mineralogical study on clays of Japan. Maruzen Co. Ltd., Tokyo, Japan.

Sudo, T., S. Shimoda, H. Yotsumoto, and S. Aita. 1981. Electron micrographs of clay minerals. Developments in sedimentology. Vol. 31. Elsevier Scientific Publishing Co., New York.

Talvenheimo, G., and J.L. White. 1952. Quantitative analysis of clay minerals with the X-ray spectrometer. Anal. Chem. 24:1784–1789.

Tamm, O. 1922. Um bestamning av de oorganiska komponenterna i markens gelkomplex. Medd. Statens Skogsfoersoeksansalt. 19:385–404.

Tazaki, K., and R. Asada. 2007. Transmission electron microscopic observation of mercury-bearing bacterial clay minerals in a small-scale gold mine in Tanzania. Geomicrobiol. J. 24:477–489.

Tessier, D. 1990. Behaviour and microstructure of clay mineral, p. 387–415. *In* M. De Boodt, M. Hayes, and A. Herbillon (eds.) Soil colloids and their association in aggregates. Plenum Publishing Corporation, New York.

Theisen, A.A., and E. Bellis. 1964. Quantitative analysis of clay mineral mixtures by X-ray diffraction. Nature 204:1228–1230.

Thorez, J. 1975. Phyllosilicates and clay minerals: A laboratory handbook for their X-ray diffraction analysis. G. Lelotte, Belgique, Belgium.

Tuddenham, W.M., and R.J.P. Lyon. 1959. Relation of infrared spectra and chemical analysis for some chlorites and related minerals. Anal. Chem. 31:377–380.

Vali, H., and H.M. Koster. 1986. Expanding behaviour, structural disorder, regular and random irregular interstratifications of 2:1 layer-silicates studied by high-resolution images of transmission electron microscopy. Clay Miner. 21:827–859.

Vanders, I., and P. Kerr. 1967. Mineral recognition. John Wiley & Sons, Inc., New York.

Vedder, W. 1964. Correlations between infrared spectrum and chemical composition of mica. Am. Mineral. 49:736–768.

Vedder, W., and R.S. McDonald. 1963. Vibrations of the OH ions in muscovite. J. Chem. Phys. 38:1583–1590.

Velde, B. 1992. Introduction to clay minerals. Chapman and Hall, London, U.K.

Wada, K. 1959. Oriented penetration of ionic compounds between the silicate layers of halloysite. Am. Mineral. 44:153–165.

Wada, K. 1961. Lattice expansion of kaolin minerals by treatment with potassium acetate. Am. Mineral. 44:1237–1247.

Wada, K., and D.J. Greenland. 1970. Selective dissolution and differential infrared spectroscopy for characterization of 'amorphous' constituents in soil clays. Clay Miner. 8:241–254.

Wada, K., and C. Mizota. 1982. Iron-rich halloysite (10Å) with crumped lamellar morphology from Hokkaido, Japan. Clays Clay Miner. 30:315–317.

Walker, G.F. 1949. Distinction of vermiculite, chlorite and montmorillonite in clays. Nature 164:577–578.

Wang, C., H. Kodama, and N.M. Miles. 1981. Effect of various pretreatments on X-ray diffraction patterns of clay fractions of podzolic B horizons. Can. J. Soil Sci. 61:311–316.

Weaver, C.E. 1958. Geologic interpretation of argillaceous sediments. Part I. Origin and significance of clay minerals in sedimentary rocks. Am. Assoc. Pet. Geol. Bull. 42:254–271.

Weaver, C.E. 1989. Clays, muds and shales. Developments in sedimentology. Vol. 44. Elsevier Scientific Publishing Co., New York.

Weiss, A., W. Thielepape, G. Gorning, W. Ritter, and H. Schafer. 1963. Kaolinite intercalation compounds, p. 287–305. Proc. Int. Clay Conf. Vol. 1. Stockholm, Sweden.

Wicks, F.J., and E.J.W. Whittaker. 1975. A reappraisal of the structure of the serpentine minerals. Can Mineral. 13:227–243.

Wicks, F.J., and J. Zussman. 1975. Microbeam X-ray diffraction patterns of the serpentine minerals. Can Mineral. 13:244–258.

Williams, P.P. 1959. Direct quantitative diffractometric analysis. Anal. Chem. 31:1842–1844.

Wilson, M.J. (ed.). 1992. Clay mineralogy: Spectroscopic and chemical determinative methods, 367 pp. Chapman and Hall, London, U.K.

Wilson, M.J., and D.R. Clark. 1978. X-ray identification of clay minerals in thin sections. J. Sediment. Petrol. 48:656–660.

Yada, K. 1967. Study of chrysotile asbestos by a high resolution electron microscope. Acta Crystallogr. 23:704–707.

Yong, R.N., and B.P. Warkentin. 1975. Soil properties and behaviour. Developments in geotechnical engineering 5. Elsevier Scientific Publishing Co., Amsterdam, the Netherlands.

Yuan, H., and D.L. Bish. 2010. NEWMOD[+], a new version of the NEWMOD program for interpreting X-ray powder diffraction patterns from interstratified clay minerals. Clays Clay Miner. 58:318–326.

22

Oxide Minerals in Soils

Nestor Kämpf
*Universidade Federal do
Rio Grande do Sul*

Andreas C. Scheinost
Institute of Radiochemistry

Darrell G. Schulze
Purdue University

22.1 Introduction

Oxide minerals or oxides comprise the oxides, hydroxides, oxyhydroxides, and hydrated oxides of Fe, Mn, Al, Si, and Ti. They commonly occur in soils, particularly those in advanced stages of weathering where they can sometimes account for as much as 50% of the total soil mass. In the discussion to follow, a polyhedral approach will be used to introduce the mineral species and highlight their major differences and similarities (Figure 22.1). For additional structural details, the reader is referred to the literature cited. Thereafter, their occurrence and formation in soils and their influence on soil properties will be discussed, concluding with the major techniques for identifying these minerals in soils.

22.2 Iron Oxides

Iron oxide minerals are ubiquitous in soils and sediments. All of the Fe oxide minerals are strongly colored and, when finely dispersed throughout the matrix, even small amounts impart bright colors to soils. Iron oxides often cement other soil minerals into stable aggregates, and when they reach massive proportions, such cementations are called laterite or ferricrete. Because iron oxide surfaces have a strong affinity for the oxyanions of phosphorus and some transition metals, they play a significant role in the environmental cycling of these elements. A list of the Fe oxides is given in Table 22.1. Recent comprehensive reviews include Schwertmann and Taylor (1989), Cornell and Schwertmann (1996, 2003), and Bigham et al. (2002).

22.2.1 Mineral Phases

The structures of most Fe oxides can be described in terms of close-packed arrays of O and OH ions, with Fe occupying interstitial octahedral sites. Structures based on hexagonal close packing (hcp) of the anions are called α-phase, and those with cubic close packing (ccp), γ-phase (note that this convention is not necessarily followed for oxides of other elements). Both hcp and ccp structures also contain tetrahedral interstices. In two minerals, magnetite and maghemite, Fe is present in some of these tetrahedral sites as well. The basic structural unit for most of the Fe oxides, however, is an octahedron in which each Fe atom is surrounded either by six O or by both O and OH ions. The various Fe oxides differ mainly in the arrangement of these octahedra and in how they are linked.

22.2.1.1 Oxyhydroxides

Goethite (α-FeOOH) consists of double chains of edge-shared octahedra that are joined to other double chains by sharing corners and by hydrogen bonds (Figure 22.1; Cornell and Schwertmann, 1996). *Lepidocrocite* (γ-FeOOH) also contains double chains of octahedra, but they are joined by shared edges, resulting in corrugated sheets of octahedra. These corrugated sheets are stacked one on top of the other and are held together by hydrogen bonds (Figure 22.1; Ewing, 1935a, 1935b; Fasiska, 1967). *Akaganéite* (β-FeOOH) consists of double chains that share corners with adjacent chains to give a 3D structure containing tunnels 0.5 nm in cross section (Figure 22.1). These tunnels contain Cl$^-$ ions that stabilize the structure (Post and Buchwald, 1991). *Schwertmannite* ($Fe_8O_8(OH)_6(SO_4)$) is

Mn: Pyrolusite, MnO_2
     Manganite, $\gamma\text{-}MnOOH$
Ti: Rutile, TiO_2

Fe: Goethite, $\alpha\text{-}FeOOH$
Mn: Ramsdellite, MnO_2
     Groutite, $MnOOH$
Al: Diaspore, $\alpha\text{-}AlOOH$

Fe: Lepidocrocite, $\gamma\text{-}FeOOH$
Al: Boehmite, $\gamma\text{-}AlOOH$

Fe: Akaganéite, $\beta\text{-}FeOOH$
     Schwertmannite, $Fe_8O_8(OH)_6(SO_4)$
Mn: Hollandite, $Ba_x(Mn^{IV}Mn^{III})_8O_{16}$, $x = 1$
     Cryptomelane, $K_x(Mn^{IV}Mn^{III})_8O_{16}$, $x = 1.3\text{--}1.5$
     Coronadite, $Pb_x(Mn^{IV}Mn^{III})_8O_{16}$, $x = 1.0\text{--}1.4$

Mn: Romanechite, $Ba_{0.66}(Mn^{IV}Mn^{III})_5O_{10} \cdot 1.3H_2O$

Mn: Todorokite,
     $(Na,Ca,K)_{0.3\text{--}0.5}(Mn^{IV}Mn^{III})_6O_{12} \cdot 3.5H_2O$

Fe: Feroxyhyte, $\delta'\text{-}FeOOH$ (A = B = Fe)
Mn: Buserite (A = B = Mn)
     Lithiophorite, $(Li,Al)(Mn^{III},Mn^{IV})O_2(OH)_2$
     (A = Mn,Co; B = Al,Li,Ni,Cu,Zn)
     Asbolane, $(Co,Li)MnO_2(OH)_2$
     (A = Mn; B = Co,Ni)
Al: Gibbsite, bayerite, nordstrandite, $Al(OH)_3$
     (A = B = Al)

Fe: Green rust, $Fe_4^{II} Fe_2^{III}(OH)_{12}(Cl,SO_4,CO_3)$

FIGURE 22.1 **(See color insert.)** Structural schemes for oxide minerals in soils.

Mn: Chalcophanite, $ZnMn_3O_7 \cdot 3H_2O$
Birnessite, $(Na,Ca,Mn^{II}) Mn_7O_4 \cdot 2.8H_2O$

Si: α-Cristobalite, SiO_2

Fe: Hematite, $\alpha\text{-}Fe_2O_3$
Al: Corundum, $\alpha\text{-}Al_2O_3$
Ti,Fe: Ilmenite, $FeTiO_3$

Ti: Anatase, TiO_2

Fe: Magnetite, Fe_3O_4
Maghemite, $\gamma\text{-}Fe_2O_3$
Mn: Hausmannite, Mn_3O_4
Ti,Fe: Titanomagnetite, $Fe_{3-x}Ti_xO_4$
Titanomaghemite, $Fe_{2-x}Ti_xO_3$

Ti: Brookite, TiO_2

Ti,Fe: Pseudobrookite, $Fe_2^{III}TiO_5$

Ti,Fe: Pseudorutile, $Fe_2^{III}Ti_3O_8(OH)_2$

Si: α-Quartz, SiO_2
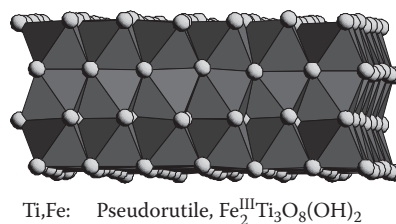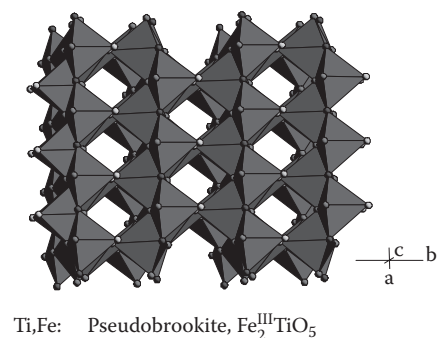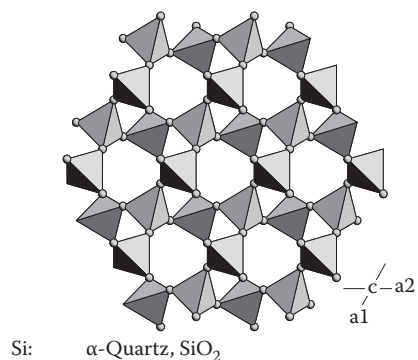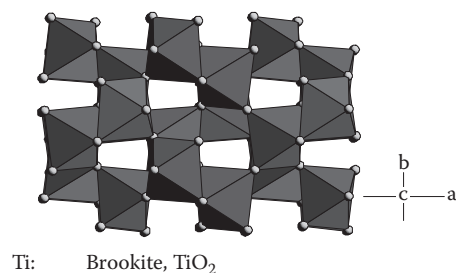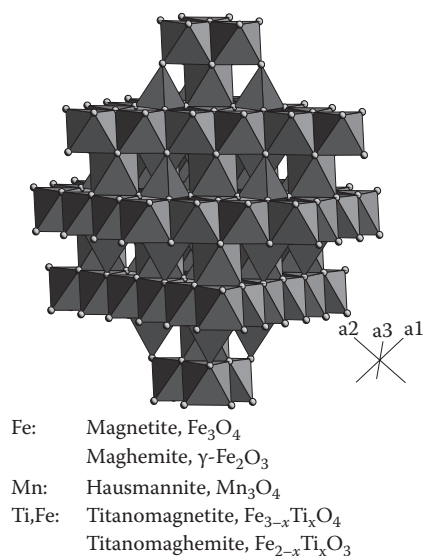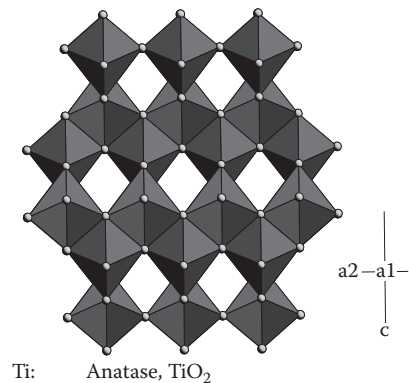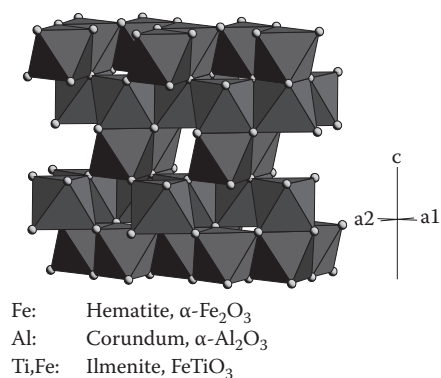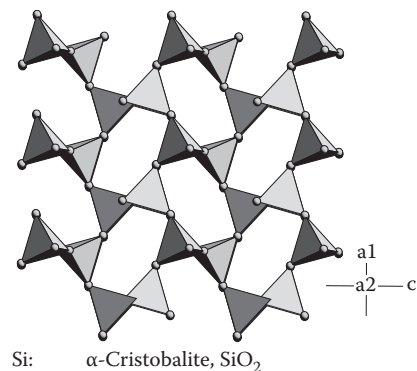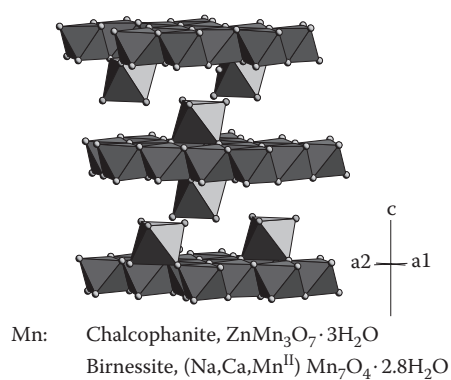
FIGURE 22.1 (continued)

TABLE 22.1 Iron Oxide Minerals: Crystallographic Properties and d Values for the Six Most Intense Diffraction Lines

Mineral	Chemical Formula	Crystal System, Space Group	Unit-Cell Dimensions (Å)	Six Most Intense Diffraction Lines d Value (Å), Relative Intensity						References[a]
Goethite	α-FeOOH	Orthorhombic, *Pbnm*	a = 4.608 b = 9.956 c = 3.021	4.183 100	2.450 50	2.693 35	1.719 20	2.190 18	2.253 14	29-0713
Lepidocrocite	γ-FeOOH	Orthorhombic, *Cmcm*	a = 3.07 b = 12.53 c = 3.88	6.260 100	3.290 90	2.470 80	1.937 70	1.732 40	1.524 40	Cornell and Schwertmann (2003); 08-0098
Akaganéite	β-FeOOH	Monoclinic, *I2/m*	a = 10.56 b = 3.031 c = 10.483 β = 90°63′	3.333 100	2.550 5	7.467 40	2.295 35	1.643 35	5.276 30	Post and Buchwald (1991); 34-1266
Schwertmannite	$Fe_8O_8(OH)_6SO_4$	Tetragonal, *P4/m*	a = 10.66 c = 6.04	2.55 100	3.39 46	4.86 37	1.51 24	2.28 23	1.66 21	Bigham et al. (1994)
Feroxyhite	δ'-FeOOH	Hexagonal, *P3ml*	a = 2.93 c = 4.56	2.545 100	2.255 100	1.685 100	1.471 100	4.610 20	1.271 20	Cornell and Schwertmann (1996); 13-0087
Ferrihydrite	$Fe_5HO_8 \cdot 4H_2O$	Hexagonal	a = 5.08 c = 9.4	2.50 100	2.21 80	1.96 80	1.48 80	1.51 70	1.72 50	Towe and Bradley (1967); 29-0712
Hematite	α-Fe_2O_3	Hexagonal, *R3c*	a = 5.034 c = 13.752	2.700 100	2.519 70	1.694 45	1.840 40	3.684 30	1.486 30	33-0664
Magnetite	Fe_3O_4	Cubic, *Fd3m*	a = 8.3967	2.532 100	1.484 40	2.967 30	1.616 30	2.099 20	1.093 12	19-0629
Maghemite	γ-Fe_2O_3	Cubic, *Fd3m*	a = 8.35	2.518 100	2.953 40	1.476 30	1.607 20	2.088 15	1.704 10	Goss (1988); 39-1346
Bernalite	$Fe(OH)_3$	Orthorhombic, *Immm*	a = 7.544 b = 7.560 c = 7.558	3.784 100	1.692 17	2.393 16	2.676 15	1.892 10	1.545 9	Birch et al. (1993)
Fougerite	$(Fe^{2+},Mg)_6$ $Fe_2^{3+}(OH)_{18} \cdot 4H_2O$	Trigonal–hexagonal, *R3m*	a = 3.125 c = 22.5	7.97 100	2.692 34	3.97 32	2.027 19	1.563 10	1.595 9	http://webmineral.com
$Fe(OH)_2$	$Fe(OH)_2$	Hexagonal, *P3ml*	a = 3.262 c = 4.596	4.597 vs[b]	2.403 vs	2.817 s	1.782 s	1.629 s	1.535 w	Miyamoto (1976)
Green rust I	$Fe(OH)_2Fe(OH)_3Cl$ (variable)	Rhombohedral	a = 3.198 c = 24.21	8.02 vs	4.01 s	2.701 m	2.408 m	2.037 w	1.487 w	Bernal et al. (1959)
Green rust II	$Fe(OH)_2Fe(OH)_3SO_4$ (variable)	Hexagonal, *R3/m*	a = 3.174 b = 10.94	10.92 vs	5.48 s	3.65 s	2.747 m	2.660 ms	2.459 ms	Bernal et al. (1959)

[a] Numbers of the format XX-XXXX indicate ICDD file number (ICDD, 1994).

[b] vs, very strong; s, strong; ms, moderately strong; m, moderate; w, weak.

isostructural with akaganéite, but instead of Cl^-, SO_4^{2-} occupies the tunnels (Figure 22.1; Bigham et al., 1990, 1994). *Feroxyhite* (δ'-FeOOH) consists of sheets of edge-sharing octahedra (Figure 22.1), with the presence of face-sharing octahedra (Manceau and Combes, 1988; Drits et al., 1993a; Manceau and Drits, 1993). In naturally occurring feroxyhite, Fe^{3+} ions are randomly distributed over the octahedral sites, while in synthetic δ-FeOOH, the Fe^{3+} ions are orderly distributed over half of the octahedral sites (Waychunas, 1991). *Ferrihydrite* ($Fe_5HO_8 \cdot 4H_2O$) is a poorly ordered Fe oxide, with a variable degree of ordering. The structure is still being investigated (Drits et al., 1993b; Manceau and Drits, 1993; Manceau and Gates, 1997) but can be visualized as a defective hematite structure containing both edge- and face-sharing octahedra (Towe and Bradley, 1967). The Fe^{3+} ions are randomly distributed over the octahedral interstices, with

many sites vacant and more OH^- and H_2O and less Fe^{3+} than in hematite (Cornell and Schwertmann, 1996).

22.2.1.2 Hydroxides

Green rusts are a group of Fe^{2+}–Fe^{3+} hydroxy salts with a structure consisting of sheets of edge-shared $Fe^{2+}(OH)_6$ octahedra in which some of the Fe^{2+} is replaced by Fe^{3+}, creating a positive layer charge. The charge is balanced by anions such as Cl^-, SO_4^{2-}, and CO_3^{2-} located between the octahedral sheets (Figure 22.1; Brindley and Bish, 1976; Taylor and McKenzie, 1980). *Bernalite*, $Fe(OH)_3 \cdot nH_2O$ (Birch et al., 1993), is a rare Fe hydroxide and has not been found in soils. *Fougerite*, $(Fe^{2+}Mg)_6 \, Fe_2^{3+}(OH)_{18} \cdot 4H_2O$, has recently been identified as a green rust mineral in soils (Trolard et al., 2007), whereas the $Fe(OH)_2$ compound has not been found as a mineral (Cornell and Schwertmann, 1996).

22.2.1.3 Oxides

Hematite (α-Fe_2O_3) consists of sheets of edge-shared octahedra with two-thirds of the available octahedral sites filled with Fe^{3+} ions. The unoccupied sites are regularly arranged to form sixfold rings of occupied octahedra analogous to dioctahedral phyllosilicate sheets (Chapter 21) and are stacked along the c-axis. Each plane of O is shared by two adjacent dioctahedral sheets. Each octahedron shares three edges with three neighboring octahedra in the same sheet and a face and six corners with nine octahedra in adjacent sheets (Figure 22.1; Blake et al., 1966; Maslen et al., 1994). *Magnetite* (Fe_3O_4), which differs from most other Fe oxides in that it contains both Fe^{2+} and Fe^{3+} ions, has an inverse spinel structure consisting of octahedral and mixed tetrahedral/octahedral layers stacked along the [111] plane, with Fe^{3+} occupying tetrahedral sites and both Fe^{2+} and Fe^{3+} in octahedral sites (Figure 22.1). *Titanomagnetites* ($Fe_{3-x}Ti_xO_4$) are solid solutions of magnetite with ulvöspinel (Fe_2TiO_4), with Ti^{4+} occupying only octahedral sites (Wechsler et al., 1984). *Maghemite* (γ-Fe_2O_3) has the chemical composition of hematite but has a structure analogous to magnetite (Figure 22.1). Most or all of the Fe occurs as Fe^{3+} and cation vacancies preserve charge balance caused by the oxidation of Fe^{2+} to Fe^{3+}. The cations are randomly distributed over the tetrahedral and octahedral sites. The vacancies are also randomly distributed but are confined to the octahedral sites. *Titanomaghemites* are oxidation products of titanomagnetites, but their structure needs clarification (Waychunas, 1991). Maghemite–magnetite solid solutions are formed by varying degrees of oxidation of magnetite.

Other metallic cations with an ionic diameter similar to Fe^{3+} (Al, Ni, Ti, Mn, Co, Cr, Cu, Zn, V) may replace the Fe ion in the structure of various Fe oxides. Isomorphous substitution of Al^{3+} for Fe^{3+} occurs most frequently and to the greatest extent (Schwertmann and Carlson, 1994). In addition to M^{3+} cations, M^{2+} and M^{4+} cations may also enter Fe^{3+} oxide structures, but the uptake is usually less than $0.1\,mol^{-1}$ (Cornell and Schwertmann, 1996).

The typical shapes of well-crystallized Fe oxide samples are listed in Table 22.2. These shapes, usually observed in samples from synthetic preparations, rocks, ferricretes, and bauxites, are frequently less well expressed by soil Fe oxides. Soil hematites usually have a granular texture, and goethites show almost no acicularity (Figure 22.2), particularly in highly weathered soils (Schwertmann and Kämpf, 1985; de Brito Galvão and Schulze, 1996). Thus, morphology alone is unreliable for distinguishing these minerals by transmission electron microscopy (TEM) (Anand and Gilkes, 1987a; Singh and Gilkes, 1992), although soil lepidocrocites often appear as thin laths, very similar to synthetic lepidocrocites (Cornell and Schwertmann, 1996). Soil Fe oxides may range from a few to several hundred nanometers in length, but data are scarce because of their tendency to occur as microaggregates (Schwertmann, 1988).

22.2.2 Occurrence and Formation

The concentration of Fe oxides in soil ranges between <1 and $>500\,g\,kg^{-1}$ and is related to parent material, degree of weathering, and pedogenic accumulation or depletion processes. Iron oxides

TABLE 22.2 Diagnostic Criteria for Iron Oxide Minerals

Mineral	Color and Munsell Hues Range	Typical Crystal Shape	DTA Events (°C)	IR Bands (cm⁻¹)	Magnetic Hyperfine Field (kOe)		
					295 K	77 K	4 K
Goethite	Strong brown to yellowish brown, 7.5YR–2.5Y	Needles, laths	Endotherm, 280–400	890, 797	382	503	506
Lepidocrocite	Reddish yellow ("orange"), 5YR–7.5YR	Laths	Endotherm, 300–350 Exotherm, 370–500	1026, 1161, 753			458
Akaganéite	Strong brown to reddish yellow, 2.5YR–7.5YR	Spindle-shaped rods					489/478/473
Schwertmannite	Reddish yellow to yellow, 5YR–10YR	Fibers	Endotherm, 100–300, 650–710 Exotherm, 540–580	3300, 1634, 1186, 1124, 1038, 976, 608			456
Ferrihydrite	Dark red to strong brown, 2.5YR–7.5YR	Spherical	Endotherm, 150				470/500
Feroxyhite	Dark red to dark reddish brown, 1.5YR–5YR	Fibers, needles	Endotherm, 250	1110, 920, 790, 670	420	530	535
Hematite	Dusky red to dark red, 5R–2.5YR	Hexagonal plates	Nil	345, 470, 540	518	542/535[a]	542/535[a]
Maghemite	Dark reddish brown, 2.5YR–5YR	Cubes	Exotherm, 600–800	400, 450, 570, 590, 630	500		526
Magnetite	Black	Cubes	[b]	400, 590	491/460[c]		

[a] With and without Morin transition, respectively.

[b] Magnetite converts via maghemite or directly to hematite, depending on particle size.

[c] For tetrahedral and octahedral Fe, respectively.
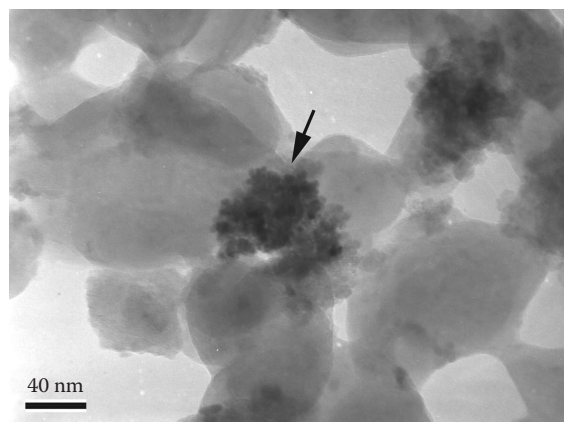
FIGURE 22.2 Transmission electron micrograph of the <0.2 μm fraction of a Rhodic Hapludox showing a goethite or hematite particle (arrow) surrounded by clean kaolinite particles. (Modified from de Brito Galvão, T.C., and D.G. Schulze. 1996. Mineralogical properties of a collapsible lateritic soil from Minas Gerais, Brazil. Soil Sci. Soc. Am. J. 60:1969–1978.)

may occur evenly distributed in the matrix or concentrated as ferricretes, layers, bands, horizons, nodules, mottles, plinthite, etc.

Iron, present as Fe^{2+} in primary minerals (mainly silicates) of most rocks, is released during weathering through protolysis and oxidation processes, hydrolyzes in contact with water, and then forms Fe^{3+} oxides. In most aerobic environments, the Fe oxides are evenly distributed resulting in homogeneous soil color. Under anaerobic conditions, however, Fe oxides may be reduced and dissolved, thus leading to a heterogeneous distribution of Fe oxides and color (redoximorphic features; Schwertmann, 1993). The formation of each of the Fe oxide minerals requires specific conditions that have been established for soils by laboratory and field observations (Schwertmann and Taylor, 1989).

In aerobic environments, goethite and hematite are the most common Fe oxides due to their high thermodynamic stability. Their formation pathway starts with Fe^{2+} released by weathering and its immediate oxidation to Fe^{3+} or by any other source of Fe^{3+} (e.g., dissolution of Fe oxides), which precipitates as ferrihydrite or goethite, depending on which of the solubility products is exceeded first (ferrihydrite, Ksp ~ 10^{-39} or goethite, 10^{-44} < Ksp < 10^{-41}). Hematite forms from its necessary precursor, ferrihydrite, through a solid-state reaction in which individual hematite crystals nucleate and grow within individual ferrihydrite aggregates by a dehydration and rearrangement process. All of the iron in each individual hematite crystal is derived from a single-ferrihydrite aggregate (Cornell and Schwertmann, 1996). The transformation of ferrihydrite to goethite, however, proceeds via a dissolution–precipitation process. The presence of face-sharing $Fe(O,OH)_6$ octahedra in ferrihydrite, as in hematite, prevents the direct solid-state transformation of ferrihydrite into goethite, which consists only of edge- and corner-sharing octahedra. Thus, the transformation of ferrihydrite to goethite requires the breakdown of the oxo-bridges of face-sharing octahedra (Combes et al., 1989). Goethite forms from Fe^{3+} ions in solution via a nucleation–crystal growth process. Hence, any Fe source (minerals, biological

exudates, organic compounds) able to keep a low Fe^{3+} activity in solution will favor goethite. Thus, the formation pathways of hematite and goethite are different but competitive.

Environmental conditions that favor the formation of ferrihydrite and its subsequent transformation to hematite are high Fe content in the parent rock (resulting in higher Fe release rate), near neutral pH (pH of minimum solubility of ferrihydrite), higher temperature or lower water activity (favoring the dehydration step) (Torrent et al., 1982), and rapid turnover of biomass (low Fe complexation) (Schwertmann, 1988). Such environmental conditions are usually related to climate (Kämpf and Schwertmann, 1983), landscape (Curi and Franzmeier, 1984; Schwertmann and Latham, 1986), landscape associated with drainage (Macedo and Bryant, 1987; da Motta and Kämpf, 1991; Peterschmitt et al., 1996; Fritsch et al., 2007), and soil depth (Bigham et al., 1978; Kämpf and Schwertmann, 1983).

In soils, the transformation rate of ferrihydrite to hematite is probably very rapid, whereas that of ferrihydrite to goethite, which proceeds through dissolution–recrystallization, is slower. This explains the widespread association of ferrihydrite–goethite, while that of ferrihydrite–hematite has been rarely found (Parfitt et al., 1988). There is no pedogenic indication for a solid-state transformation of goethite to hematite by simple dehydration, nor the inverse by simple hydration. However, high temperatures caused by forest or bush fires may transform goethite into hematite and lepidocrocite into maghemite (Stanjek, 1987; see maghemite below).

Both hematite and goethite have a similarly low solubility (Lindsay, 1979). Under moderately reducing conditions, however, a transformation of red into yellow soils (yellowing or xanthization) may take place due to a preferential dissolution of hematite over goethite (Macedo and Bryant, 1989; Jeanroy et al., 1991; Peterschmitt et al., 1996; Fritsch et al., 2005). The increased resistance of goethite to dissolution is explained by a higher Al for Fe substitution (Torrent et al., 1987; Trolard and Tardy, 1987). The solubility and dissolution of Fe oxides is reviewed by Schwertmann (1991). When Fe oxides are reduced in higher landscape positions, the soluble Fe^{2+} may be transported to lower landscape positions and reprecipitate as Fe oxides (Schwarz, 1994). In aquic environments with prolonged waterlogging, the complete removal of Fe oxides may take place, resulting in soil bleaching (chroma < 2) (da Motta and Kämpf, 1991; Peterschmitt et al., 1996). The bleached color is due to the matrix minerals (phyllosilicates, quartz, etc.) in the absence of Fe oxides.

Many species of microorganisms, mainly anaerobic bacteria, are capable of reducing Fe oxides (Lovley, 1995). As heterotrophic organisms, they depend on available biomass for metabolic oxidation so that reduction in soils is most intense in the upper horizons. The process usually involves enzymatic transfer of electrons by microorganisms from the decomposing biomass to Fe^{3+} (Ghiorse and Ehrlich, 1992; Lovley, 1992, 1995).

Some Fe and Mn oxides may be formed by microorganisms in a biologically induced form outside the cell, or matrix mediated, by internal cellular precipitation (Ghiorse and Ehrlich, 1992). The most common occurrence of biotic formation of

Fe oxides is Fe ochre in field drains (Houot and Berthelin, 1992), the Fe plaque in the root zone of wetland plants (Weiss et al., 2003), and the Fe precipitate, yellow boy, in acid mine drainages (Bigham et al., 1992). Aspects of biomineralization of Fe and Mn in soils are reviewed in Skinner and Fitzpatrick (1992).

The typical Fe oxides of seasonally reduced soil environments are goethite, lepidocrocite, and ferrihydrite, formed by abiotic or biotic processes. Therefore, goethite and lepidocrocite are widely associated in reductomorphic soils, whereas hematite is restricted to mottles and nodules (Schwertmann and Kämpf, 1983; Fitzpatrick et al., 1985; Wang et al., 1993; dos Anjos et al., 1995). While goethite can form from either Fe^{2+} or Fe^{3+} ions in solution, lepidocrocite in soils seems to require the presence of Fe^{2+} ions (Schwertmann and Taylor, 1989). Thus, lepidocrocite, recognized by its bright orange color, indicates prevailing reductomorphic conditions in a soil profile leading to the formation of Fe^{2+}. Factors that favor the formation of goethite over lepidocrocite, however, are a higher partial pressure of CO_2, normally found closer to roots, the presence of HCO_3^- or CO_3^{2-}, an increasing rate of oxidation, and the presence of Al in the system (Schwertmann and Taylor, 1989; Carlson and Schwertmann, 1990).

Ferrihydrite has been reported in ochreous precipitates from the oxidation of emerging, Fe^{2+}-containing waters (Schwertmann and Fischer, 1973; Carlson and Schwertmann, 1981; Schwertmann and Kämpf, 1983), in bog and lake Fe ores (Schwertmann et al., 1982), in podzol B horizons (Adams and Kassim, 1984), in placic horizons (Campbell and Schwertmann, 1984), in Andepts (Parfitt et al., 1988), in soil iron pans (Childs et al., 1990), and in constructed wetlands for acid drainage treatment (Karathanasis and Thompson, 1995). These occurrences reflect an environment where Fe^{2+} is rapidly oxidized (abiotically or biotically) in the presence of high concentrations of organic matter and/or Si. These compounds and possibly others hinder the immediate formation of FeOOH phases and the subsequent transformation of ferrihydrite to more stable Fe oxides (Schwertmann, 1966; Cornell, 1987). In contrast, in tropical soils, the concentration of ferrihydrite is very low because of low interference of Si and organics. The presence of ferrihydrite together with FeOOH phases may indicate that environmental conditions are not favorable for crystal development or that the formations are relatively young, as in paddy soils (Wang et al., 1993) and constructed wetlands (Karathanasis and Thompson, 1995).

Reductomorphic soils often display greenish-blue colors that change rapidly to yellowish brown on exposure to the air. These colors are indicative of the presence of green rusts, recently identified as fougerite (Trolard et al., 2007). Iron-reducing bacteria appear as the main factor involved in green rust formation (Berthelin et al., 2006). Green rusts occur under reducing and weakly acid to weakly alkaline conditions as intermediate phases in the abiotic formation of goethite, lepidocrocite, and magnetite (Schwertmann and Fechter, 1994).

Schwertmannite has been found in strongly acid, sulfate-rich waters associated with mining activities (Bigham et al., 1990, 1992, 1994, 1996b; Fanning et al., 1993), a Histosol that received runoff from lead smelting activities (Gao and Schulze,

2010a, 2010b), and in an acid alpine stream draining a pyritic schist (Schwertmann et al., 1995a). Most of these occurrences reflect acid environments where bacteria catalyze the oxidation of FeS_2 releasing Fe^{3+} and SO_4^{2-} that, in the pH range 2.8–4.0, precipitate as schwertmannite (Bigham et al., 1992, 1996b). At lower pH, jarosite is favored, while at higher pH values, goethite and ferrihydrite form (Bigham et al., 1996a). Schwertmannite is metastable and converts to goethite over time. Bigham et al. (1992) proposed a biogeochemical model for the precipitation of jarosite, schwertmannite, and ferrihydrite, and their conversion to goethite. The model considers the oxidation of Fe^{2+} by *Thiobacillus ferrooxidans* or oxygen, the concentration of SO_4^{2-}, the pH, and the presence of other cations (K, Na) in the system.

Magnetite in soils is usually inherited from the parent rock (lithogenic), but both biologically (Fassbinder et al., 1990) and abiotically formed magnetite (Maher and Taylor, 1988) has been reported. The alteration of magnetite to hematite via solid-state transformation, with no evidence for maghemite development, is described by Gilkes and Suddhiprakarn (1979) and Anand and Gilkes (1984a). In laboratory studies, particle size determines whether hematite or maghemite forms when magnetite is oxidized below 220°C (Egger and Feitknecht, 1962; Gallagher et al., 1968). Particles less than 300 nm in diameter transform to maghemite, while larger particles oxidize to hematite. This may explain why soil maghemites typically occur in the clay fraction.

Maghemite is common in many different soils, especially in the tropics and subtropics, occurring dispersed or concentrated in concretions (Taylor and Schwertmann, 1974; Curi and Franzmeier, 1984; Anand and Gilkes, 1987a; Fontes and Weed, 1991). The two major possible pathways for maghemite formation in soils are the aerial oxidation of lithogenic magnetite (Fontes and Weed, 1991) and the transformation of other pedogenic Fe oxides by heating (between 300°C and 425°C) in the presence of organic compounds (Schwertmann and Fechter, 1984; Anand and Gilkes, 1987b; Stanjek, 1987).

Akaganéite and feroxyhite are both rare minerals. Akaganéite has been found in environments with high chloride concentrations (e.g., 0.1 M), low pH (e.g., 3–4), and high temperature (e.g., 60°C), as found in some hot springs and volcanic deposits (Schwertmann and Fitzpatrick, 1992) and has only recently been reported in soils (Fitzpatrick et al., 2008; Gao and Schulze, 2010a, 2010b). Laboratory synthesis showed that chloride is essential for the crystallization of akaganéite (Shah Singh and Kodama, 1994). The mechanism of feroxyhite formation in nature is unknown. It has been observed in rusty precipitates, formed in the interstices of sand grains from rapidly flowing Fe^{2+}-containing water, which was quickly oxidized (Carlson and Schwertmann, 1980). A possible association with dominant ferrihydrite in some typic Hydrandepts of Hawaii is reported by Parfitt et al. (1988). The similarity of the x-ray diffraction (XRD) patterns of feroxyhite and ferrihydrite makes it extremely difficult, however, to distinguish these minerals in natural samples (Carlson and Schwertmann, 1981).

The different types of Fe oxides may show a partial replacement of Fe by other cations. This substitution is conditioned

by the availability of cations that, in turn, makes it representative of specific environments of the Fe oxide formation, like a mineralogical–pedochemical signature. So far, such relationships have been established for Al substitution in goethites and for V^{3+} in goethite and hematite (Schwertmann and Pfab, 1994, 1996). Medium to high Al substitution (0.15–0.33 mol fraction) is usually observed in goethites from environments with low Si and high Al activity, found in highly weathered tropical and subtropical soils, bauxites, and saprolites (Fitzpatrick and Schwertmann, 1982; Curi and Franzmeier, 1984; Schwertmann and Kämpf, 1985; Anand and Gilkes, 1987a; Fontes et al., 1991; Singh and Gilkes, 1992; Muggler, 1998), whereas goethite with low substitution (0–0.15 mol fraction) prevails in slightly acidic, eutrophic soils and in redoximorphic soils (Fitzpatrick and Schwertmann, 1982). The occurrence of goethites with highly contrasting Al substitution indicates changes in the weathering rate or soil redox conditions (da Motta and Kämpf, 1991; Muggler, 1998). The presence of V^{3+} in goethite and hematite may be used as an indicator for former anoxic environments (Schwertmann and Pfab, 1996).

22.2.3 Influence on Soil Properties

Crystals of Fe oxides in soils are often extremely small, sometimes as small as 5 nm in diameter for poorly crystalline minerals like ferrihydrite or as small as 150 nm in diameter for better crystalline minerals like goethite and hematite. Structural disorder is common. Fe oxides, therefore, exhibit a large specific surface area (70–250 $m^2\,g^{-1}$). They usually occur in higher amounts in comparison to Mn and Al oxides, they have a high point of zero charge (PZC; pH 7–9) and a pH-dependent surface charge. All these factors explain the substantial influence of Fe oxides on the physical and chemical properties of soils.

Significant correlations have been found between the sorption of heavy metals (Cu, Pb, Zn, Cd, Co, Ni, Mn), anions (PO_4, SO_3, MoO_4, AsO_4, Se_4O, S_4O, and organic anions), and the content and mineralogy of Fe oxides (Chapter 17).

The Munsell color ranges for the various Fe oxides are listed in Table 22.2. Detailed aspects of that subject are reviewed by Schwertmann (1993) and Cornell and Schwertmann (1996). The iron oxides may show a variation of color with their crystal size and the isomorphous substitution of Fe by foreign ions (Schwertmann, 1993). Basically, reddish soil color (Munsell hues 5YR and redder) is due to the presence of hematite (and maghemite in some tropical soils), masking the presence of goethite. Thus, the hematite content determines the redness of a soil (Torrent et al., 1983), while the yellow color due to goethite (Munsell hue between 7.5YR and 2.5Y) is expressed only in the absence of hematite. The presence of lepidocrocite is indicated by an orange color, normally restricted to mottles or localized spots in aquic environments. In surface soils, the colors due to Fe oxides may be masked by black organic compounds. Information about aeration and soil drainage can be inferred from the distribution or absence of different Fe oxide minerals in soils, according to their specific formation conditions described above.

Landscape sequences of red soils (hematite and goethite) on well-drained hilltops, through yellow soils (goethite) on moderately drained midslopes, to mottled and gray soils in poorly drained valleys are examples of Fe oxides acting as indicators of aerobic and anaerobic environments (Peterschmitt et al., 1996). Localized accumulations of Fe oxides (mottling, plinthite) and bleached matrix colors (chroma < 2), indicative of seasonal to permanent waterlogged soil environments, are used as diagnostic criteria for redoximorphic features and aquic regimes (Chapter 2).

The aggregating effect of Fe oxides in soils is indicated by significant correlations between the fraction of water-stable aggregates or related structural properties, and the content of Fe oxides (see Schwertmann and Taylor, 1989; Barthès et al., 2008), by electron microscopy (EM) observations of Fe oxide deposits on kaolinite platelets (Fordham and Norrish, 1979), and by the dispersion of aggregated soils after removal of their Fe oxides with a reducing agent (McNeal et al., 1968). Although an aggregating effect of the Fe oxides in soils is generally accepted, the exact mechanism remains obscure. Thin sections and scanning electron microscopy (SEM) observations of nodules, concretions, and ferricretes suggest that cementation develops through the growth of Fe oxide crystals in place between matrix particles, leading to a very stable, nondispersible association of matrix particles (Shafdan et al., 1985). Aggregates, on the other hand, seem to form through an attraction between positively charged Fe oxide particles and negatively charged matrix particles, mainly clay-sized phyllosilicates. A typical example of aggregation is the highly stable microaggregates (coffee powder structure) found in Oxisols. Because these aggregates are not dispersible in water, they have a significant influence on the water-holding capacity and the hydraulic conductivity of these soils (van Wambeke, 1992). On the other hand, these aggregates can be dispersed by organic ligands (oxalate, citrate) without solubilizing much iron (Pinheiro-Dick and Schwertmann, 1996).

22.2.4 Identification

22.2.4.1 Field Techniques

Fe oxides show striking colors, ranging between yellow and red. Particular colors are typical of the various forms and are helpful for their identification in the field (Table 22.2). Correlations have been established between Munsell hue and a redness or yellowness index based on hematite/goethite ratios and hematite content in soils (Torrent et al., 1983). Color alone is, however, not sufficient to unequivocally identify specific Fe oxide minerals. The presence of the magnetic Fe oxides, magnetite, and maghemite, can be easily detected in soils with a hand magnet (Schulze, 1988).

22.2.4.2 Chemical Dissolution Techniques

Chemical dissolution procedures for Fe oxides, initially proposed to eliminate their interference in the examination of phyllosilicates, were later developed to dissolve, and thus

to quantify, specific phases (see reviews by Borggaard, 1988; Parfitt and Childs, 1988). However, the identity of the dissolved phases must be determined by noninvasive techniques, such as XRD and Mössbauer spectroscopy. Potassium or Na pyrophosphate solutions have been used for estimating Fe (and Al) organic complexes (McKeague, 1967; Bascomb, 1968). The soil is usually shaken for 16 h with 0.1 M pyrophosphate at pH 10, but the extraction and clarification procedures are highly technique dependent (Schuppli et al., 1983; Loveland and Digby, 1984). For the extraction of short-range order Fe oxides (predominantly ferrihydrite), a 2 (Schwertmann, 1964) or 4 h (McKeague and Day, 1966) extraction with 0.2 M ammonium oxalate at pH 3 in the dark is widely used. This technique also extracts Fe from organic complexes and poorly crystalline lepidocrocite (Ohta et al., 1993), whereas some magnetites are partly dissolved. Feroxyhite is more resistant to the 2 h oxalate treatment than ferrihydrite (Carlson and Schwertmann, 1980), and for schwertmannite, a 15 min treatment is used (Murad et al., 1994; Bigham et al., 1996b). According to Wang et al. (1993), a shorter oxalate treatment (~10–30 min) is more appropriate to measure ferrihydrite in <2 μm reductomorphic soil fractions than the much longer treatments commonly used for soils. For the complete dissolution of the pedogenic Fe oxide minerals, without dissolution of other minerals, a treatment with sodium dithionite–citrate–bicarbonate (DCB) is used. The extraction is carried out for 15 min at 75°C (Mehra and Jackson, 1960) or by shaking for 16 h with a dithionite–citrate solution at room temperature (Holmgren, 1967). However, as dissolution is affected by particle size, crystallinity (particularly for lithogenic magnetite and hematite) and high Al substitution (goethite and hematite), several DCB treatments may be required.

The ratio of oxalate-extractable Fe (Fe_o) to DCB-extractable Fe (Fe_d) quantifies the proportion of the more and less active fractions, respectively. A high Fe_o/Fe_d ratio and a loss in redness after oxalate treatment give first indication of ferrihydrite in a sample. Additionally, it is a useful parameter for characterizing soil properties (e.g., P sorption) and pedogenic processes (McKeague and Day, 1966; Blume and Schwertmann, 1969). The ratio of Fe_d to Fe_t (total Fe content by HCl or HF digestion) quantifies the proportion of Fe in primary silicate minerals, which has been released by weathering and precipitated as Fe^{3+} oxides, hence estimating Fe sources for potential Fe oxide formation and/or soil weathering stage. Advanced soil weathering stages, indicated by Fe_d/Fe_t ratios > 0.8, are usually shown by Oxisols and Ultisols.

A field test for the presence of Fe^{2+} ions and ferric-organic complexes in reductomorphic soils can be performed with a solution of α,α'-dipyridil (Childs, 1981; Kennedy et al., 1982) or with a solution of 1:10-phenanthroline (Richardson and Hole, 1979). Lovley and Phillips (1987) proposed the use of hydroxylamine hydrochloride to determine whether Fe^{3+} is available for microbial reduction. The amount of reducible Fe in natural wetlands and rice paddies is highly correlated to the labile and stable Fe^{3+} oxides (van Bodegom et al., 2003).

22.2.4.3 X-Ray Diffraction and Other Spectroscopic Techniques

The identification of the various Fe oxides requires physical methods, like XRD, differential thermal analysis (DTA), infrared spectroscopy (IR) (Cambier, 1986), Mössbauer spectroscopy (Murad, 1988, 1996), electron microscopy (EM) (Boudeulle and Muller, 1988; Eggleton, 1988), or magnetic measurements (Coey, 1988). A review of characterization methods for Fe oxides is given by Cornell and Schwertmann (1996) and Schwertmann and Taylor (1989). The most important diagnostic criteria are summarized in Tables 22.1 and 22.2. XRD patterns for the major soil Fe oxides are shown in Figure 22.3.

For powder XRD, unless one uses a monochromator, CoKα or FeKα radiation should be employed to avoid the high-fluorescence background that results when Fe-rich samples are irradiated with CuKα radiation. The sensitivity of XRD is improved by using a differential XRD method (Schulze, 1981) and selective dissolution treatments (Wang et al., 1993). Due to the low concentration of Fe oxides in soils, concentration techniques may be necessary (Schulze, 1988), like particle-size separation, magnetic separation (Schulze and Dixon, 1979), and, for goethite and hematite in kaolinitic soils, a 5 M NaOH treatment (Norrish and Taylor, 1961; Kämpf and Schwertmann, 1982a). A quantitative determination of Fe oxide minerals in soils is generally possible using the intensities of selected x-ray lines, DTA peaks, or Mössbauer hyperfine sextets. Suitable standards with the same characteristics may be isolated from soils or can be synthesized. The accuracy of the quantification can be controlled by comparison with the chemically determined total amount of Fe oxides (Fe_d) (Kämpf and Schwertmann, 1982b).

The Al substitution in goethite can be calculated from the unit-cell c-dimension by Al (mole fraction) = 17.30−57.20c (nm) (Schulze, 1984). The c value is obtained from the XRD positions d(110) and d(111) by $c = [1/d(111)^2 − 1/d(110)^2]^{-1/2}$, with all goethite data based on space group *Pbnm* (Table 22.1). In hematite, Al substitution is given by Al (mole fraction) = 31.09−61.71a (nm), where a = d(110)2, or, preferably, a = d(300)2√3 (Schwertmann et al., 1979).

22.3 Manganese Oxides

Manganese oxides are usually minor components in soils but with a significant influence on the soil chemical properties. While Mn is a micronutrient essential for plants and animals, it is the only plant essential element that also frequently occurs in toxic concentration in acid soils. The mineralogy of Mn oxides is complicated by the large number of minerals and the lack of precise knowledge of some of their structures, leading to uncertainties as to whether certain forms should be regarded as distinct mineral species, or merely as variants, or mixtures of other mineral forms (McKenzie, 1989; Post, 1992). The poor crystallinity and low concentration of these minerals in most soils is a further challenge for their characterization. A list of Mn oxides is given in Table 22.3. More information can be found in McKenzie (1989), Graham et al. (1988), and Dixon and White (2002).
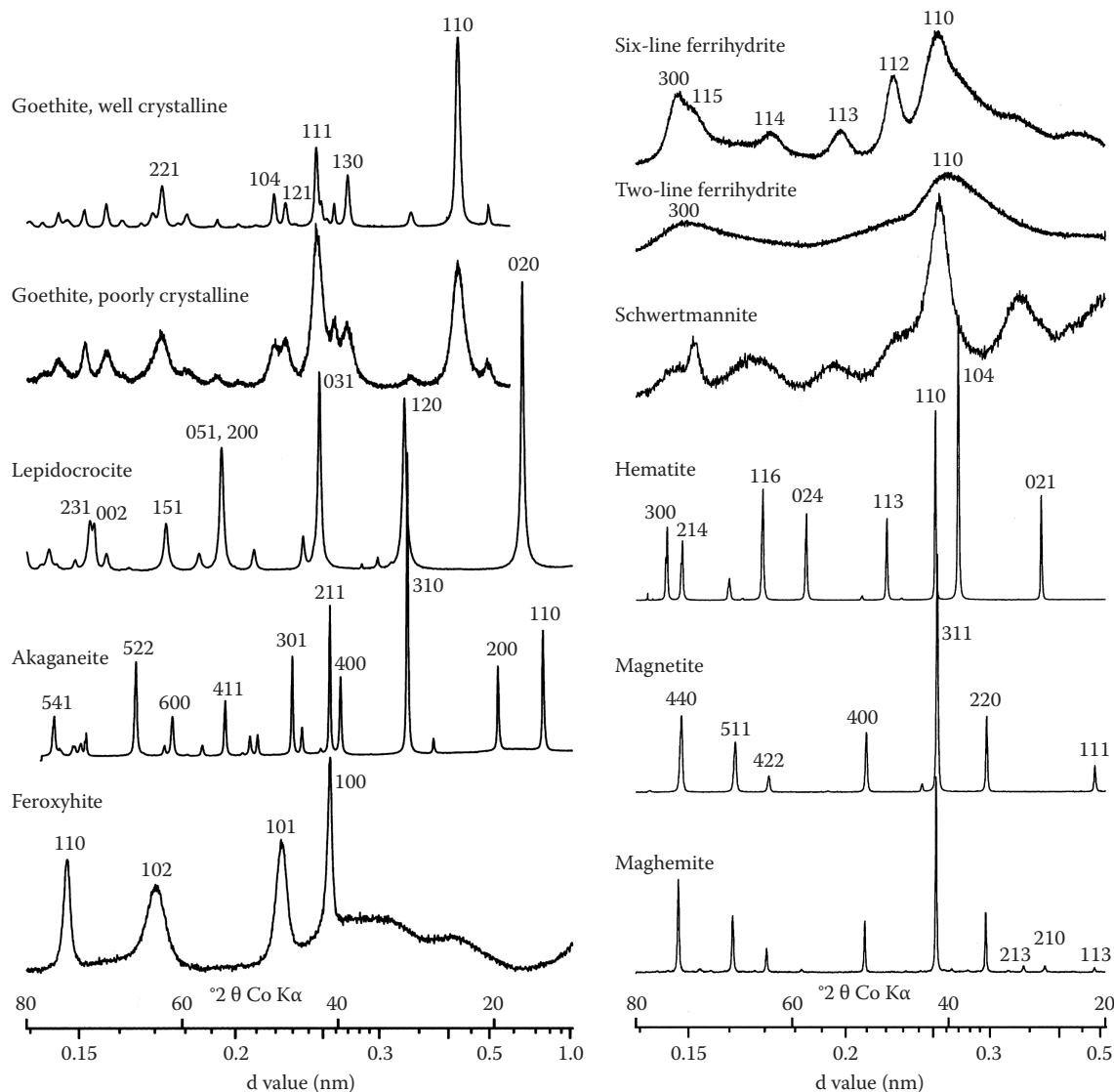
FIGURE 22.3 XRD patterns for the major soil iron oxide minerals (Co Kα, peaks are labeled with their Miller indices). (From Cornell, R.M., and U. Schwertmann. 1996. The iron oxides. Wiley-VCH Verlag, Weinheim, Germany.)

22.3.1 Mineral Phases

Manganese oxides consist mainly of Mn in octahedral coordination. The various minerals differ in the arrangement and linkage of the octahedra. The Mn oxides may be placed into three groups: (1) Tectomanganates or tunnel structures, (2) phyllomanganates or layer structures, both groups with mainly tetravalent Mn, and (3) the trivalent oxides or lower oxides.

22.3.1.1 Tectomanganates (Tunnel Structures)

Tectomanganates are formed from single, double, or wider chains of MnO_6 octahedra that are linked into a framework, thus forming tunnels (or channels) through the structures. These tunnels are occupied by large foreign cations and water molecules. Pyrolusite, ramsdellite, and nsutite are also referred

as chain structures because they do not contain tunnels but are included in the group of tectomanganates because of their structural similarity to the channel or tunnel structure members.

Pyrolusite, MnO_2, consists of single chains of edge-sharing MnO_6 octahedra linked by sharing corners to form 1×1 pseudotunnels (Figure 22.1; Baur, 1976). The single rows of unoccupied octahedral sites are not wide enough to accommodate foreign ions or water molecules. High-resolution transmission electron microscopy (HRTEM) of pyrolusite crystals showed lamellae with the structure of ramsdellite (Yamada et al., 1986). *Ramsdellite*, MnO_2, consists of double chains of MnO_6 octahedra forming 1×2 tunnels (Figure 22.1; Post, 1992), that possibly contain water molecules (Potter and Rossman, 1979). *Nsutite* has a structure formed by the intergrowth of pyrolusite-like single chains and ramsdellite-like double chains of octahedra

TABLE 22.3   Manganese Oxide Minerals: Crystallographic Properties and d Values for the Six Most Intense Diffraction Lines

Mineral (Synthetic Equivalent)[a]	Chemical Formula	Crystal System, Space Group	Unit-Cell Dimensions (Å)	Six Most Intense Diffraction Lines d Value (Å), Relative Intensity						References[b]
Pyrolusite; (β-MnO_2)	MnO_2	Tetragonal, $P4_2/mnm$	a = 4.398	3.11	2.407	1.623	1.306	1.304	2.110	Baur (1976); 24-0735
			c = 2.873	100	55	55	20	20	16	
Ramsdellite	MnO_2	Orthorhombic, $Pnma$	a = 9.27	4.06	1.647	2.438	2.344	1.660	1.906	39-0375
			b = 2.866	100	57	49	46	34	31	
			c = 4.533							
Nsutite (γ-MnO_2)	$Mn^{4+}Mn^{3+}$ (O,OH)$_2$	Hexagonal, P	a = 9.65	1.635	4.00	2.33	2.42	2.13	1.60	Zwicker et al. (1962); 17-0510
			c = 4.43	100	95	70	65	45	45	
Hollandite (α-MnO_2)	$Ba_x(Mn^{4+}Mn^{3+})_8O_{16}$ (x=1)	Monoclinic, $I2/m$	a = 10.026	3.10	3.14	3.172	2.412	3.069	3.459	Post et al. (1982); 38-0476
			b = 2.8782	100	88	40	37	32	23	
			c = 9.729							
			β = 91°03′							
Cryptomelane	$K_x(Mn^{4+}Mn^{3+})_8O_{16}$ (x=1.3–1.5)	Monoclinic, $I2/m$	a = 9.956	2.40	3.122	3.09	2.41	4.852	7.01	Post et al. (1982); 44-1386
			b = 2.8705	100	51	45	40	32	30	
			c = 9.706							
			β = 90°95′							
Coronadite	$Pb_x(Mn^{4+}Mn^{3+})_8O_{16}$ (x=1–1.4)	Monoclinic, $I2/m$	a = 9.938	3.124	3.491	2.209	2.409	6.98	1.548	Post and Bish (1989); 41-0596
			b = 2.8678	100	30	18	17	7	6	
			c = 9.834							
			β = 90°39′							
Romanèchite	$Ba_{0.66}(Mn^{4+}Mn^{3+})_5O_{10} \cdot 1.34H_2O$	Monoclinic, $C2/m$	a = 13.919	2.408	2.188	3.481	6.96	2.366	2.882	Turner and Post (1988); 14-0627
			b = 2.8459	100	85	60	55	50	40	
			c = 9.678							
			β = 92°39′							
Todorokite	$(Na,Ca,K)_{0.3–0.5}$ $(Mn^{4+}Mn^{3+})_6O_{12} \cdot 3.5H_2O$	Monoclinic, $P2/m$	a = 9.764	9.55	2.399	2.388	2.355	4.77	2.345	Post and Bish (1988); 38-0475
			b = 2.8416	100	36	25	24	24	25	
			c = 9.551							
			β = 94°06′							
Chalcophanite	$ZnMn_3O_7 \cdot 3H_2O$	Trigonal, $R3$	a = 7.533	6.93	2.228	4.07	1.590	3.507	2.550	Post and Appleman (1988); 45-1320
			c = 20.794	100	40	29	23	19	17	

(*continued*)

TABLE 22.3 (continued)   Manganese Oxide Minerals: Crystallographic Properties and d Values for the Six Most Intense Diffraction Lines

Mineral (Synthetic Equivalent)[a]	Chemical Formula	Crystal System, Space Group	Unit-Cell Dimensions (Å)	Six Most Intense Diffraction Lines d Value (Å), Relative Intensity						References[b]
Birnessite	$(Na,Ca,Mn^{2+})Mn_7O_4 \cdot 2.8\ H_2O$	Monoclinic, $C2/m$	Na: a = 5.175	7.14	3.57	2.519	2.429	2.154	2.222	Post and Veblen (1990);
			b = 2.850	100	27	14	13	7	5	43-1456
			c = 7.337							
			β = 103°18′							
Vernadite (δ-MnO₂)	$MnO_2 \cdot nH_2O$	(Pseudo) tetragonal $I4/m$	a = 9.866	2.39	3.11	2.15	1.827	1.537	1.422	Chukhrov et al. (1980);
			c = 2.844	100	60	60	40	40	40	15-0604
Rancieite	$(Ca,Mn)Mn_4O_9 \cdot nH_2O$	Hexagonal, P	a = 8.68	7.49	3.74	2.463	2.342	1.425	2.064	22-0718
			c = 9.00	100	14	10	6	4	2	
Buserite	$Na_4Mn_{14}O_{27} \cdot 21H_2O$	Orthorhombic	a = 17.5	10.1	5.01	3.34	1.46	2.56	2.47	32-1128
			b = 30.7	100	70	50	50	30	30	
			c = 10.2							
Lithiophorite	$LiAl_2Mn_2^{4+}Mn^{3+}O_6(OH)_6$	Trigonal, $R3m$	a = 2.924	4.71	9.43	2.371	1.880	3.143	1.453	Post and Appleman (1994);
			c = 28.169	100	68	24	14	7	4	41-1378
Groutite (α-MnOOH)	MnOOH	Orthorhombic, $Pbnm$	a = 4.560	4.20	2.81	2.67	2.30	1.695	2.38	Glasser and Ingram (1968);
			b = 10.70	100	70	70	60	50	40	12-0733
			c = 2.870							
Manganite (γ-MnOOH)	MnOOH	Monoclinic, pseudoorthorhombic, $B2_1m$	a = 8.88	3.40	2.64	1.782	2.417	1.672	2.414	Dachs (1963); in Bricker
			b = 5.25	100	24	21	17	17	16	(1965); 41-1379
			c = 5.71							
			β = ~90°							
Feitknechtite (β-MnOOH)	MnOOH	Tetragonal, P	a = 8.6	4.62	2.635	2.36	1.96	1.55	1.50	Bricker (1965); 18-0804
			c = 9.30	100	50	20	10	1	1	
Hausmannite	Mn_3O_4	Tetragonal, $I4_1/amd$	a = 5.7621	2.847	2.768	1.544	3.089	4.924	1.799	Bricker (1965); 24-0734
			c = 9.4696	100	85	50	40	30	25	
Manganosite	$Mn^{2+}O$	Cubic, $Fm3m$	a = 4.4448	2.223	2.568	1.571	1.34	0.994	0.907	Sasaki et al. (1980); 7-0230
				100	60	60	20	18	16	
Bixbyite (α-Mn₂O₃)	$(Fe,Mn)_2O_3$	Cubic, $Ia3$	a = 9.4091	2.716	1.663	3.842	2.352	1.418	1.845	Geller (1971); 41-1442
				100	28	16	14	13	9	

[a] Designation commonly used in the literature for a synthetic compound with the same crystal structure as the natural mineral are given in parentheses.

[b] Numbers of the format XX-XXXX indicate ICDD file number (ICDD, 1994).

in random fashion (Zwicker et al., 1962; Potter and Rossman, 1979; Turner and Buseck, 1983). Inclusions of todorokite have also been found in nsutite (Turner and Buseck, 1983). *Hollandite*, *cryptomelane*, and *coronadite*, sometimes grouped as α-MnO_2, consist of double chains of edge-sharing MnO_6 octahedra linked to form 2 × 2 tunnels (Figure 22.1; Post et al., 1982; Post and Burnham, 1986; Post and Bish, 1989). These minerals have the general formula $A_{0-2}(Mn^{4+},Mn^{3+})_8(O,OH)_{16}$. The tunnels contain water molecules along with the A cation, which is primarily Ba^{2+} in hollandite, K^+ in cryptomelane, and Pb^{2+} in coronadite. Natural samples usually contain a variety of cations in the tunnels. These large cations are located at specific tunnel sites, and their presence is necessary to prevent the structure from collapsing (Giovanoli, 1985a). The charges of the tunnel cations are balanced by the substitution of Mn^{4+} by Mn^{3+} ions. *Romanèchite* consists of double and triple chains of edge-sharing octahedra that share corners to form a framework containing 2 × 3 tunnels. The tunnels contain Ba^{2+} (and a variety of other large cations, such as Na, K, Sr) and water molecules (Figure 22.1; Turner and Post, 1988). The charges of the tunnel cations are balanced by the substitution of Mn^{4+} by Mn^{3+} ions. HRTEM images show that intergrowths of romanèchite and hollandite are common and are produced by the sharing of the double chains common to both minerals with double or triple chains occurring at random (Turner and Buseck, 1979). While romanèchite has also been called psilomelane, McKenzie (1989) used psilomelane for a structural type and Waychunas (1991) for mixtures of Mn oxide minerals. *Todorokite*, which had its structure and existence as a single mineral questioned until recently (Burns et al., 1985; Giovanoli, 1985b), consists of triple chains of edge-sharing MnO_6 octahedra linked to form large 3 × 3 tunnels (Post and Bish, 1988) containing Na, Ca, K, Ba, Sr, and water molecules (Figure 22.1). The octahedra at the edges of the triple chains are larger than those in the middle and, therefore, probably accommodate the larger, lower-valence cations (Mg^{2+}, Mn^{3+}, Cu^{2+}, Ni^{2+}, etc.) found in todorokite samples (Post and Bish, 1988). The occurrence of variable tunnel widths (3 × 2, 3 × 3, 3 × 4, and 3 × 5) observed in HRTEM images suggests that todorokite represents a family rather than a single mineral (Turner and Buseck, 1981).

22.3.1.2 Phyllomanganates (Layer Structures)

Chalcophanite has a layer structure composed of sheets of edge-sharing MnO_6 octahedra alternating with planes of Zn cations (but also Mn^{2+}, Ba, Ca, Mg, K, Pb, Cu, etc.) and water molecules (Figure 22.1; Ostwald, 1985; Post and Appleman, 1988). One of every seven octahedral sites in the Mn–O sheet is vacant, and the Zn cations are situated above and below the vacancies. The other phyllomanganates are structurally analogous to chalcophanite, with Na^+, Ca^{2+}, K^+, and Mn^{2+} as interlayer cations. *Birnessite*, consists of a layer structure analogous to chalcophanite but with fewer vacancies in the octahedral sheets and with Na, K, or Mg replacing the Zn cations (Figure 22.1; Post and Veblen, 1990). The interlayer region contains water molecules as in chalcophanite. *Rancieite* has a layer

structure similar to birnessite, with Ca^{2+} as the main interlayer cation and interlayer water molecules (Bardossy and Brindley, 1978; Potter and Rossman, 1979; Chukhrov et al., 1980). The structure and existence in nature of *buserite*, also known as 10 Å manganite, are still unresolved. Buserite appears to have a layer structure similar to birnessite, with the larger, 10 Å layer spacing probably due to interlayer water (Waychunas, 1991). *Lithiophorite* has a layer structure comprising sheets of edge and corner-sharing MnO_6 octahedra alternating with sheets of $(Al,Li)(OH)_6$ octahedra (Figure 22.1; Post and Appleman, 1994). The cation sites in the Mn–O octahedral sheet are fully occupied, 2/3 with Mn^{4+} and 1/3 with Mn^{3+}. In the (Al,Li)–OH octahedral sheet, 2/3 of the sites are occupied by Al and 1/3 by Li cations. Lithiophorites with only traces of Li and others with wide variations in Ni, Co, Cu, and Zn concentrations inversely related to the Al content have been found (Ostwald, 1984a). According to Manceau et al. (1987, 1990), Co can occur within the octahedral Mn sheets, while Ni and Cu are located in the (Al,Li)–OH octahedral sheet, probably replacing Li, whereas in asbolane, Ni builds partial $Ni(OH)_2$ sheets. Manceau et al. (1987) also found evidence for mixed layering between lithiophorite and asbolane, as observed by Ostwald (1984a). *Asbolane* has a layer structure with alternating sheets of Mn^{4+}–O octahedra and Co–Ni–OH octahedra (Figure 22.1; Chukhrov and Gorshkov, 1981). The Co–Ni sheet may be discontinuous (island-like). The positive charge of the Mn^{4+} sheets is balanced by the negative charge of the Co–Ni sheets. Hydrogen bonding occurs between the oxygen atoms of the Mn sheets and the hydroxyl groups of the Co–Ni sheets. Most asbolanes are fine-grained, poorly crystalline minerals. *Vernadite* has been considered a disordered birnessite, lacking regular stacking in the c-axis. However, according to Chukhrov et al. (1980) and Chukhrov and Gorshkov (1981), it is a distinct mineral species with a disordered structure that has some similarity to that of birnessite, represented by layers of hexagonal close-packed O and water molecules in which less than half of the octahedra are occupied by Mn ions. Diffraction patterns of vernadite have only two d-spacings at ~2.4 and ~1.4 Å, whereas birnessite yields additional reflections at 7.0–7.2, 3.5–3.6 Å, and other spacings (Chukhrov et al., 1980).

22.3.1.3 Trivalent Manganese Oxides and Oxyhydroxides

Groutite (α-MnOOH) is isostructural with goethite and ramsdellite (Figure 22.1; Glasser and Ingram, 1968). *Feitknechtite* (β-MnOOH) has a structure similar to lepidocrocite (Figure 22.1), but its structure has not been refined (Waychunas, 1991). *Manganite* (γ-MnOOH) is similar to pyrolusite and rutile (Figure 22.1; Huebner, 1976). *Hausmannite* (Mn_3O_4) has a disordered spinel structure analogous to magnetite (Figure 22.1; Huebner, 1976). *Manganosite* has the NaCl structure and is of limited occurrence in nature (Huebner, 1976). *Bixbyite* (α-$(Fe,Mn)_2O_3$) has an anion-deficient fluorite structure (Geller, 1971), while pure (synthetic) α-Mn_2O_3 is orthorhombic (Huebner, 1976; Waychunas, 1991).

22.3.2 Occurrence and Formation

Many studies have been made on the synthesis of Mn oxides (McKenzie, 1971; Hem and Lind, 1983; Giovanoli, 1985a) and the transformation of one form to another (Faulring et al., 1960; Bricker, 1965; Glasser and Smith, 1968; Rask and Buseck, 1986). Some of these studies, even if more applicable to diagenetic and hydrothermal conditions and to the formation of ore deposits, may be extrapolated to soil environments. Of the variety of Mn oxides found in terrestrial environments, only a few have been positively identified in soils. For example, nsutite and pyrolusite are common in Mn ore deposits (Zwicker et al., 1962; Varentsov, 1982) but have not been found in soils. Synthesis studies show that large amounts of foreign ions prevent the formation of nsutite, while even small amounts of foreign ions prevent the formation of pyrolusite. In soils, the foreign ions released by mineral weathering probably limit the formation of these Mn oxides, whereas in ore deposits, the abundance of Mn would aid in the removal of foreign ions, thus allowing the formation of nsutite and pyrolusite (McKenzie, 1989). According to Ostwald (1984b), the mineral associations birnessite–montmorillonite, chalcophanite–kaolinite, and lithiophorite–gibbsite observed in Australian Mn ore deposits may represent special cases of metasomatic replacements of clay minerals by phyllomanganate minerals, whereas the occurrence of vernadite in a wide range of substrates is explained by its formation as a product of rapid microbial oxidation of Mn^{2+} (Chukhrov and Gorshkov, 1981).

Laboratory experiments show that Mn_3O_4 (hausmannite) is the initial product of the chemical oxidation of aqueous Mn^{2+} at 25°C in moderately alkaline solutions (pH 8.5–9.5). It converts to β-MnOOH (feitknechtite) and then to γ-MnOOH (manganite) by aging in solution (Hem and Lind, 1983; Murray et al., 1985). At 0°C, the initial product is feitknechtite, which converts to MnO_2, identified as ramsdellite or birnessite (Hem and Lind, 1983). The formation and transformation of Mn oxides through successively increasing oxidation states are compatible with the common distribution of Mn oxides observed in supergene Mn ores (Bricker, 1965) and in lateritic weathering sequences (Parc et al., 1989). Minor inversions of that Mn oxidation sequence occur, however, due to the presence of foreign ions, thus promoting the formation of cryptomelane and lithiophorite instead of pyrolusite (Parc et al., 1989). The influence of microorganisms and organic compounds additionally increases the complexity of Mn oxide formation in soil environments.

Manganese is one of the first elements released during weathering of primary minerals, which explains its common accumulation in saprolites. Mobile in solution as Mn^{2+}, its oxidation to Mn^{3+} and Mn^{4+} and its subsequent precipitation is often accelerated by microorganisms (Ghiorse, 1988). Most of the Mn oxides found in surface environments are compounds of Mn^{4+} with some Mn^{3+}. The typical fine-grain size and poor crystallinity of the Mn oxides in soils are probably related to seasonal moisture changes, the presence of interfering organic and inorganic components in the soil solution, and the presence of multiple Mn oxidation states.
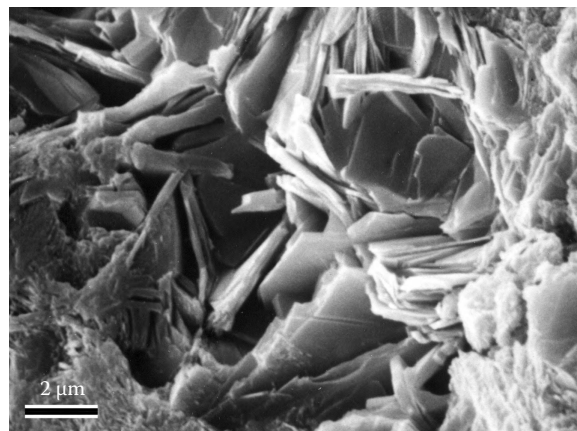
FIGURE 22.4 Scanning electron micrograph of the fractured surface of a nodule from a Tropeptic Eutrustox showing crystals of the Mn oxide mineral lithiophorite. (From Golden, D.C., J.B. Dixon, and Y. Kanehiro. 1993. The manganese oxide mineral, lithiophorite, in an Oxisol from Hawaii. Aust. J. Soil Res. 31:51–66.)

Manganese oxides are commonly of authigenic origin in soils, formed by direct chemical or biochemical precipitation from solution and by crystallization of poorly organized colloidal sols. However, little is known about the processes involved in their pedogenic formation. In most soils, Mn oxides occur as finely dispersed particles but may also be found as discontinuous black–brown coatings of ped surfaces (mangans), as pore fillings, and as concretions and nodules (Figure 22.4). Their presence and concentration are more likely in soil environments with alternate reducing and oxidizing conditions, which affect the mobility and the precipitation of Mn (White and Dixon, 1996). Thus, higher accumulations of Mn oxides are more frequent in soils with restricted drainage such as the aquic suborders (Aqualfs, Aquolls, Aquults, etc.), and intergrades (Aquic Hapludalfs, Aquic Dystrochrepts, etc.), and are also commonly found in saprolites. However, Mn accumulations in a soil are not necessarily an indication of a present redox environment but may reflect former wetter conditions (Allen and Hajek, 1989).

Mn and Fe oxides are usually associated in nodules with dominant silicates (quartz, clay minerals) that probably act as primary templates for the Mn and Fe oxidation and precipitation (Schwertmann and Fanning, 1976). Golden et al. (1986) demonstrated experimentally that Mn^{4+} oxide minerals can also act as oxidizing agents of Fe^{2+} ions in solution, causing the precipitation of Fe^{3+} oxides in association with Mn oxides. The crystal growth and progressive cementation of the nodules can be explained by an autocatalytic oxidation of Mn^{2+} adsorbed onto the formed Mn (and Fe) oxides (McBride, 1994). Thus, abiotic oxidation accelerates in response to the increased surface available for selective adsorption of Mn^{2+}. The overall process, of Mn oxidation and nodule formation, may also be promoted by Mn-oxidizing bacteria (Robbins et al., 1992; Ehrlich, 1996; Leinemann et al., 1997) and fungi (Thompson et al., 2005, 2006). Furthermore, microorganisms may indirectly control local environments, through elevation of pH and Eh, which favor

oxidation of Fe and Mn (Ghiorse and Ehrlich, 1992). Models of biotic oxidation of Mn (and Fe) are described by Ghiorse and Ehrlich (1992), Ehrlich (1990), and Tebo et al. (2004).

In soil samples from Australia, the abundance of Mn oxides decreased from birnessite and lithiophorite, over hollandite, to only one occurrence of each of pyrolusite and todorokite (Taylor et al., 1964). A high frequency of birnessite was also found in soils of other regions (Taylor, 1968; Ross et al., 1976). Todorokite formation from Mn released by the weathering of Mn-substituted siderite in lignite overburden was reported by Senkayi et al. (1986). Birnessite, lithiophorite, and lattice fringes attributed to todorokite were observed in nodules of a Vertic Argiustoll, while lithiophorite only was found in a Rhodic Paleudult (Uzochukwu and Dixon, 1986). Romanèchite was identified in nodules of a Typic Ochraquult (Robbins et al., 1992), and vernadite associated with feroxyhite was identified in a Scottish Gleyic Cambisol (Birnie and Paterson, 1991). Chukhrov and Gorshkov (1981), as cited by McKenzie (1989), reported that the most common Mn oxide in soils was vernadite, followed by birnessite, cryptomelane, todorokite, and hausmannite. Chukhrov and Gorshkov (1981) reported the presence of hausmannite in unspecified soils with pH ~ 8.2.

22.3.3 Influence on Soil Properties

The wide range of PZC values reported for synthetic samples of Mn oxides (Healy et al., 1966; Crowther and Dillard, 1983; Oscarson et al., 1983), from 1.5–3.5 for birnessite to 2.8–4.6 for the hollandite group and 6.4–7.3 for pyrolusite, is probably related to the synthesis conditions. In general, most Mn oxides have a PZC at a pH < 4.0, a high negative charge, a large range of surface area (5–$360\,m^2\,g^{-1}$), and show a strong specific adsorption of cations.

A high sorption capacity of Mn oxides for metal ions has been found by many authors (see Murray, 1975; McKenzie, 1989). In general, metal ions are sorbed in the increasing order Mg < Ca < Sr < Ba < Ni < Zn < Co < Mn < Cu < Pb (Murray, 1975), leading to the accumulation of relatively high concentrations of heavy metals (Childs, 1975; Sidhu et al., 1977) and actinides in the Mn oxides (Means et al., 1978; Cerling and Turner, 1982).

Manganese oxides are strong inorganic oxidants, thus affecting the availability or the hazard of particular metals. In the case of Co adsorption onto the birnessite surface at pH <7, Co^{2+} is oxidized to Co^{3+} by Mn^{4+}, with formation of Mn^{2+} in the process (Crowther and Dillard, 1983). The strong adsorption mechanism has a significant influence on the availability of Co to plants (Adams et al., 1969; McKenzie, 1978).

The Mn oxides have a major influence on the toxicity and bioavailability of As and Cr in terrestrial and aquatic environments. The reduced species As^{3+}, which is more toxic, more soluble, and more mobile than the oxidized As^{5+} species, is effectively adsorbed and oxidized by Mn^{4+} oxides at pH <6 (Oscarson et al., 1983; Thanabalasingam and Pickering, 1986; Scott and Morgan, 1995). However, the ability to deplete the concentration of As^{3+} in solution varies between different types of Mn oxides and is

related to the crystallinity, specific surface areas, and the PZC of the oxides. In contrast to other transition elements, Cr toxicity and mobility increase with its oxidation state. Thus, the presence of Mn oxides, as oxidizing agents for Cr^{3+} in soil systems, should be considered in disposal studies (Fendorf et al., 1992).

Manganese and Fe oxides also act as final electron acceptors oxidizing organic compounds and are conversely dissolved in the process. Organic compounds that form inner-sphere complexes with the oxide surface (e.g., catechol) dissolve the Mn oxide more quickly than those compounds that form outer-sphere complexes (e.g., hydroquinone) (Stone and Morgan, 1984a, 1984b; McBride, 1987). Lovley and Phillips (1988) found evidence that microorganisms can obtain energy for growth by coupling the oxidation of organic matter to the dissimilatory reduction of Fe^{3+} and Mn^{4+}. They proposed a microbial food-chain model, in which fermentative organisms initially metabolize complex organic matter, and in the next stage, a separate group of bacteria oxidizes the fermentation products to CO_2 while reducing Fe^{3+} and Mn^{4+}. The overall process may be an important degradative pathway for organic compounds and the formation of humic material, while the reduction and dissolution of Mn oxides increases the Mn mobility and bioavailability to organisms. Hui et al. (2003), for example, showed that aromatic amines are oxidized by soil Mn oxides resulting in the release of Mn^{2+} to solution.

22.3.4 Identification

Accumulations of Mn oxides in soils are readily identified by their characteristic black–brown coloration. The effervescence observed with the addition of H_2O_2 is a usual field criterion that helps to confirm the presence of these minerals. The small amount of Mn oxides in soils, the diffuse nature of the x-ray patterns of some of the minerals, and the coincidence of diagnostic lines with those of associated matrix minerals can make their identification by XRD difficult. Thus, XRD, preferably with Fe Kα radiation, can be used only where natural segregations of the minerals occur, such as nodules, veins, or coatings. Even then, a concentration pretreatment may be necessary. Powder XRD patterns for some of the more important Mn oxide minerals are shown by Post (1992) and the most intense XRD lines of the major minerals are listed in Table 22.3. Because of the lower crystallinity, soil Mn oxides may, however, show some deviation from these patterns. A selective dissolution of Mn oxides with acidified hydroxylamine hydrochloride was proposed by Chao (1972), whereas Tokashiki et al. (1986) characterized Mn oxides by using successive sodium hydroxide, hydroxilamine hydrochloride, and DCB treatments. Concentration procedures for Mn oxides, selective chemical treatments, and identification by XRD and thermal methods are also reported by Uzochukwu and Dixon (1986). A comprehensive compilation of IR powder-absorption spectra of synthetic and naturally occurring Mn oxides is given by Potter and Rossman (1979).

Crystal structure and crystal chemistry refinements have been performed with HRTEM, electron-microprobe analysis, electron diffraction, powder XRD with Rietveld analysis

(Post and Appleman, 1988; Post and Bish, 1988; Post and Veblen, 1990), and x-ray absorption spectroscopy (Manceau et al., 1987, 1992a, 1992b; Manceau and Combes, 1988). HRTEM has led to the discovery of complex intergrowths of variable lattice periodicities in some Mn oxides (Turner and Buseck, 1979, 1981, 1983; Yamada et al., 1986).

22.4 Aluminum Oxides

Of six different Al oxide minerals, gibbsite and, less commonly, boehmite form under soil conditions. Nordstrandite and bayerite have only been identified in restricted geologic environments. Diaspore and corundum are occasionally found in bauxite deposits but are rare in soils. The relative rarity of Al oxide minerals in soils, which is in contrast to their structural Fe analogs, may be explained by the competitive formation of aluminosilicates and by the difficulty of identifying small amounts of these minerals by XRD (Taylor, 1987). More information on Al oxides is given by Hsu (1989) and Huang et al. (2002).

22.4.1 Mineral Phases

Crystallographic properties for the Al oxide minerals likely to occur in soils are listed in Table 22.4. Aluminum occurs exclusively in octahedral coordination in the Al oxides, and the

structures can be described in terms of the arrangement and linkage of Al-containing octahedra.

22.4.1.1 Hydroxides

Gibbsite (γ-Al(OH)$_3$) consists of sheets of edge-shared octahedra with two-thirds of the available octahedral sites filled with Al^{3+} ions. The occupied octahedral sites form sixfold rings, analogous to the dioctahedral sheet of phyllosilicates. The octahedral sheets are stacked along the c-axis, and the OH groups in one sheet reside directly above those in the next, rather than in the position of closest packing. The interlayer attractions are weak and are dominated by hydrogen bonding (Figure 22.1; Hsu, 1989). *Bayerite*, a second polymorphic form of Al(OH)$_3$, is analogous to gibbsite except that the OH groups of one sheet reside in the depressions between the O atoms of the next sheet and are in the close-packing position (Figure 22.1; Hsu, 1989). *Nordstrandite*, a third polymorphic form of Al(OH)$_3$, has a structure in which the stacking of the octahedral sheets alternate between the gibbsite and bayerite arrangements (Hsu, 1989). In addition to these crystalline phases, amorphous Al(OH)$_3$ may occur in soils (Süsser and Schwertmann, 1991). *Doyleite*, a fourth polymorphic form of Al(OH)$_3$, has a structure of sheets of edge-sharing octahedra (Chao et al., 1985) and is a rare hydrothermal mineral not found in soils.

TABLE 22.4   Aluminum Oxide Minerals: Crystallographic Properties and d Values for the Six Most Intense Diffraction Lines

Mineral	Chemical Formula	Crystal System, Space Group	Unit-Cell Dimensions (Å)	Six Most Intense Diffraction Lines d Value (Å), Relative Intensity						References[a]
Gibbsite	γ-Al(OH)$_3$	Monoclinic, $P2_1/n$	a = 8.6552	4.85	4.371	2.385	4.32	2.45	2.05	33-0018
			b = 5.0722	100	70	55	50	40	40	
			c = 9.7161							
			β = 94°607′							
Nordstrandite	Al(OH)$_3$	Triclinic, $P1$	a = 5.082	4.79	2.27	4.32	2.393	2.016	1.902	24-0006
			b = 5.127	100	30	25	25	25	20	
			c = 4.980							
			α = 93°40′							
			β = 118°55′							
			γ = 70°16′							
Bayerite	Al(OH)$_3$	Monoclinic, $P21/a$	a = 5.062	2.222	4.71	4.35	1.723	3.20	1.333	20-0011
			b = 8.671	100	90	70	40	30	18	
			c = 4.713							
			β = 90°27′							
Doyleite	Al(OH)$_3$	Triclinic, $P/1$ or $P1$	a = 5.00	4.794	2.36	1.972	1.857	1.842	4.29	Chao et al. (1985)
			b = 5.17	100	40	30	30	30	20	
			c = 4.98							
			α = 97.5°							
			β = 118.6°							
			γ = 104.74°							
Diaspore	α-AlOOH	Orthorhombic, $Pbnm$	a = 4.396	3.99	2.317	2.131	2.077	1.633	2.558	05-0355
			b = 9.426	100	56	52	49	43	30	
			c = 2.844							
Boehmite	γ-AlOOH	Orthorhombic, $Amam$	a = 3.70	6.11	3.164	2.346	1.86	1.85	1.453	21-1307
			b = 12.227	100	65	55	30	25	16	
			c = 2.868							
Corundum	α-Al$_2$O$_3$	Rhombohedral, $R3c$	a = 4.7592	2.086	2.551	1.60	3.48	1.374	1.74	43-1484
			c = 12.992	100	98	96	72	57	48	

[a] Numbers of the format XX-XXXX indicate ICDD file number (ICDD, 1994).

22.4.1.2 Oxyhydroxides

Diaspore (α-AlOOH) is isostructural with goethite (Figure 22.1; Ewing, 1935a, 1935b; Busing and Levy, 1958). *Boehmite* (γ-AlOOH) has the same structure as lepidocrocite (Figure 22.1). Poorly crystalline boehmite, formerly also called pseudo-boehmite, has a restricted number of unit cells along the b-axis (Tettenhorst and Hofmann, 1980).

22.4.1.3 Oxides

Corundum (α-Al$_2$O$_3$) is isostructural with hematite (Figure 22.1; Maslen et al., 1993).

22.4.2 Occurrence and Formation

Aluminum is a common element in primary silicate minerals, from which it may be released by weathering. Depending on the environmental conditions, the free Al hydrolyzes and precipitates as Al hydroxide or oxyhydroxide, and then crystallizes. The reason this fails to occur in many soils is because aluminum usually binds to silicon to form aluminosilicate clay minerals. Only in the most intense leaching environments is the Si concentration low enough that gibbsite may form. However, the detailed chemistry of Al oxide formation in soils remains obscure, mainly because of the uncertainties in the formation and the type of hydroxy polymers involved.

Synthesis studies have, so far, given an approximate outline of the probable processes of Al-hydroxide formation. The development of Al-hydroxide polymorphs appears to be related to the rate of precipitation, pH of the system, clay mineral surfaces, and the nature and concentration of inorganic and organic anions (Huang, 1988; Hsu, 1989). Only synthesis experiments related to pedogenic environments are summarized here. At room temperature, gibbsite forms in acid solutions (pH < 6) under slow hydrolysis, nordstrandite in neutral to alkaline solutions (pH > 7), and bayerite in alkaline solutions under fast hydrolysis (Hsu, 1966; Schoen and Roberson, 1970). These conditions agree with the natural occurrence of gibbsite in highly weathered acidic soils, whereas nordstrandite and bayerite have been found in association with limestone materials. Thermodynamically, gibbsite is the most stable of the three Al(OH)$_3$ polymorphs (Hemingway et al., 1991).

Anions that have strong affinity for Al^{3+}, such as SO$_4^{2-}$, CO$_3^{2-}$, PO$_4^{4-}$, and SiO$_3^{2-}$, may interfere with the crystallization of Al(OH)$_3$ (Huang, 1988; Hsu, 1989). Organic ligands like citric, malic, tannic, aspartic, and fulvic acids block the coordination sites of polynuclear Al ions and restrain the hydrolysis, inhibiting the crystallization of Al(OH)$_3$ and influencing the nature of the precipitated compound (Kodama and Schnitzer, 1980; Violante and Violante, 1980; Violante and Huang, 1985; Singer and Huang, 1990). This explains why little or no gibbsite is found in the A horizon of acidic soils high in organic matter and exchangeable Al, while it is found in greater quantities in deeper horizons.

The pathway for gibbsite formation by direct desilication of primary Al silicates or through clay mineral intermediates (Jackson, 1964) is conditioned by the intensity of leaching, which,

in turn, is affected by rainfall, temperature, parent rock, topography, groundwater table, vegetation, and time. Environments with warm temperatures, high rainfall, and free drainage favor desilication and leaching of ions, as well as mineralization of organic matter. Oxisols and laterites found in these environments may have significant amounts of gibbsite associated with kaolinite and Fe oxides (Curi and Franzmeier, 1984; Macedo and Bryant, 1987). Gibbsite is frequently a minor component in ultisols, which are widespread in humid tropical, subtropical, and temperate regions (Allen and Hajek, 1989). High amounts of gibbsite may be found under temperate climates in deeper parts of soil profiles submitted to higher leaching rates (Norfleet et al., 1993; Ogg and Baker, 1999). The porosity of the soil parent material is important in determining whether or not gibbsite forms. In Kenya, for example, soils formed in porous volcanic ash on the slopes of Mt. Kenya contain considerably more gibbsite than soils from western Kenya formed from hard igneous and metamorphic rocks (Obura, 2008).

Nordstrandite has been found in nature as crystals radiating into solution cavities in limestone from Guam (Hathaway and Schlanger, 1965) and as small pellets in a limestone soil from Borneo (Wall et al., 1962). Bayerite found in calcareous material in Israel probably formed under hydrothermal conditions (Gross and Heller, 1963). The apparent rarity of bayerite and nordstrandite in soils may reflect identification difficulties, due to their low concentration and/or masking by the presence of gibbsite.

The same factors may be responsible for the rare reports of boehmite in soils (Taylor, 1987). Boehmite has been identified in lateritic materials (Gilkes et al., 1973) and together with diaspore in bauxites (Vgenopoulos, 1984). Boehmite often occurs in bauxite deposits in association with gibbsite, from which it is considered to form through partial dehydration by diagenesis or hydrothermal alteration. Hughes et al. (1994) identified boehmite in native American artifacts from an Ordovician-age paleosol in northwestern Illinois. The formation of poorly crystalline boehmite has been attributed to the presence of foreign ions during crystallization (Hsu, 1967). Since poorly crystalline boehmite alters to crystalline Al(OH)$_3$ polymorphs on removal of the foreign ions from the environment, the occurrence of crystalline boehmite in soils, other than those derived from bauxite, may be less common than believed (Taylor, 1987).

Diaspore in association with minor amounts of goethite has been identified as a surface weathering product formed by the desilication of a kaolinitic clay in Missouri (Keller, 1978). The possible influence of Fe in diaspore formation is supported by synthesis experiments and natural rock weathering and bauxite occurrences (Taylor, 1987). The formation of Fe-substituted diaspore is, however, unlikely because the large Fe^{3+} does not fit into the diaspore structure, in contrast to the substitution of the smaller Al^{3+} in the larger goethite structure (Davies and Navrotsky, 1983). In confirmation, Fe in a diaspore crystal has been identified as a hematite-like cluster (Hazemann et al., 1992).

Corundum occurs in pegmatites and other rocks associated with nepheline syenites, in high-grade aluminous metamorphic rocks, in aluminous xenoliths in igneous rocks, in

metamorphosed bauxitic deposits, and as a detrital mineral in sediments (Deer et al., 1992). Reports of corundum in soils are scarce, but because of its high resistance to weathering, it can be found as a residual mineral of igneous or thermally metamorphosed rocks, like in lateritic bauxites of Australia (Taylor et al., 1983) and in Brazilian Oxisols (Macedo and Bryant, 1987). In some Australian soils derived from corundum-free rocks, the presence of corundum is attributed to the heating of the soil by bush fires (Wells et al., 1989). The corundum may have formed in association with hematite by the dehydroxylation of Al-goethite. The associated crystallization of isostructural hematite may have acted as a template for the nucleation of corundum, thereby reducing the activation energy and enabling its formation at relatively low temperature via the epitaxial growth on hematite.

22.4.3 Influence on Soil Properties

Because of the high surface area (100–220 m^2 g^{-1}), high PZC (pH 5–9), and a variable surface charge, Al oxides favorably chemisorb heavy metals (Cu, Pb, Zn, Ni, Co, Cd) and anions (PO$_4^{4-}$, SiO$_3^{2-}$, MoO$_4^{2-}$, SO$_4^{2-}$, catechols) (Hingston et al., 1974; McBride and Wesselink, 1988; McBride, 1989). The adsorption is related to the reactive (singly coordinated OH$^-$ groups at crystallite edges) rather than the total surface area. The high P retention in Oxisols and Ultisols is usually related to the content of gibbsite, in addition to that of Fe oxides (Juo, 1981; van Wambeke, 1992). Synthesis studies have shown the effectiveness of Al(OH)$_3$ precipitates in improving the physical properties of soils and clay minerals (see Hsu, 1989). Aluminum oxides, mainly gibbsite, in addition to Fe oxides, are strongly associated with the aggregation of Oxisols and Ultisols, but the mechanism of aggregation is unclear.

22.4.4 Identification

Crystalline Al oxides can be identified and quantified by XRD, thermal analysis, IR, and EM with a combination of several methods being necessary. Currently, no selective dissolution methods are known for the specific dissolution of crystalline and noncrystalline Al forms in soil samples, but acid NH$_4^+$ oxalate treatment (Schwertmann, 1964) is frequently used to estimate the content of undefined, poorly crystalline Al forms in soils. The problems inherent in the identification of poorly crystalline Al oxides in the presence of clay minerals are illustrated by Violante and Huang (1994). By XRD analysis, the detection limit of poorly crystalline boehmite in samples containing kaolinite or montmorillonite was approximately 30%–40%. When both kaolinite and montmorillonite were present, however, the identification by DTA, IR, or TEM often failed even for concentrations of boehmite as high as 50%–80%.

Under the electron microscope, well-crystallized gibbsite occurs as hexagonal plates (Hsu, 1989) or as elongated hexagonal rods in synthetic crystals (Tait et al., 1983). Euhedral gibbsite crystals from a bauxite zone in a deep lateritic weathering profile are shown in Figure 22.5. Bayerite frequently occurs as a triangular pyramid, with its long direction perpendicular to the basal
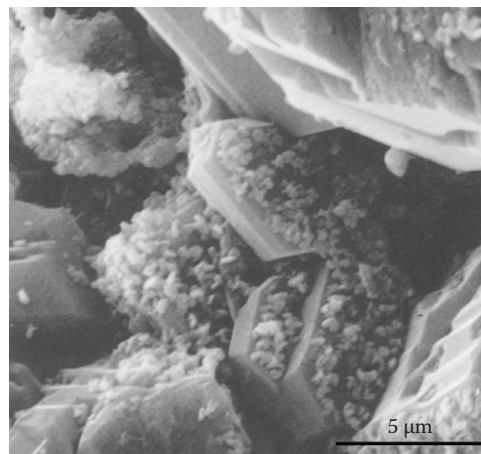
FIGURE 22.5   Scanning electron micrograph from a bauxitic zone in a deep lateritic profile from the Jos plateau, Nigeria, showing hematite crystal aggregates on large euhedral gibbsite crystals. (From Zeese, R., U. Schwertmann, G.F. Tietz, and U. Jux. 1994. Mineralogy and stratigraphy of three deep lateritic profiles of the Jos plateau (central Nigeria). Catena 21:195–214.)

plane (Schoen and Roberson, 1970; Violante and Jackson, 1981; Hsu, 1989). Synthetic nordstrandite shows rectangular plates, elongated parallelograms, ill-defined ovoidal particles, or clusters of acicular crystals (Shoen and Roberson, 1970; Violante and Jackson, 1979, 1981; Violante et al., 1982; Tait et al., 1983). With the exception of gibbsite, many of these crystal shapes seem to occur only in synthetic Al oxides.

22.4.4.1 X-Ray Diffraction

The main XRD peaks of the Al oxides are listed in Table 22.4. For specimens with poor crystallinity, these diffraction peaks may not be completely resolved. Gibbsite in soils can be quantified by XRD, but it will not be detected if its content is below about 50 g kg^{-1} (Jackson, 1969). The strongest peak for boehmite is the d(020) at 6.11 Å, while very fine boehmite particles show a broadening and an apparent line shift of the d(020) peak to larger d-spacings due to a diffraction effect (Tettenhorst and Hoffman, 1980).

22.4.4.2 Thermal Analysis

DTA is frequently used for the identification of Al hydroxides and oxyhydroxides (Mackenzie, 1970, 1972; Karathanasis and Harris, 1994). The presence of 10 g kg^{-1} gibbsite can be detected by DTA, and its amount can be estimated by comparing the 300°C–330°C endothermic peak with those of standards. The peak temperature for bayerite is only slightly lower than that for gibbsite, so that their differentiation in natural samples is impossible (Karathanasis and Harris, 1994). In soil samples, the 300°C–330°C peak may overlap with that of goethite, requiring an additional DTA curve after the removal of goethite with sodium dithionite (Jackson, 1969). Diaspore yields an endothermic peak at 540°C, whereas boehmite has a peak that occurs in the range of 450°C–580°C depending on crystallinity and particle size. The overlap with the kaolinite peak at 550°C limits the identification of these oxyhydroxides.

22.4.4.3 Infrared Absorption Analysis

Characteristic OH-stretching absorption spectra have been used to identify $Al(OH)_3$ and AlOOH polymorphs in monomineralic samples. Gibbsite shows three absorption bands and another doublet, with reported bands for well-crystallized natural gibbsites at 3622, 3627, 3460, 3396, and 3384 cm^{-1}. Each of the absorption bands may shift slightly to lower wave numbers with decreasing crystallinity (Elderfield and Hem, 1973). The 3622 cm^{-1} absorption band is coincident with that of kaolinite and micaceous clays (Farmer and Russel, 1967), thus precluding its use for the identification of gibbsite. Boehmite and diaspore each show two OH-stretching bands, boehmite at 3087 and 3283 cm^{-1} and diaspore at 2922 and 2990 cm^{-1} (Ryskin, 1974). Raman microprobe spectroscopy offers an alternative procedure in characterizing $Al(OH)_3$ polymorphs in soils (Rodgers, 1993).

22.5 Silicon Oxides

The Si oxides occurring in soils are given in Table 22.5. In nature, the SiO_2 polymorphs (quartz, tridymite, and cristobalite) occur as both higher temperature or β-phases and lower temperature or α-phases. In soils, however, only the α-phases are usually found. Opal is a hydrated form of Si of both biogenic and inorganic origin. Microcrystalline quartz, or microquartz, occurs in the form of chalcedony and chert (Heaney, 1994). Microcrystalline fibrous α-quartz, in the form of chalcedony, usually occurs as a secondary infilling of seams and cavities within rocks, sometimes creating concentrically banded agates or geodes. An intergrowth of authigenic microcrystalline α-quartz grains, which range in size from <1 to 50 μm, occurs in highly silicified sedimentary rocks in the form of chert (Knauth, 1994). More information about Si oxides can be found in Drees et al. (1989), Heaney et al. (1994), and Monger and Kelly (2002).

22.5.1 Mineral Phases

The Si oxides are tectosilicates. The repeating unit is a SiO_4 tetrahedron in which each O is linked to Si atoms of adjacent tetrahedra, forming a 3D framework structure. This contrasts with the Fe, Al, Mn, and Ti oxides in which the basic unit is a cation in octahedral coordination. The pattern of the tetrahedral linkage is different for each Si oxide polymorph, and this difference is reflected in their structural, physical, and chemical properties. The structures of the α-phases are closely related to those of their high-temperature β-phase equivalents. *Quartz* consists of paired helical chains of corner sharing SiO_4 tetrahedra that spiral along the Z-axis (Heaney, 1994). The intertwined chains produce open channels parallel to c (Z) that appear hexagonal in projection (Figure 22.1). In *tridymite* and *cristobalite*, the idealized fundamental module consists of a sheet containing six-member rings of SiO_4 tetrahedra, with the tetrahedra alternately pointing above and below the plane defined by the basal oxygen atoms (Heaney and Banfield, 1993). Tridymite and cristobalite are differentiated by a different stacking arrangement of this tetrahedral sheet. In tridymite, the sheets are stacked such that the hexagonal rings (ditrigonal and oval rings in α-tridymite structures) lie directly over one another, creating continuous tunnels normal to the sheets (Figure 22.1). These tunnels account for the lower density of tridymite (2.26 g cm^{-3}) as compared to quartz (2.65 g cm^{-3}). In cristobalite, the stacking involves three sheets that are translated relative to one another such that the hexagonal rings

TABLE 22.5 Silicon Oxide Minerals: Crystallographic Properties and d Values for the Six Most Intense Diffraction Lines

Mineral	Chemical Formula	Crystal System, Space Group	Unit-Cell Dimensions (Å)	Six Most Intense Diffraction Lines[a] d Value (Å), Relative Intensity						References[b]
α-Quartz	SiO_2	Trigonal, $P3_121$	a = 4.912 c = 5.403	3.342 100	4.257 22	1.818 14	1.542 9	2.457 8	2.282 8	Will et al. (1988, in Heaney, 1994); 33-1161
α-Tridymite	SiO_2	Monoclinic, $C222_1$	a = 10.04 b = 17.28 c = 8.20 β = 91°50′	4.08 100	4.28 93	3.80 68	3.24 48	2.48 35	2.382 21	Heaney (1994); 42-1401
α-Cristobalite	SiO_2	Tetragonal $P4_12_12$	a = 4.969 c = 6.925	4.04 100	2.487 13	2.841 9	3.136 8	2.467 4	1.929 4	Schmahl et al. (1992, in Heaney, 1994); 39-1425
Opal (natural)				4.08 100	2.51 30	2.86 10	3.14 9	1.937 5	1.878 5	38-0448
Opal-C			~α-Cristobalite + 4.3							Drees et al. (1989)
Opal-CT				4.10 vs	4.29 s	2.50 s	3.34 w	3.18 w	2.85 w	Drees et al. (1989)
Opal-A				4.1 sb	2.0 wd	1.5 wd	1.2 wd			Drees et al. (1989)

[a] vs, Very strong; s, strong; w, weak; sb, strong broad; wd, weak diffuse.
[b] Numbers of the format XX-XXXX indicate ICDD file number (ICDD, 1994).

(oval rings in α-cristobalite) do not superimpose (Figure 22.1). Thus, cristobalite lacks the continuous tunnels normal to the layers as in tridymite, but tunnel structures are formed parallel to the layers. Thus, the density of cristobalite is only slightly higher than that of tridymite ($2.32\,g\,cm^{-3}$).

Opal is classified into three structural groups based on XRD powder patterns: Opal-C (well-ordered α-cristobalite), opal-CT (disordered α-cristobalite, α-tridymite), and opal-A (highly disordered, nearly amorphous) (Jones and Segnit, 1971; Deer et al., 1992), in a sequence of decreasing structural order. In opal-C, the stacking is predominantly cristobalitic, which is revealed by weak superstructure XRD reflections characteristic of α-cristobalite. Opal-CT consists of random stackings of α-cristobalite- and α-tridymite-like arrangements, yielding a disordered crystal structure. The microstructure of opal-CT is composed of small spheres less than $5\,\mu m$ in diameter, called lepispheres, which consist of an interpenetrative growth of tiny cristobalite and tridymite blades (Hesse, 1988; Graetsch, 1994). Opal-A is a noncrystalline hydrous Si polymorph ($SiO_2 \cdot nH_2O$) differing from the crystalline polymorphs in that it lacks long-range atomic order. The microstructure of opal-A shows the closest packing of spheres of Si with diameters ranging from 10 to $50\,nm$ (Greer, 1969) with water filling the interstices (Graetsch, 1994). The specific gravity of biogenic opal from soils and plants varies continuously over the range 1.5–2.3, with modal values from 2.10 to $2.15\,g\,cm^{-3}$ (Drees et al., 1989). The broad range is a function of the submicron opal structure, H_2O content, occluded organic matter, and microscopic voids.

Quartz is one of the purest minerals known, and due to its more closed structure, it is purer than the other Si oxide polymorphs (Drees et al., 1989). Nevertheless, common trace elements in quartz, either interstitial or as substitutions, occur and include Al, Ti, Fe, Na, Li, K, Mg, Ca, OH. Both cristobalite and tridymite, which are also the major structural constituents of opal-CT, may accommodate more impurities than quartz because of their more open framework structure. The predominant impurity is Al, with lesser amounts of Fe, Ti, K, Na, Ca, and Mg. Opal-A contains 850–950 g kg^{-1} of SiO_2, 40–90 g kg^{-1} of bound H_2O, and significant amounts of occluded, chemisorbed, or solid-solution impurities of Al, Fe, Ti, Mn, P, Cu, N, C, alkalies, and alkaline earths (Drees et al., 1989).

22.5.2 Occurrence and Formation

Quartz is a common constituent in many igneous, sedimentary, and metamorphic rocks, and also occurs as a secondary mineral, forming cementing materials in sediments. Tridymite is a typical mineral of acid volcanic rocks together with cristobalite that also occurs in basaltic rocks. Tridymite is also a common constituent of highly metamorphosed impure limestones and arkoses adjacent to basic igneous intrusions, while cristobalite occurs in metamorphosed sandstones and sandstone xenoliths in basic rocks. Opal is found in sedimentary, volcanic, and marine environments (Deer et al., 1992). Thus, quartz is the most abundant Si oxide in soil environments, cristobalite occurs

in soils developed from volcanic materials, opal may be a significant component of soils, but tridymite occurs only rarely.

Quartz in soils is mainly a primary mineral, inherited from the parent material. The higher stability of quartz compared to other silicate minerals is explained by its crystallization from magma closer to present earth-surface conditions. The higher stability of quartz relative to the other Si oxide polymorphs is due to a denser packing of the crystal structure and a higher energy required to break the Si–O–Si bond. Quartz in chert forms through a diagenetic transformation from opal-A via opal-CT and opal-C (Hesse, 1988). As demonstrated experimentally, quartz may also precipitate directly (Mackenzie and Gees, 1971). Authigenic quartz in the form of chemical precipitations and grain overgrowths frequently occurs in sediments (carbonates, sandstones) and is believed to form in soils, too (Drees et al., 1989).

Quartz is generally concentrated in the sand and silt fractions of soils, with a lower frequency in the coarse-clay fraction ($0.2–2\,\mu m$). The depth distribution of quartz is a function of parent material and degree of weathering (Drees et al., 1989). In relatively undifferentiated soils (Entisols and Inceptisols), distributions of quartz reflect mainly variations in parent material. Quartz may comprise >90% of the inorganic fraction of Quartzipsamments. In Aridisols, quartz is generally abundant due to restricted weathering, while in moderately weathered soils (Spodosols, Alfisols, Mollisols, and Ultisols), the eluvial horizons are enriched in quartz relative to the parent material due to the weathering and removal of less resistant minerals. Conversely, illuvial horizons have lower quartz contents than eluvial horizons or parent materials due to the dilution by silicate clays, carbonates, or oxides. Highly weathered and leached Oxisols have only very low amounts of quartz left in the clay fraction.

Opal-CT, opal-C, and α-cristobalite are usually limited to specific geographic regions and rock-stratigraphic units. Opal-C and α-cristobalite have been reported in Andepts and other soils derived from volcanic material in South America, New Zealand, Central America, and Japan. Opal-CT has been identified in bentonites, siliceous shales, and indurated silicates. As a consequence, soils developed in these materials commonly contain opal-CT. It is also a primary constituent of many cherts, porcelanites, fossil wood, silcretes, and some duripans (Drees et al., 1989). According to synthesis experiments and evidence from diagenetic environments, opal-A transforms to opal-CT and quartz by a series of dissolution–precipitation reactions (Kastner et al., 1977; Williams and Crerar, 1985; Williams et al., 1985, Cady et al., 1996). The conversion is dependent on temperature, pressure, pH, pore solution chemistry, and specific surface area. As the transformation is mediated by clays, oxides, and carbonates, it is assumed to occur in soils (Drees et al., 1989).

Opal-A may form by both organic and inorganic processes in pedogenic environments. Biogenic opal-A originates from Si accumulated by plants and aquatic organisms, and thus occurs under a wide range of environmental conditions (Jones and Beavers, 1964; Wilding and Drees, 1971).
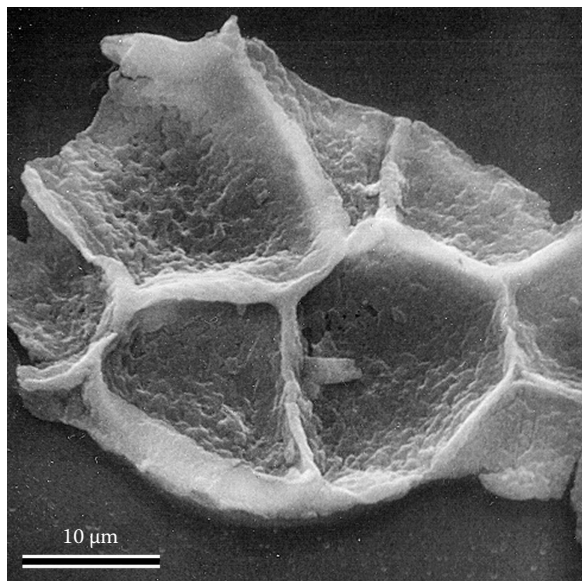
FIGURE 22.6 Scanning electron micrograph of an opal phytolith isolated from a forest soil. (From Drees, L.R., L.P. Wilding, N.E. Smeck, and A.L. Senkayi. 1989. Silica in soils: Quartz and disordered silica polymorphs, p. 913–974. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.)

Biogenic opal is a minor but nearly ubiquitous constituent of soils, with opal phytoliths (Figure 22.6) representing the major form in nonaquatic environments, while sponge spicules, diatoms, and radiolaria are the main opal form in aquatic environments and limestone residuum (Jones and Beavers, 1964; Wilding and Drees, 1971). Amounts of opal phytoliths in soils commonly range from <1 to 30 g kg^{-1}, usually occurring in the 5–20 and 20–50 μm size fractions and decreasing with soil depth (Drees et al., 1989). Phytoliths are recognized as the principal sink and source of Si in soil solutions (Lucas, 2001; Farmer et al., 2005).

Opal-A, which forms when the conditions favor supersaturation of Si in soil solution (Jones and Handreck, 1967), occurs as nodules and as a primary cement for indurated soil horizons (duripans) and silcretes. Such soils usually have pH and moisture regimes favoring solubilization and translocation of Si to lower positions in the soil profile. Geographically, duripans are restricted to materials that provide soluble Si, such as pyroclastics and intermediate and basic igneous rocks (Flach et al., 1969; Soil Survey Staff, 1975). According to Flach et al. (1969), pedogenic transformation of opal to quartz is believed to occur in duripans and in Si-cemented soils. Silica released by weathering of silicate minerals may precipitate as amorphous SiO$_2$ on soil particles, forming bridges between mineral grains, leading to the hardness and brittleness of fragipans (Franzmeier et al., 1989; Karathanasis, 1989).

At ambient temperature and neutral pH, the solubility of amorphous Si in soils is approximately 50–60 mg Si L^{-1} and that of quartz is commonly 3–7 mg Si L^{-1}. The solubility of Si oxides is a function of temperature, pH, particle size, chemical composition, and the presence of a disrupted surface layer. For both crystalline and amorphous Si polymorphs, it is essentially constant between the pH limits of 2 and 8.5, increasing rapidly above pH 9 due to the ionization of monosilicic acid.

The presence of organic molecules greatly enhances the dissolution rates of Si oxides. The dissolution of biogenic opal and quartz increases as particle size decreases due to the increase in surface area (Drees et al., 1989). Chemisorption of metallic ions such as Al, Fe, Mg, Ca, Cu, Pb, and Hg to Si surfaces reduces the dissolution rates of Si due to the formation of relatively insoluble Si coatings (Drees et al., 1989), thus increasing the stability of Si oxides. Oxides of Fe and Al, by acting as a soluble Si sink, greatly increase the dissolution rates of amorphous Si oxides and reactive uncoated quartz surfaces. Dissolution (weathering) of Si oxides is induced by the reduction of Si levels in soil solution by leaching and plant uptake. Dissolution of quartz initiates below approximately 3 mg of Si L^{-1} (Kittrick, 1969).

22.5.3 Influence on Soil Properties

Noncemented soils comprised dominantly of quartz are nonplastic due to the weak cohesion (van der Waals forces) that develops between Si particles and show low water-retention capacity and high hydraulic conductivity. Duripans, cemented by even small amounts of Si, are hard to extremely hard when dry, have high unconfined compressive strength, and resist dispersion in Na hexametaphosphate (Flach et al., 1974).

Due to their essentially uncharged, poorly hydrated surface, crystalline Si oxides have relatively small effects on the physicochemical activity affecting the soil–plant relationship and act as a diluent to the much more reactive clay minerals and oxides of Fe, Al, and Mn. Silica is not essential for the growth of most plants, but orthosilicic acid, H$_4$SiO$_4$, has a beneficial effect on the growth of some plants (Jones and Handreck, 1967). Rice and sugarcane, for example, often respond to fertilization with very soluble Si sources such as Si slag.

Due to its ubiquity, abundance, resistance to weathering, and immobility, quartz has been often used as an index mineral for weathering and soil formation (Barshad, 1964; Sudom and St. Arnaud, 1971). Extensive leaching may, however, drastically increase the solubility of primary quartz (Little et al., 1978; Asumadu et al., 1988; Pye and Mazzullo, 1994), and authigenic (Breese, 1960) as well as biogenic quartz may form in soils (Wilding and Drees, 1974), thus restricting the use of quartz as an index mineral.

The isotopic composition of quartz (^{18}O), which depends on the temperature of formation, has been used to identify eolian additions to soils (Clayton et al., 1972; Mizota and Matsuhisa, 1995) and the formation of authigenic opaline Si in volcanic ash soils (Mizota et al., 1991).

Opal phytoliths have been used to trace the origin of colluvial sediments (Lutwick and Johnston, 1969) and to identify surface horizons of paleosols (Dormaar and Lutwick, 1969). Depth distributions of sponge spicules may be useful to identify lithological discontinuities, and opal of aquatic organisms (diatoms,

sponge spicules, and siliceous shells) provides direct evidence that the parent material of the soil is of marine or lacustrine origin (Jones and Beavers, 1963, 1964; Wilding and Drees, 1968, 1971). Biogenic opal has been extensively used to reconstruct the vegetative history of soils (Drees et al., 1989; Fisher et al., 1995; Staller, 2002). The $^{13}C/^{12}C$ isotope ratio of C occluded in opal phytoliths was used to establish the succession of C3 and C4 grasses, providing a quantitative method for monitoring climatic changes (Kelly et al., 1991).

22.5.4 Identification

Silica oxides can be identified by XRD, SEM, microprobe analysis, IR, and light microscopy. Opal-CT and opal-A may need to be concentrated by particle-size fractionation, specific gravity, and/or differential dissolution. Counting particle size separates under the light microscope is the simplest method of quantifying Si oxides but is limited in resolution (5–10 μm) (Drees et al., 1989). A review of methodologies for extraction of amorphous Si from soils and aquatic sediments is given by Sauer et al. (2006) and Sacconne et al. (2007).

22.5.4.1 Crystal Form, Habit, and Color

In soils, quartz generally occurs as anhedral grains, rarely exhibiting the characteristic prismatic habit of macrocrystals. Angular grains are generally the result of mechanical fracturing, while rounding is a result of physical attrition during both transport and solution. Overgrowths on quartz grains may suggest long-term stable environments. Particles <100 μm are commonly shaped like flat plates due to cleavage (Drees et al., 1989). Thus, quartz grain surface morphology may be used to elucidate mineral origin and past and/or present chemical and physical environment.

Opal-CT is commonly observed in sediments as small (<10 μm) lepispheres. The morphology of opal-A of biogenic origin is closely related to the biological cell or structure in which it originates: opal of forest origin consisting of cellular incrustations with numerous thin-sheet structures (Figure 22.6); opal of grass consisting of solid polyhedral structures resulting from the silicification of the entire cell; and opaline microfossils of sponge spicules, diatoms, and shells (Drees et al., 1989).

Pure Si oxides are colorless but chemical impurities may impart various colors. Quartz is usually colorless and transparent, or white, with a vitreous luster. Cristobalite is white to milky white and ranges from translucent to opaque. The color and degree of translucency of opal-CT seem to depend on the aggregation of lepispheres. For opal-A, color is not a diagnostic criterion because it is strongly affected by occluded chemical impurities and light interference and scattering. In transmitted light, biogenic opal isolated from soils ranges from colorless or light tan to various shades of brown or black (Drees et al., 1989).

22.5.4.2 X-Ray Diffraction

The diagnostic parameters for Si oxides are presented in Figure 22.7 and Table 22.5. Quartz is easily identified by its most intense reflections at 4.26 and 3.34 Å, even in mixed mineral
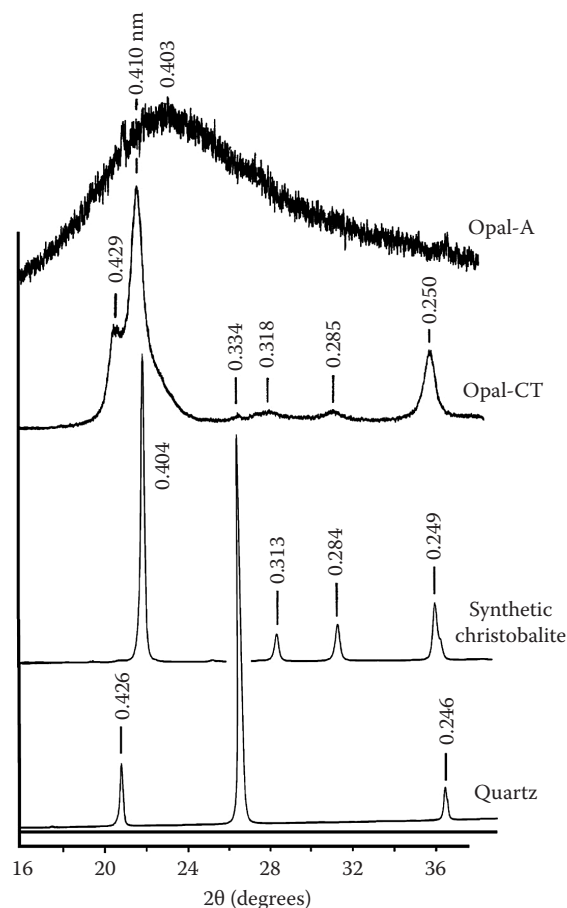
FIGURE 22.7 XRD patterns for quartz, synthetic cristobalite, opal-CT, and opal-A (Cu Kα radiation, peak position in nanometers). (Modified from Drees, L.R., L.P. Wilding, N.E. Smeck, and A.L. Senkayi. 1989. Silica in soils: Quartz and disordered silica polymorphs, p. 913–974. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.)

assemblages. The accuracy of the quantitative determination of quartz depends on sample preparation and the diffraction peak used (Rowse and Jepson, 1972). For well-ordered or α-cristobalite, the 4.04 Å peak is the main clue in mixed samples provided that no feldspars are present. Opal-C gives a pattern resembling that of α-cristobalite, except for slight line broadening and minor evidence of tridymite stacking. Opal-CT usually exhibits broad cristobalite peaks at about 4.1 and 2.5 Å and a peak at 4.3 Å, which is attributed to tridymite. The wide variations in the intensity and line profiles make its quantification by XRD difficult. The short-range order in opal-A produces a diffuse, broad x-ray hump centered at about 4.0 Å.

22.5.4.3 Thermal Properties

Quartz exhibits a sharp endothermic DTA peak at about 570°C (Drees et al., 1989), which may not be evident for microcrystalline quartz (<0.05 μm). Well-ordered synthetic cristobalite has a characteristic endothermic peak at about 260°C, whereas opal-CT and opal-A do not yield a characteristic endotherm below 600°C.

According to Rowse and Jepson (1972), DTA is better than either XRD or chemical techniques for detecting small quantities of quartz in clay materials.

22.5.4.4 Infrared Spectroscopy

All tectosilicates are characterized by Si–O–Si stretching vibrations between 950 and 1200 cm^{-1} and by O–Si–O bending vibrations between 400 and 550 cm^{-1}. Quartz is distinguished from the other Si polymorphs by a distinctive absorption band at 692 cm^{-1} and two strong doublets, one at 395 and 370 cm^{-1} and the other at 800 and 780 cm^{-1} (Drees et al., 1989). The latter doublet has been used also for the quantification of quartz (Chester and Green, 1968). This method is, however, restricted to samples <1 μm in particle size and free of other Si polymorphs.

The disordered Si polymorphs, opal-C, opal-CT, and opal-A, have a single band around 800 cm^{-1} in common. Well-ordered cristobalite and opal-C show additional bands at 620 and 380 cm^{-1}. Opal-A of biogenic origin is characterized by much weaker and broader absorption bands at 460 and 1100 cm^{-1} as compared to the better crystallized polymorphs and an additional weak band at 965 cm^{-1} (Drees et al., 1989).

22.6 Titanium and Zirconium Minerals

Of all Ti and Zr compounds in soil and weathered materials, the Ti oxides (rutile, anatase, and ilmenite) and the Zr silicate (zircon) are the most ubiquitous. The principal Ti and Zr minerals are listed in Table 22.6. Additional information about Ti and Zr minerals can be found in Milnes and Fitzpatrick (1989), Lindsley (1976, 1991), and Fitzpatrick and Chittleborough (2002). Titanomagnetite and titanomaghemite are described in Section 22.2.

22.6.1 Mineral Phases

Titanium occurs primarily in octahedral coordination, and like the Fe and Mn oxides, the structures of the various Ti oxide minerals can be described in terms of the different arrangement of the Ti-containing octahedra (Lindsley, 1976; Waychunas, 1991; Heaney and Banfield, 1993).

Rutile (TiO$_2$) is isostructural with pyrolusite and manganite, consisting of single chains of edge-sharing TiO$_6$ octahedra (Figure 22.1; Heaney and Banfield, 1993). *Anatase* consists of TiO$_6$ octahedra that share four O–O edges, two at the top and two at right angles at the bottom. The octahedra outline a 3D framework, rather than distinct chains (Figure 22.1; Waychunas, 1991). Schwertmann et al. (1995b) reported the substitution of Ti^{4+} by Fe^{3+} up to 0.1 mol fraction in pedogenic and synthetic anatase. Charge compensation is achieved by the incorporation of structural OH. *Brookite* has a more complex structure than rutile or anatase, consisting of deformed TiO$_6$ octahedra sharing three O–O edges to form staggered, cross-linked chains that are oriented along the c-axis (Figure 22.1; Lindsley, 1976). *Ilmenite* (FeTiO$_3$) is almost isostructural with hematite, with one-half of the Fe atoms replaced by Ti, so that the Fe^{3+}–O$_3$–Fe^{3+} units in hematite become Fe^{2+}–O$_3$–Ti^{4+} units in ilmenite (Figure 22.1; Lindsley, 1976). *Pseudobrookite* (ideally Fe$_2$TiO$_5$) has strongly distorted octahedra but can be described in terms of an ideal cubic close packing anion arrangement (Figure 22.1; Waychunas, 1991). The two types of octahedral sites are M1 and M2 in the ratio 1:2, ideally occupied by Fe^{3+} and Ti^{4+}, respectively. *Pseudorutile* (Fe$_2$O$_3$·nTiO$_2$·mH$_2$O; $3 < n < 5$ and $1 < m < 2$) is a structurally disordered and poorly characterized mineral formed by the alteration of ilmenite. Its structure is based on a hexagonal closest packing of O^{2-} anions and has been described as an intergrowth of a rutile-type structure with a goethite-type

TABLE 22.6 Titanium and Zr Oxide Minerals: Crystallographic Properties and d Values for the Six Most Intense Diffraction Lines

Mineral	Chemical Formula	Crystal System, Space Group	Unit-Cell Dimensions (Å)	Six Most Intense Diffraction Lines d Value (Å), Relative Intensity						References[a]
Rutile	TiO$_2$	Tetragonal, *P4$_2$/mnm*	a = 4.5933	3.247	1.687	2.487	2.188	1.624	1.36	21-1276
			c = 2.9592	100	60	50	25	20	20	
Anatase	TiO$_2$	Tetragonal, *I4$_1$/amd*	a = 3.7853	3.52	1.892	2.378	1.70	1.666	1.481	21-1272
			c = 9.5139	100	35	20	20	20	14	
Brookite	TiO$_2$	Orthorhombic, *Pbca*	a = 5.4558	3.152	2.90	3.465	1.893	1.66	2.476	29-1360
			b = 9.1819	100	90	80	30	30	25	
			c = 5.1429							
Ilmenite	Fe^{2+}TiO$_3$	Hexagonal, *R3*	a = 5.0884	2.754	2.544	1.726	1.868	1.468	3.737	29-0733
			c = 14.093	100	70	55	40	35	30	
Pseudobrookite	Fe$_2$$^{3+}TiO_5$	Orthorhombic, *Bbmm*	a = 9.7965	3.486	2.752	4.90	2.458	1.865	2.407	41-1432
			b = 9.9805	100	77	42	23	23	22	
			c = 3.7301							
Pseudorutile	FeIII$_2$Ti$_3$O$_8$(OH)$_2$	Hexagonal, *P6$_3$22*	a = 14.486	3.50	2.66	2.51	1.687	3.67	3.23	Waychunas
			c = 4.467	100	90	80	70	40	30	(1991)
Ulvöspinel	Fe$_2$TiO$_4$	Cubic, *Fd3m*	a = 8.393–8.536	2.573	1.509	3.018	1.642	2.134	1.742	34-0177
				100	39	33	33	19	10	
Zircon	ZrSiO$_4$	Tetragonal, *I4$_1$/amd*	a = 6.604	3.30	4.434	2.518	1.712	2.066	1.908	6-266
			c = 5.979	100	45	45	40	20	14	

[a] Numbers of the format XX-XXXX indicate ICDD file number (ICDD, 1994).

structure (Figure 22.1; Grey et al., 1983). *Zircon* ($ZrSiO_4$) is an orthosilicate with a structure consisting of chains of alternating edge-sharing SiO_4 tetrahedra and $[ZrO_8]$ triangular dodecahedra. The chains are joined laterally by dodecahedra-sharing edges in an arrangement that produces unoccupied octahedral voids (Speer, 1982).

22.6.2 Occurrence and Formation

Rutile, anatase, ilmenite, and less frequently brookite are found as accessory minerals in many igneous and metamorphic rocks and as detrital minerals in sediments, although anatase is often also of authigenic origin (Deer et al., 1992). The Ti and Zr minerals in soils may be residual minerals inherited from the parent material, formed through weathering of Ti- and Zr-bearing minerals, or authigenic (Milnes and Fitzpatrick, 1989). Rutile, anatase, ilmenite, and zircon are commonly residual minerals occurring in the sand and silt fractions of a variety of soils. Ilmenite may weather to pseudorutile and mixtures of rutile, anatase, and Fe oxides. Some evidence indicates that anatase and ilmenite weathering occurs by organic acids in A horizons of Scottish podzols (Berrow et al., 1978), and weathering features were observed on the surface of zircon and rutile grains of Australian Spodic Quartzipsamments (Tejan-Kella et al., 1991).

Many examples of secondary Ti oxides formed through weathering of primary minerals are found in saprolites and in soils. The data on the secondary or authigenic formation of Zr minerals are, however, controversial (Milnes and Fitzpatrick, 1989). The alteration of ilmenite under oxidizing conditions leads to the formation of pseudorutile (Grey and Reid, 1975; Anand and Gilkes, 1984b) but, at this stage, there is no clear evidence that it also takes place under pedogenic conditions (Berrow et al., 1978). Authigenic formation of poorly crystalline anatase as an alteration product of sphene was observed in peaty podzols from Scotland (Berrow et al., 1978). In conclusion, basic aspects of the Ti and Zr minerals, such as whether the minerals are relict or pedogenic, and the conditions responsible for their weathering and formation in soils, are still unresolved (Taylor, 1987; Milnes and Fitzpatrick, 1989).

22.6.3 Influence on Soil Properties

Because of their generally low concentrations in soils, there is little evidence of the effect of Ti and Zr minerals on soil reactivity. Only in tropical soils where their concentrations are much higher some effects may be expected. The broken-edge bonds of anatase are hydroxylated and can exhibit variable charge characteristics (Fitzpatrick et al., 1978). The rutile and anatase surface has hydroxyl groups with different reactivities (Tanaka and White, 1982) that can adsorb and retain phosphate and arsenate (Cabrera et al., 1977; Fordham and Norrish, 1983).

Because ilmenite and zircon are generally very resistant in soils, they may be used as reference minerals in weathering and soil genesis studies (Bleeker, 1972; Mitchell, 1975; Claridge and Weatherhead, 1978; Tejan-Kella et al., 1991). The immobility of these minerals must, however, be previously assured (Colin et al., 1993; Anda et al., 2009).

22.6.4 Identification

As Ti and Zr minerals can occur in different particle-size fractions and exhibit variable crystallinity, a variety of techniques may be necessary to identify and characterize them. Optical microscopy is a useful technique for examining and identifying these minerals, both in undisturbed form in thin sections and as components separated by physical or chemical techniques (Mitchell, 1975). A combination of XRD, IR absorption, DTA, and electron microprobe techniques is useful for identifying the Ti and Zr minerals in the sand and silt fractions of soils but may be less satisfactory for the clay fractions due to the interference of layer silicates (Milnes and Fitzpatrick, 1989). TEM along with energy-dispersive x-ray analysis of individual particles can be used to identify Ti oxide crystals in soil clay fractions (Figure 22.8). Titanium oxides may be concentrated in kaolinitic soils by dissolution of clay minerals in boiling 5 M NaOH alone (Norrish and Taylor, 1961) or by a combination of boiling 5 M KOH and DCB treatments (Zeese et al., 1994). Amorphous Ti oxides can be separated from more crystalline forms by extraction in acid ammonium oxalate (Fitzpatrick et al., 1978). Minor amounts of anatase (>0.02%) can be identified by Raman spectroscopy (Murad, 1997). Tejan-Kella et al. (1991) described an SEM method to characterize microtextural features of zircon and rutile grains. In XRD of the clay fraction, the strong, sharp anatase peak at 3.52 Å (Table 22.6) is often clearly resolved after the interfering 3.59 Å kaolinite line has been eliminated by heating at 550°C.
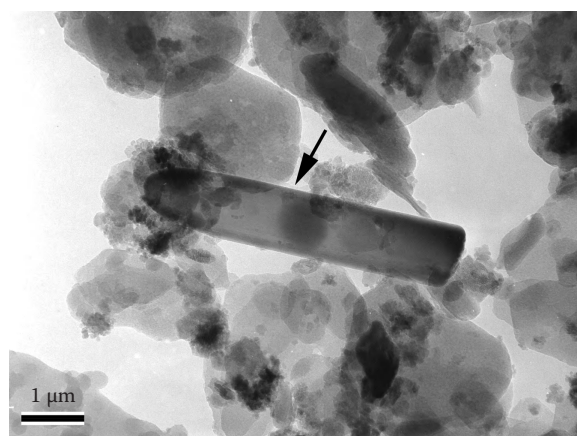
FIGURE 22.8 Transmission electron micrograph of the 2–0.2 μm fraction of a Rhodic Hapludox showing an anatase or rutile particle (arrow) surrounded by kaolinite and iron oxide particles. (Modified from de Brito Galvão, T.C., and D.G. Schulze. 1996. Mineralogical properties of a collapsible lateritic soil from Minas Gerais, Soil Sci. Soc. Am. J. 60:1969–1978.)

References

Adams, S.N., J.L. Honeysett, K.G. Tiller, and K. Norrish. 1969. Factors controlling the increase of cobalt in plants following the addition of a cobalt fertilizer. Aust. J. Soil Res. 7:29–42.

Adams, W.A., and J.K. Kassim. 1984. Iron oxyhydroxides in soils developed from lower palaeozoic sedimentary rocks in mid-Wales and implications for some pedogenetic processes. J. Soil Sci. 35:117–126.

Allen, B.L., and B.F. Hajek. 1989. Mineral occurrence in soil environments, p. 199–278. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Anand, R.R., and R.J. Gilkes. 1984a. Mineralogical and chemical properties of weathered magnetite grains from lateritic saprolite. J. Soil Sci. 35:559–567.

Anand, R.R., and R.J. Gilkes. 1984b. Weathering of ilmenite in a lateritic pallid zone. Clays Clay Miner. 32:363–374.

Anand, R.R., and R.J. Gilkes. 1987a. Variations in the properties of iron oxides within individual specimens of lateritic duricrust. Aust. J. Soil Res. 25:287–302.

Anand, R.R., and R.J. Gilkes. 1987b. The association of maghemite and corundum in Darling Range laterites, Western Australia. Aust. J. Soil Res. 25:303–311.

Anda, M., D.J. Chittleborough, and R.W. Fitzpatrick. 2009. Assessing parent material uniformity of a red and black soil complex in the landscapes. Catena 78:142–153.

Asumadu, K., R.J. Gilkes, T.M. Armitage, and H.M. Churchward. 1988. The effects of chemical weathering on the morphology and strength of quartz grains—An example from S.W. Australia. Eur. J. Soil Sci. 39:375–383.

Bardossy, G., and G.W. Brindley. 1978. Rancieite associated with a karstic bauxite deposit. Am. Mineral. 63:762–767.

Barshad, I. 1964. Chemistry of soil development, p. 1–70. *In* F.E. Bear (ed.) Chemistry of soil. Reinhold Publishing Corporation, New York.

Barthès, B.G., E. Kouakoua, M.-C. Larré-Larrouy, T.M. Razafimbelo, E.F. de Luca, A. Azontonde, C.S.V.J. Neves, P.L. de Freitas, and C.L. Feller. 2008. Texture and sesquioxide effects on water-stable aggregates and organic matter in some tropical soils. Geoderma 143:14–25.

Bascomb, C.L. 1968. Distribution of pyrophosphate-extractable iron and organic carbon in soils. J. Soil Sci. 19:251–268.

Baur, W.H. 1976. Rutile-type compounds. V. Refinement of MnO_2 and MgF_2. Acta Crystallogr. B 32:2200–2204.

Bernal, J.D., D.R. Dasgupta, and A.L. Mackay. 1959. The oxides and hydroxides of iron and their structural interrelationships. Clay Miner. Bull. 4:15–30.

Berrow, M.L., M.J. Wilson, and G.A. Reaves. 1978. Origin of extractable titanium and vanadium in the A horizon of Scottish podzols. Geoderma 21:89–103.

Berthelin, J., G. Ona-Nguema, S. Stemmler, C. Quantin, M. Abdelmoula, and F. Jorand. 2006. Bioreduction of iron species and biogenesis of green rusts in soils. C.R. Geosci. 338:447–455.

Bigham, J.M., L. Carlson, and E. Murad. 1994. Schwertmannite, a new iron oxyhydroxy-sulfate from Pyhäsalmi, Finland and other localities. Mineral. Mag. 58:641–648.

Bigham, J.M., R.W. Fitzpatrick, and D.G. Schulze. 2002. Iron oxides, p. 323–366. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Bigham, J.M., D.C. Golden, L.H. Bowen, S.W. Buol, and S.B. Weed. 1978. Mössbauer and X-ray evidence for the pedogenic transformation of hematite to goethite. Soil Sci. Soc. Am. J. 42:979–981.

Bigham, J.M., U. Schwertmann, and L. Carlson. 1992. Mineralogy of precipitates formed by biogeochemical oxidation of Fe(II) in mine drainage. Catena Suppl. 1:219–232.

Bigham, J.M., U. Schwertmann, L. Carlson, and E. Murad. 1990. A poorly crystallized oxyhydroxide of iron formed by bacterial oxidation of Fe(II) in acid mine waters. Geochim. Cosmochim. Acta 54:2743–2758.

Bigham, J.M., U. Schwertmann, and G. Pfab. 1996a. Influence of pH on mineral speciation in a bioreactor simulating acid mine drainage. Appl. Geochem. 11:845–849.

Bigham, J.M., U. Schwertmann, S.J. Traina, R.L. Winland, and M. Wolf. 1996b. Schwertmannite and goethite solubilities and the chemical modelling of iron in acid sulfate waters. Geochim. Cosmochim. Acta 60:2111–2121.

Birch, W.D., A. Pring, A. Reller, and H. Schmalle. 1993. Bernalite, $Fe(OH)_3$, a new mineral from Broken Hill, New South Wales: Description and structure. Am. Mineral. 78:827–834.

Birnie, A.C., and E. Paterson. 1991. The mineralogy and morphology of iron and manganese oxides in an imperfectly-drained Scottish soil. Geoderma 50:219–237.

Blake, R.L., R.E. Hessevick, T. Zoltai, and L.W. Finger. 1966. Refinement of the hematite structure. Am. Mineral. 51:123–129.

Bleeker, P. 1972. The mineralogy of eight latosolic and related soils from Papua-New Guinea. Geoderma 8:191–205.

Blume, H.P., and U. Schwertmann. 1969. Genetic evaluation of the profile distribution of aluminum, iron and manganese oxides. Soil Sci. Soc. Am. Proc. 33:438–444.

Borggaard, O.K. 1988. Phase identification by selective dissolution techniques, p. 83–98. *In* J.W. Stucki, B.A. Goodman, and U. Schwertmann (eds.) Iron in soils and clay minerals. D. Reidel Publishing Co., Dordrecht, the Netherlands.

Boudeulle, M., and J.-P. Muller. 1988. Structural characteristics of hematite and goethite and their relationship with kaolinite in a laterite from Cameroon. A TEM study. Bull. Minér. 111:149–166.

Breese, G.F. 1960. Quartz overgrowths as evidence of silica deposition in soils. Aust. J. Sci. 23:18–20.

Bricker, O. 1965. Some stability relations in the system Mn–O_2–H_2O at 25°C and one atmosphere total pressure. Am. Mineral. 50:1296–1354.

Brindley, G.W., and D.L. Bish. 1976. Green rust: A pyroaurite type structure. Nature 263:353.

Burns, R.G., V.M. Burns, and H.W. Stockman. 1985. The todor-okite–buserite problem: Further considerations. Am. Mineral. 70:205–208.

Busing, W.R., and H.A. Levy. 1958. A single crystal neutron diffraction study of diaspore, AlO(OH). Acta Crystallogr. 11:798–803.

Cabrera, F., L. Madrid, and P. de-Armbarri. 1977. Adsorption of phosphate by various oxides: Theoretical treatment of the adsorption envelope. J. Soil Sci. 28:306–313.

Cady, S.L., H.-R. Wenk, and K.H. Downing. 1996. HRTEM of microcrystalline opal in chert and porcelanite from the Monterey Formation, California. Am. Mineral. 81:1380–1395.

Cambier, P. 1986. Infrared studies of goethites of varying crystallinity and particle size: I. Interpretation of OH and lattice vibration frequencies. Clay Miner. 21:191–200.

Campbell, A.S., and U. Schwertmann. 1984. Iron oxide mineralogy of placic horizons. J. Soil Sci. 35:569–582.

Carlson, L., and U. Schwertmann. 1980. Natural occurrence of feroxyhite (δ′-FeOOH). Clays Clay Miner. 28:272–280.

Carlson, L., and U. Schwertmann. 1981. Natural ferrihydrites in surface deposits from Finland and their association with silica. Geochim. Cosmochim. Acta 45:421–429.

Carlson, L., and U. Schwertmann. 1990. The effect of CO_2 and oxidation rate on the formation of goethite versus lepidocrocite from an Fe(II) system at pH 6 and 7. Clay Miner. 25:65–71.

Cerling, T.E., and R.R. Turner. 1982. Formation of freshwater Fe–Mn coatings on gravel and the behaviour of ^{60}Co, ^{90}Sr, and ^{137}Cs in a small watershed. Geochim. Cosmochim. Acta 46:1333–1343.

Chao, T.T. 1972. Selective dissolution of manganese oxides from soils and sediments with acidified hydroxylamine hydrochloride. Soil Sci. Soc. Am. J. 36:764–768.

Chao, G.Y., J. Baker, A.P. Sabina, and A.C. Green. 1985. Doyleite, a new polymorph of Al(OH)$_3$, and its relationship to bayerite, gibbsite, and nordstrandite. Can. Mineral. 23:21–28.

Chester, R., and R.N. Green. 1968. The infra-red determination of quartz in sediments and sedimentary rocks. Chem. Geol. 3:199–212.

Childs, C.W. 1975. Composition of iron–manganese concretions from some New Zealand soils. Geoderma 13:141–152.

Childs, C.W. 1981. Field tests for ferrous iron and ferric-organic complexes (on exchange sites or in water-soluble forms) in soils. Aust. J. Soil Res. 19:175–180.

Childs, C.W., R.W.P. Palmer, and C.W. Ross. 1990. Thick iron oxide pans in soils of Taranaki, New Zealand. Aust. J. Soil Res. 28:245–257.

Chukhrov, F.V., and A.I. Gorshkov. 1981. On the nature of some hypergene manganese minerals. Chem. Erde 40:207–216.

Chukhrov, F.V., A.I. Gorshkov, E.S. Rudnitskaya, V.V. Beresovskaya, and A.V. Sivtsov. 1980. Manganese minerals in clays: A review. Clays Clay Miner. 28:346–354.

Claridge, G.G.C., and A.V. Weatherhead. 1978. Mineralogy of silt fractions of New Zealand soils. N.Z. J. Sci. 21:413–423.

Clayton, R.N., R.W. Rex, J.K. Syers, and M.L. Jackson. 1972. Oxygen isotope abundance in quartz from Pacific pelagic sediments. J. Geophys. Res. 77:3907–3915.

Coey, J.M.D. 1988. Magnetic properties of iron in soil iron oxides and clay minerals, p. 397–466. *In* J.W. Stucki, B.A. Goodman, and U. Schwertmann (eds.) Iron in soils and clay minerals. D. Reidel Publishing Company, Dordrecht, the Netherlands.

Colin, F., C. Alarçon, and P. Vieillard. 1993. Zircon: An immobile index in soils? Chem. Geol. 107:273–276.

Combes, J.M., A. Manceau, G. Calas, and J.Y. Bottero. 1989. Formation of ferric oxides from aqueous solutions: A polyhedral approach by X-ray absorption spectroscopy: I. Hydrolysis and formation of ferric gels. Geochim. Cosmochim. Acta 53:583–594.

Cornell, R.M. 1987. Comparison and classification of the effects of simple ions and molecules upon the transformation of ferrihydrite into more crystalline products. Z. Pflanzenernähr. Bodenkd. 150:304–307.

Cornell, R.M., and U. Schwertmann. 1996. The iron oxides. Wiley-VCH Verlag, Weinheim, Germany.

Cornell, R.M., and U. Schwertmann. 2003. The iron oxides. 2nd edn. Wiley-VCH Verlag, Weinheim, Germany.

Crowther, D.L., and J.G. Dillard. 1983. The mechanism of Co(II) oxidation on synthetic birnessite. Geochim. Cosmochim. Acta 47:1399–1403.

Curi, N., and D.P. Franzmeier. 1984. Toposequence of Oxisols from the Central Plateau of Brazil. Soil Sci. Soc. Am. J. 48:341–346.

Dachs, H. 1963. Neutronen- und Rontgenuntersuchungen an Manganit, MnOOH. Z. Krist. 118:303–326.

da Motta, P.E.F., and N. Kämpf. 1991. Iron oxide properties as support to soil morphological features for prediction of moisture regimes in Oxisols of central Brazil. Z. Pflanzenernähr. Bodenkd. 155:385–390.

Davies, P.K., and A. Navrotsky. 1983. Quantitative correlations of deviations from ideality in binary and pseudobinary solid solutions. J. Solid State Chem. 46:1–22.

de Brito Galvão, T.C., and D.G. Schulze. 1996. Mineralogical properties of a collapsible lateritic soil from Minas Gerais, Brazil. Soil Sci. Soc. Am. J. 60:1969–1978.

Deer, W.A., R.A. Howie, and J. Zussman. 1992. An introduction to the rock-forming minerals. 2nd edn. Longman Scientific and Technical, Essex, UK.

Dixon, J.B., and G.N. White. 2002. Manganese oxides, p. 367–388. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Dormaar, J.F., and L.E. Lutwick. 1969. Infrared spectra of humic acids and opal phytoliths as indicators of paleosols. Can. J. Soil Sci. 49:29–37.

dos Anjos, L.H.C., D.P. Franzmeier, and D.G. Schulze. 1995. Formation of soils with plinthite on a toposequence in Maranhão State, Brazil. Geoderma 64:257–279.

Drees, L.R., L.P. Wilding, N.E. Smeck, and A.L. Senkayi. 1989. Silica in soils: Quartz and disordered silica polymorphs, p. 913–974. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Drits, V.A., B.A. Sakharov, and A. Manceau. 1993a. Structure of feroxyhite as determined by simulation of X-ray curves. Clay Miner. 28:209–221.

Drits, V.A., B.A. Sakharov, A.L. Salyn, and A. Manceau. 1993b. Structural model for ferrihydrite. Clay Miner. 28:185–207.

Egger, K., and W. Feitknecht. 1962. Über die Oxidation von Fe_3O_4 zu γ- und α-Fe_2O_3. Die differenzthermoanalytische (DTA) und thermogravimetrische (TG) Verfolgung des Reaktionsablaufes an künstlichen Formen von Fe_3O_4. Helv. Chim. Acta 45:2042–2057.

Eggleton, R.A. 1988. The application of micro-beam methods to iron minerals in soils, p. 165–201. *In* J.W. Stucki, B.A. Goodman, and U. Schwertmann (eds.) Iron in soils and clay minerals. D. Reidel Publishing Company, Dordrecht, the Netherlands.

Ehrlich, H.L. 1990. Geomicrobiology. 2nd edn. Marcel Dekker, New York.

Ehrlich, H.L. 1996. How microbes influence mineral growth and dissolution. Chem. Geol. 132:5–9.

Elderfield, H., and H.D. Hem. 1973. The development of crystalline structure in aluminum hydroxide polymorphs on aging. Mineral. Mag. 39:89–96.

Ewing, F.J. 1935a. The crystal structure of lepidocrocite. J. Chem. Phys. 3:420–424.

Ewing, F.J. 1935b. The crystal structure of diaspore. J. Chem. Phys. 3:203–207.

Fanning, D.S., M.C. Rabenhorst, and J.M. Bigham. 1993. Colors of acid sulfate soils, p. 91–108. *In* J.M. Bigham and E.J. Ciolkosz (eds.) Soil color. SSSA Special Publication No. 31. SSSA, Madison, WI.

Farmer, V.C., E. Delbos, and J.D. Miller. 2005. The role of phytolith formation and dissolution in controlling concentrations of silica in soil solution and streams. Geoderma 127:71–79.

Farmer, V.C., and J.D. Russel. 1967. Infrared absorption spectrometry in clay studies. Clays Clay Miner. 15:121–141.

Fasiska, E.J. 1967. Structural aspects of the oxides and oxidehydrates of iron. Corros. Sci. 7:833–839.

Fassbinder, J.W.E., H. Stanjek, and H. Vali. 1990. Occurrence of magnetic bacteria in soil. Nature 343:161–163.

Faulring, F.M., W.K. Zwicker, and W.D. Forgeng. 1960. Thermal transformations and properties of cryptomelane. Am. Mineral. 45:946–959.

Fendorf, S.E., M. Fendorf, D.L. Sparks, and R. Gronsky. 1992. Inhibitory mechanisms of Cr(III) oxidation by γ-MnO_2. J. Colloid Interface Sci. 153:37–54.

Fisher, R.F., C.N. Bourn, and W.F. Fisher. 1995. Opal phytoliths as an indicator of the floristics of prehistoric grasslands. Geoderma 68:243–255.

Fitzpatrick, R.W., and D.J. Chittleborough. 2002. Titanium and zirconium minerals, p. 667–690. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Fitzpatrick, R., B. Degens, A. Baker, M. Raven, P. Shand, M. Smith, S. Rogers, and R. George. 2008. Avon Basin, WA Wheatbelt: Acid sulfate soils and salt efflorescences in open drains and receiving environments, p. 189–204. *In* R. Fitzpatrick and P. Shand (eds.) Inland acid sulfate soil systems across Australia. CRC LEME open file report No. 249 (thematic volume). CRC LEME, Perth, Australia.

Fitzpatrick, R.W., J. LeRoux, and U. Schwertmann. 1978. Amorphous and crystalline iron–titanium oxides in synthetic preparation, at near ambient conditions, and in soil clays. Clays Clay Miner. 26:189–201.

Fitzpatrick, R.W., and U. Schwertmann. 1982. Al-substituted goethite—An indicator of pedogenic and other weathering environments in South Africa. Geoderma 27:335–347.

Fitzpatrick, R.W., R.M. Taylor, U. Schwertmann, and C.W. Childs. 1985. Occurrence and properties of lepidocrocite in some soils of New Zealand, South Africa and Australia. Aust. J. Soil Res. 23:543–567.

Flach, K.W., W.D. Nettleton, L.H. Gile, and J.C. Cady. 1969. Pedocementation: Induration by silica, carbonates, and sesquioxides in the quaternary. Soil Sci. 107:442–453.

Flach, K.W., W.D. Nettleton, and R.E. Nelson. 1974. The micromorphology of silica-cemented soil horizons in western North America, p. 714–729. *In* G.K. Rutherford (ed.) Soil microscopy. The Limestone Press, Kingston, Canada.

Fontes, M.P.F., L.H. Bowen, and S.B. Weed. 1991. Iron oxides in selected Brazilian Oxisols: II. Mössbauer studies. Soil Sci. Soc. Am. J. 55:1150–1155.

Fontes, M.P.F., and S.B. Weed. 1991. Iron oxides in selected Brazilian Oxisols: I. Mineralogy. Soil Sci. Soc. Am. J. 55:1143–1149.

Fordham, A.W., and K. Norrish. 1979. Electron microprobe and electron microscope studies of soil clay particles. Aust. J. Soil Res. 17:283–306.

Fordham, A.W., and K. Norrish. 1983. The nature of soil particles particularly those reacting with arsenate in a series of chemically treated samples. Aust. J. Soil Res. 21:455–477.

Franzmeier, D.P., L.D. Norton, and G.C. Steinhardt. 1989. Fragipan formation in loess of the midwestern United States, p. 69–97. *In* N.E. Smeck and E.J. Ciolkosz (eds.) Fragipans: Their occurrence, classification, and genesis. SSSA Special Publication No. 24. SSSA, Madison, WI.

Fritsch, E., A.J. Herbillon, N.R. do Nascimento, M. Grimaldi, and A.J. Melfi. 2007. From plinthic acrisols to plinthosols and gleysols: Iron and groundwater dynamics in the tertiary sediments of the upper Amazon basin. Eur. J. Soil Sci. 58:989–1006.

Fritsch, E., G. Morin, A. Bedidi, D. Bonnin, E. Balan, S. Caquineau, and G. Calas. 2005. Transformation of haematite and Al-poor goethite to Al-rich goethite and associated yellowing in a ferrallitic clay soil profile in the middle Amazon Basin (Manaus, Brazil). Eur. J. Soil Sci. 56:575–588.

Gallagher, K.J., W. Feitknecht, and U. Mannweiler. 1968. Mechanism of oxidation of magnetite to γ-Fe_2O_3. Nature 217:1118–1121.

Gao, X., and D.G. Schulze. 2010a. Chemical and mineralogical characterization of arsenic, lead, chromium, and cadmium in a metal-contaminated Histosol. Geoderma 156:278–286.

Gao, X., and D.G. Schulze. 2010b. Precipitation and transformation of secondary Fe oxyhydroxides in a Histosol impacted by runoff from a lead smelter. Clays Clay Miner. 58:377–387.

Geller, S. 1971. Structures of α-Mn_2O_3, $(Mn_{0.983}Fe_{0.017})_2O_3$ and $(Mn_{0.37}Fe_{0.63})_2O_3$ and relation to magnetic ordering. Acta Crystallogr. B 27:821–828.

Ghiorse, W.C. 1988. The biology of manganese transforming microorganisms in soil, p. 75–85. *In* R.D. Graham, R.J. Hannam, and N.C. Uren (eds.) Manganese in soils and plants. Kluwer Academic Publishers, Boston, MA.

Ghiorse, W.C., and H.L. Ehrlich. 1992. Microbial biomineralization of iron and manganese. Catena Suppl. 21:75–99.

Gilkes, R.J., G. Scholz, and G.M. Dimmock. 1973. Lateritic deep weathering of granite. J. Soil Sci. 24:523–536.

Gilkes, R.J., and A. Suddhiprakarn. 1979. Magnetite alteration in deeply weathered adamellite. J. Soil Sci. 30:357–361.

Giovanoli, R. 1985a. Layer structures and tunnel structures in manganates. Chem. Erde 44:227–244.

Giovanoli, R. 1985b. A review of the todorokite–buserite problem: Implications to the mineralogy of marine manganese nodules: Discussion. Am. Mineral. 70:202–204.

Glasser, L.S.D., and L. Ingram. 1968. Refinement of the crystal structure of groutite, α-MnOOH. Acta Crystallogr. B 24:1233–1236.

Glasser, L.S.D., and I. Smith. 1968. Oriented transformations in the system $MnO–O–H_2O$. Mineral. Mag. 36:976–987.

Golden, D.C., C.C. Chen, J.B. Dixon, and Y. Tokashiky. 1986. Pseudomorphic replacement of manganese oxides by iron oxide minerals. Geoderma 42:199–211.

Golden, D.C., J.B. Dixon, and Y. Kanehiro. 1993. The manganese oxide mineral, lithiophorite, in an Oxisol from Hawaii. Aust. J. Soil Res. 31:51–66.

Goss, C.J. 1988. Saturation magnetization, coercivity and lattice parameter changes in the system $Fe_3O_{4-}–\gamma Fe_2O_3$, and their relationship to structure. Phys. Chem. Miner. 16:164–171.

Graetsch, H. 1994. Structural characteristics of opaline and microcrystalline silica minerals. Rev. Mineral. Geochem. 29:209–232.

Graham, R.D., R.J. Hannam, and N.C. Uren. 1988. Manganese in soils and plants. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Greer, R.T. 1969. Submicron structure of amorphous opal. Nature 224:1199–1200.

Grey, I.E., C. Li, and J.A. Watts. 1983. Hydrothermal synthesis of goethite–rutile intergrowth structures and their relationship to pseudorutile. Am. Mineral. 68:981–988.

Grey, I.E., and A.F. Reid. 1975. The structure of pseudorutile and its role in the natural alteration of ilmenite. Am. Mineral. 60:898–906.

Gross, S., and L. Heller. 1963. A natural occurrence of bayerite. Mineral. Mag. 33:723–724.

Hathaway, J.C., and S.O. Schlanger. 1965. Nordstrandite $(Al_2O_3 \cdot 3H_2O)$ from Guam. Am. Mineral. 50:1029–1037.

Hazemann, J.L., A. Manceau, P. Sainctavit, and C. Malgrange. 1992. Structure of the $\alpha Fe_x Al_{1-x}OOH$ solid solution. I. Evidence by polarized EXAFS for an epitaxial growth of hematite-like clusters in diaspore. Phys. Chem. Miner. 19:25–38.

Healy, T.W., A.P. Herring, and D.W. Fuerstenau. 1966. The effect of crystal structure on the surface properties of a series of manganese dioxides. J. Colloid Interface Sci. 21:435–444.

Heaney, P.J. 1994. Structure and chemistry of the low-pressure silica polymorphs. Rev. Mineral. Geochem. 29:1–40.

Heaney, P.J., and J.A. Banfield. 1993. Structure and chemistry of silica, metal oxides, and phosphates. Rev. Mineral. Geochem. 28:185–233.

Heaney, P.J., C.T. Prewit, and G.V. Gibbs (eds.) 1994. Silica. Physical behavior, geochemistry and materials applications. Rev. Mineral. 29:375.

Hem, J.D., and C.J. Lind. 1983. Nonequilibrium models for predicting forms of precipitated manganese oxides. Geochim. Cosmochim. Acta 47:2037–2046.

Hemingway, B.S., R.A. Robie, and J.A. Apps. 1991. Revised values for the thermodynamic properties of boehmite, AlO(OH), and related species and phases in the system Al–H–O. Am. Mineral. 76:445–451.

Hesse, R. 1988. Origin of chert: Diagenesis of biogenic siliceous sediments. Geosci. Can. 15:171–192.

Hingston, F.J., A.M. Posner, and J.P. Quirk. 1974. Anion adsorption by goethite and gibbsite. I. The role of the proton in determining adsorption envelopes. J. Soil Sci. 23:177–192.

Holmgren, G.G.S. 1967. A rapid citrate–dithionite extractable iron procedure. Soil Sci. Soc. Am. Proc. 31:210–211.

Houot, S., and J. Berthelin. 1992. Submicroscopic studies of iron deposits occurring in field drains: Formation and evolution. Geoderma 52:209–222.

Hsu, P.H. 1966. Formation of gibbsite from aging hydroxy-aluminum solutions. Soil Sci. Soc. Am. Proc. 30:173–176.

Hsu, P.H. 1967. Effect of salts on the formation of bayerite versus pseudoboehmite. Soil Sci. 103:101–110.

Hsu, P.H. 1989. Aluminum hydroxides and oxyhydroxides, p. 331–378. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Huang, P.M. 1988. Ionic factors affecting aluminum transformations and the impact on soil and environmental sciences. Adv. Agron. 8:1–78.

Huang, P.M., M.K. Wang, N. Kämpf, and D.G. Schulze. 2002. Aluminum hydroxides, p. 261–289. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Huebner, J.S. 1976. The manganese oxides—A bibliographic commentary, p. SH1–SH17. *In* D. Rumble (ed.) Oxide minerals. Mineralogical Society of America Short Course Notes. Vol. 3. Mineralogical Society of America, Washington, DC.

Hughes, R.E., D.M. Moore, and H.D. Glass. 1994. Qualitative and quantitative analysis of clay minerals in soils, p. 330–359. *In* J.E. Amonette and L.W. Zelazny (eds.) Quantitative methods in soil mineralogy. SSSA, Madison, WI.

Hui, L., L.S. Lee, D.G. Schulze, and C.A. Guest. 2003. Role of soil manganese in the oxidation of aromatic amines. Environ. Sci. Technol. 37:2686–2693.

ICDD. 1994. Powder diffraction file 1995. PDF-2 database sets 1–45, Minerals. International Centre for Diffraction Data, Newtown Square, PA.

Jackson, M.L. 1964. Chemical composition of soils, p. 71–141. *In* F.E. Bear (ed.) Chemistry of the soil. Van Nostrand-Reinhold, New York.

Jackson, M.L. 1969. Soil chemical analysis. Advanced course. Parallel Press, Madison, WI.

Jeanroy, E., J.L. Rajot, P. Pillon, and A.J. Herbillon. 1991. Differential dissolution of hematite and goethite in dithionite and its implication on soil yellowing. Geoderma 50:79–94.

Jones, R.L., and A.H. Beavers. 1963. Sponge spicules in Illinois soils. Soil Sci. Soc. Am. Proc. 27:438–440.

Jones, R.L., and A.H. Beavers. 1964. Variation of opal phytolith content among some great soil groups of Illinois. Soil Sci. Soc. Am. Proc. 28:711–712.

Jones, L.H.P., and K.A. Handreck. 1967. Silica in soils, plants and animals. Adv. Agron. 19:107–149.

Jones, J.B., and E.R. Segnit. 1971. The nature of opal. I. Nomenclature and constituent phases. J. Geol. Soc. Aust. 18:57–68.

Juo, A.S.R. 1981. Chemical characteristics, p. 51–79. *In* D.J. Greenland (ed.) Characterization of soils. Oxford University Press, New York.

Kämpf, N., and U. Schwertmann. 1982a. The 5M NaOH concentration treatment for iron oxides in soils. Clays Clay Miner. 30:401–408.

Kämpf, N., and U. Schwertmann. 1982b. Quantitative determination of goethite and hematite in kaolinitic soils by X-ray diffraction. Clay Miner. 17:359–363.

Kämpf, N., and U. Schwertmann. 1983. Goethite and hematite in a climosequence in southern Brazil and their application in classification of kaolinitic soils. Geoderma 29:27–39.

Karathanasis, A.D. 1989. Solution chemistry of fragipans: Thermodynamic approach to understanding fragipan formation, p. 113–139. *In* N.E. Smeck and E.J. Ciolkosz (eds.) Fragipans: Their occurrence, classification, and genesis. SSSA Special Publication No. 24. SSSA, Madison, WI.

Karathanasis, A.D., and W.G. Harris. 1994. Quantitative thermal analysis of soil materials, p. 360–411. *In* J.E. Amonette and L.W. Zelazny (eds.) Quantitative methods in soil mineralogy. SSSA, Madison, WI.

Karathanasis, A.D., and Y.L. Thompson. 1995. Mineralogy of iron precipitates in a constructed acid mine drainage wetland. Soil Sci. Soc. Am. J. 59:1773–1781.

Kastner, M., J.B. Keene, and J.M. Gieskes. 1977. Diagenesis of siliceous oozes—1. Chemical controls on the rate of opal-A to opal-CT transformation—An experimental study. Geochim. Cosmochim. Acta 41:1041–1059.

Keller, W.D. 1978. Diaspore recrystallization at low temperature. Am. Mineral. 63:326–329.

Kelly, E.F., R.G. Amundson, B.D. Marino, and M.J. Deniro. 1991. Stable isotope ratios of carbon in phytoliths as a quantitative method of monitoring vegetation and climate change. Quat. Res. 35:222–233.

Kennedy, J.A., H.K.J. Powell, and J.M. White. 1982. A modification of Child's field test for ferrous iron and ferric-organic complexes in soils. Aust. J. Soil Res. 20:261–263.

Kittrick, J.A. 1969. Soil minerals in the Al_2O_3–SiO_2–H_2O system and a theory of their formation. Clays Clay Miner. 17:157–167.

Knauth, L.P. 1994. Petrogenesis of chert. Rev. Mineral. Geochem. 29:233–258.

Kodama, H., and M. Schnitzer. 1980. Effect of fulvic acid on the crystallization of aluminum hydroxide. Geoderma 24:195–205.

Leinemann, C.-P., M. Taillefert, D. Perret, and J.-F. Gaillard. 1997. Association of cobalt and manganese in aquatic systems: Chemical and microscopic evidence. Geochim. Cosmochim. Acta 61:1437–1466.

Lindsay, W.L. 1979. Chemical equilibria in soils. Wiley Interscience, New York.

Lindsley, D.H. 1976. The crystal chemistry and structure of oxide minerals as exemplified by the Fe–Ti oxides. Rev. Mineral. 3:L1–L60.

Lindsley, D.H. 1991. Oxide minerals: Petrologic and magnetic significance. Mineralogical Society of America, Washington, DC.

Little, I.P., T.M. Armitage, and R.J. Gilkes. 1978. Weathering of quartz in dune sands under subtropical conditions in Eastern Australia. Geoderma 20:225–237.

Loveland, P.J., and P. Digby. 1984. The extraction of Fe and Al by 0.1 M pyrophosphate solutions: A comparison of some techniques. J. Soil Sci. 35:243–250.

Lovley, D.R. 1992. Microbial oxidation of organic matter coupled to the reduction of Fe(III) and Mn(IV) oxides, p. 101–114. *In* H.C.W. Skinner and R.W. Fitzpatrick (eds.) Biomineralization. Processes of iron and manganese—Modern and ancient environments. Catena Suppl. 21. Catena Verlag, Cremlingen, Germany.

Lovley, D.R. 1995. Microbial reduction of iron, manganese, and other metals. Adv. Agron. 54:175–231.

Lovley, D.R., and E.J.P. Phillips. 1987. Rapid assay for microbially reducible ferric iron in aquatic sediments. Appl. Environ. Microbiol. 53:1536–1540.

Lovley, D.R., and E.J.P. Phillips. 1988. Novel mode of microbial energy metabolism: Organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. Appl. Environ. Microbiol. 54:1472–1480.

Lucas, Y. 2001. The role of plants in controlling rates and products of weathering: Importance of biological pumping. Annu. Rev. Earth Planet. Sci. 29:135–163.

Lutwick, L.E., and A. Johnston. 1969. Cumulic soils of the rough fescue prairie–poplar transition region. Can. J. Soil Sci. 49:199–203.

Macedo, J., and R.B. Bryant. 1987. Morphology, mineralogy, and genesis of a hydrosequence of Oxisols in Brazil. Soil Sci. Soc. Am. J. 51:690–698.

Macedo, J., and R.B. Bryant. 1989. Preferential microbial reduction of hematite over goethite in a Brazilian Oxisol. Soil Sci. Soc. Am. J. 53:1114–1118.

Mackenzie, R.C. 1970. Oxides and hydroxides of higher valency elements, p. 271–302. *In* R.C. Mackenzie (ed.) Differential thermal analysis. Vol. 1. Academic Press Ltd., New York.

Mackenzie, R.C. 1972. Soils, p. 267–297. *In* R.C. Mackenzie (ed.) Differential thermal analysis. Vol. 2. Academic Press Ltd., New York.

Mackenzie, F.T., and R. Gees. 1971. Quartz: Synthesis at earth-surface conditions. Science 173:533–535.

Maher, B.A., and R.M. Taylor. 1988. Formation of ultrafine-grained magnetite in soils. Nature 336:368–370.

Manceau, A., P.R. Buseck, D. Miser, J. Rask, and D. Nahon. 1990. Characterization of Cu in lithiophorite from a banded Mn ore. Am. Mineral. 75:490–494.

Manceau, A., and J.M. Combes. 1988. Structure of Mn and Fe oxides and oxyhydroxides: A topological approach by EXAFS. Phys. Chem. Miner. 15:283–295.

Manceau, A., and V.A. Drits. 1993. Local structure of ferrihydrite and feroxyhite by EXAFS spectroscopy. Clay Miner. 28:165–184.

Manceau, A., and W.P. Gates. 1997. Surface structural model for ferrihydrite. Clays Clay Miner. 45:448–460.

Manceau, A., A.I. Gorshkov, and V.A. Drits. 1992a. Structural chemistry of Mn, Fe, Co, and Ni in manganese hydrous oxides: I. Information from XANES spectroscopy. Am. Mineral. 77:1133–1143.

Manceau, A., A.I. Gorshkov, and V.A. Drits. 1992b. Structural chemistry of Mn, Fe, Co, and Ni in manganese hydrous oxides: II. Information from EXAFS spectroscopy and electron and X-ray diffraction. Am. Mineral. 77:1144–1157.

Manceau, A., S. Llorca, and G. Calas. 1987. Crystal chemistry of cobalt and nickel in lithiophorite and asbolane from New Caledonia. Geochim. Cosmochim. Acta 51:105–113.

Maslen, E.N., V.A. Streltsov, N.R. Streltsova, and N. Ishizawa. 1994. Synchrotron X-ray study of the electron density in α-Fe$_2$O$_3$. Acta Crystallogr. B 50:435–441.

Maslen, E.N., V.A. Streltsov, N.R. Streltsova, N. Ishizawa, and Y. Satow. 1993. Synchrotron X-ray study of the electron density in α-Al$_2$O$_3$. Acta Crystallogr. B 49:973–980.

McBride, M.B. 1987. Adsorption and oxidation of phenolic compounds by iron and manganese oxides. Soil Sci. Soc. Am. J. 51:1466–1472.

McBride, M.B. 1989. Reactions controlling heavy metal solubility in soils. Adv. Agron. 10:1–56.

McBride, M.B. 1994. Environmental chemistry of soils. Oxford University Press, New York.

McBride, M.B., and L.G. Wesselink. 1988. Chemisorption of catechol on gibbsite, boehmite, and noncrystalline alumina surfaces. Environ. Sci. Technol. 22:703–708.

McKeague, J.A. 1967. An evaluation of 0.1 M pyrophosphate and pyrophosphate-dithionite in comparison with oxalate as extractants of the accumulation products in podzols and some other soils. Can. J. Soil Sci. 47:95–99.

McKeague, J.A., and J.H. Day. 1966. Dithionite- and oxalate-extractable Fe and Al as aids in differentiating various classes of soils. Can. J. Soil Sci. 46:13–22.

McKenzie, R.M. 1971. The synthesis of cryptomelane and some other oxides and hydroxides of manganese. Mineral. Mag. 38:493–502.

McKenzie, R.M. 1978. The effect of two manganese dioxides on the uptake of Pb, Co, Ni, Cu and Zn by subterranean clover. Aust. J. Soil Res. 16:209–214.

McKenzie, R.M. 1989. Manganese oxides and hydroxides, p. 439–465. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

McNeal, B.L., D.A. Layfield, W.A. Norvell, and J.D. Rhoades. 1968. Factors influencing hydraulic conductivity of soils in the presence of mixed salt solution. Soil Sci. Soc. Am. Proc. 32:187–190.

Means, J.L., D.A. Crerar, M.P. Borcsik, and J.O. Duguid. 1978. Adsorption of cobalt and selected actinides by Mn and Fe oxides in soils and sediments. Geochim. Cosmochim. Acta 42:1763–1773.

Mehra, O.P., and M.L. Jackson. 1960. Iron oxide removal from soils and clays by a dithionite–citrate system buffered with sodium bicarbonate, p. 317–342. *In* A. Swineford (ed.) Proc. 7th Clays Clay Miner. Conf., Pergamon Press, Elmsford, New York.

Milnes, A.R., and R.W. Fitzpatrick. 1989. Titanium and zirconium minerals, p. 1131–1205. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Mitchell, W.A. 1975. Heavy minerals, p. 450–480. *In* J.E. Gieseking (ed.) Soil components. Vol. 2. Inorganic components. Springer-Verlag, Berlin, Germany.

Miyamoto, H. 1976. The magnetic properties of Fe(OH)$_2$. Mater. Res. Bull. 11:329–336.

Mizota, C., M. Itoh, M. Kusakabe, and M. Noto. 1991. Oxygen isotope ratios of opaline silica and plant opal in three recent volcanic ash soils. Geoderma 50:211–217.

Mizota, C., and Y. Matsuhisa. 1995. Isotopic evidence for the eolian origin of quartz and mica in soils developed on volcanic materials in the Canary Archipelago. Geoderma 66:313–320.

Monger, H.C., and E.F. Kelly. 2002. Silica minerals, p. 611–636. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Muggler, C.C. 1998. Polygenetic Oxisols on tertiary surfaces, Minas Gerais, Brazil: Soil genesis and landscape development. Ph.D. Thesis. Wageningen, the Netherlands.

Murad, E. 1988. Properties and behavior of iron oxides as determined by Mössbauer spectroscopy, p. 309–350. *In* J.W. Stucki, B.A. Goodman, and U. Schwertmann (eds.) Iron in soils and clay minerals. D. Reidel Publishing Company, Dordrecht, the Netherlands.

Murad, E. 1996. Magnetic properties of microcrystalline iron(III) oxides and related materials as reflected in their Mössbauer spectra. Phys. Chem. Miner. 23:248–262.

Murad, E. 1997. Identification of minor amounts of anatase in kaolins by Raman spectroscopy. Am. Mineral. 82:203–206.

Murad, E., U. Schwertmann, J.M. Bigham, and L. Carlson. 1994. Mineralogical characteristics of poorly crystallized precipitates formed by oxidation of Fe^{2+} in acid sulfate waters. ACS Symp. Ser. 550:190–200.

Murray, J.W. 1975. The interactions of metal ions at the manganese dioxide–solution interface. Geochim. Cosmochim. Acta 39:505–519.

Murray, J.W., J.G. Dillard, R. Giovanoli, H. Moers, and W. Stumm. 1985. Oxidation of Mn(II): Initial mineralogy, oxidation state and aging. Geochim. Cosmochim. Acta 49:463–470.

Norfleet, M.L., A.D. Karathanasis, and B.R. Smith. 1993. Soil solution composition relative to mineral distribution in Blue Ridge mountain soils. Soil Sci. Soc. Am. J. 57:1375–1380.

Norrish, K., and R.M. Taylor. 1961. The isomorphous replacement of iron by aluminium in soil goethites. J. Soil Sci. 12:294–306.

Obura, P.A. 2008. Effect of soil properties on bioavailability of aluminum and phosphorus in selected Kenyan and Brazilian acid soils. Ph.D. Thesis. Purdue University, West Lafayette, IN.

Ogg, C.M., and J.C. Baker. 1999. Pedogenesis and origin of deeply weathered soils formed in alluvial fans of the Virginia Blue Ridge. Soil Sci. Soc. Am. J. 63:601–606.

Ohta, S., S. Effendi, N. Tanaka, and S. Miura. 1993. Ultisols of lowland Dipterocarp forest in east Kalimantan, Indonesia. Soil Sci. Plant Nutr. 39:1–12.

Oscarson, D.W., P.M. Huang, W.K. Liaw, and U.T. Hammer. 1983. Kinetics of oxidation of arsenite by various manganese oxides. Soil Sci. Soc. Am. J. 47:644–648.

Ostwald, J. 1984a. Two varieties of lithiophorite in some Australian deposits. Mineral. Mag. 48:383–388.

Ostwald, J. 1984b. The influence of clay mineralogy on the crystallization of the tetravalent manganese layer-lattice minerals. Neues Jahrb. Mineral. Abh. 1984:9–16.

Ostwald, J. 1985. Some observations on the chemical composition of chalcophanite. Mineral. Mag. 49:752–755.

Parc, S., D. Nahon, Y. Tardy, and P. Vieillard. 1989. Estimated solubility products and fields of stability for cryptomelane, nsutite, birnessite, and lithiophorite based on natural lateritic weathering sequences. Am. Mineral. 74:466–475.

Parfitt, R.L., and C.W. Childs. 1988. Estimation of forms of Fe and Al: A review, and analysis of contrasting soils by dissolution and Moessbauer methods. Aust. J. Soil Res. 26:121–144.

Parfitt, R.L., C.W. Childs, and D.N. Eden. 1988. Ferrihydrite and allophane in four Andepts from Hawaii and implications for their classifications. Geoderma 41:223–241.

Peterschmitt, E., E. Fritsch, J.L. Rajot, and A.J. Herbillon. 1996. Yellowing, bleaching and ferritisation processes in soil mantle of the Western Ghats, South India. Geoderma 74:235–253.

Pinheiro-Dick, D., and U. Schwertmann. 1996. Microaggregates from Oxisols and Inceptisols: Dispersion through selective dissolutions and physicochemical treatments. Geoderma 74:49–63.

Post, J.E. 1992. Crystal structures of manganese oxide minerals, p. 51–73. *In* H.C.W. Skinner and R.W. Fitzpatrick (eds.) Biomineralization. Processes of iron and manganese—Modern and ancient environments. Catena Suppl. 21. Catena Verlag, Cremlingen, Germany.

Post, J.E., and D.E. Appleman. 1988. Chalcophanite, $ZnMn_3O_7\cdot3H_2O$: New crystal-structure determinations. Am. Mineral. 73:1401–1404.

Post, J.E., and D.E. Appleman. 1994. Crystal structure refinement of lithiophorite. Am. Mineral. 79:370–374.

Post, J.E., and D.L. Bish. 1988. Rietveld refinement of the todorokite structure. Am. Mineral. 73:861–869.

Post, J.E., and D.L. Bish. 1989. Rietveld refinement of the coronadite structure. Am. Mineral. 74:913–917.

Post, J.E., and V.F. Buchwald. 1991. Crystal structure refinement of akaganéite. Am. Mineral. 76:272–277.

Post, J.E., and C.W. Burnham. 1986. Modelling tunnel-cation displacements in hollandites using structure-energy calculations. Am. Mineral. 71:1178–1185.

Post, J.E., and D.R. Veblen. 1990. Crystal structure determinations of synthetic sodium, magnesium, and potassium birnessite using TEM and the Rietveld method. Am. Mineral. 75:477–489.

Post, J.E., R.B. von Dreele, and P.R. Buseck. 1982. Symmetry and cation displacements in hollandites: Structure refinements of hollandite, cryptomelane and priderite. Acta Crystallogr. B 38:1056–1065.

Potter, R.M., and G.R. Rossman. 1979. The tetravalent manganese oxides: Identification, hydration, and structural relationships by infrared spectroscopy. Am. Mineral. 64:1199–1218.

Pye, K., and J. Mazzullo. 1994. Effects of tropical weathering on quartz grain shape. An example from northeastern Australia. J. Sediment. Res. A 64:500–507.

Rask, J.H., and P.R. Buseck. 1986. Topotactic relations among pyrolusite, manganite, and Mn_5O_8: A high-resolution transmission electron microscopy investigation. Am. Mineral. 71:805–814.

Richardson, J.L., and F.D. Hole. 1979. Mottling and iron distribution in a Glossoboralf-Haplaquoll hydrosequence on a glacial moraine in northwestern Wisconsin. Soil Sci. Soc. Am. J. 43:552–558.

Robbins, E.I., J.P. D'Agostino, J. Ostwald, D.S. Fanning, V. Carter, and R.L. Van Hoven. 1992. Manganese nodules and microbial oxidation of manganese in the Huntley Meadows wetland, Virginia, USA, p. 179–202. *In* H.C.W. Skinner and R.W. Fitzpatrick (eds.) Biomineralization. Processes of iron and manganese—Modern and ancient environments. Catena Suppl. 21. Catena Verlag, Cremlingen, Germany.

Rodgers, K.A. 1993. Routine identification of aluminum hydroxide polymorphs with the laser Raman microprobe. Clay Miner. 28:85–99.

Ross, S.J., Jr., D.P. Franzmeier, and C.B. Roth. 1976. Mineralogy and chemistry of manganese oxides in some Indiana soils. Soil Sci. Soc. Am. J. 40:137–143.

Rowse, J.B., and W.B. Jepson. 1972. The determination of quartz in clay minerals: A critical comparison of methods. J. Therm. Anal. 4:169–175.

Ryskin, Y.I. 1974. The vibration of protons in minerals: Hydroxyl, water and ammonium, p. 137–181. *In* V.C. Farmer (ed.) Infrared spectra of minerals. Mineralogical Society, London, UK.

Sacconne, L., D.J. Conley, E. Koning, D. Sauer, M. Sommer, D. Kaczorek, S.W. Blecker, and E.F. Keller. 2007. Assessing the extraction and quantification of amorphous silica in soils of forest and grassland ecosystems. Eur. J. Soil Sci. 58:1446–1459.

Sasaki, S., K. Fukino, Y. Takéuchi, and R. Sadanaga. 1980. On the estimation of atomic charges by the x-ray method for some oxides and silicates. Acta Crystallogr. A 36:904–915.

Sauer, D., L. Saccone, D.J. Conley, L. Herrmann, and M. Sommer. 2006. Review of methodologies for extracting plant-available and amorphous Si from soils and aquatic sediments. Biogeochemistry 80:89–108.

Schmahl, W.W., I.P. Swainson, M.T. Dove, and A. Graeme-Barber. 1992. Landau free energy and order parameter behavior of the α/β phase transition in cristobalite. Z. Kristallogr. 201:125–145.

Schoen, R., and C.E. Roberson. 1970. Structures of aluminum hydroxide and geochemical implications. Am. Mineral. 55:43–77.

Schulze, D.G. 1981. Identification of soil iron oxide minerals by differential x-ray diffraction. Soil Sci. Soc. Am. J. 45:437–440.

Schulze, D.G. 1984. The influence of aluminum on iron oxides. VIII. Unit cell dimensions of Al substituted goethites and estimation of Al from them. Clays Clay Miner. 32:36–44.

Schulze, D.G. 1988. Separation and concentration of iron-containing phases, p. 63–81. *In* J.W. Stucki, B.A. Goodman, and U. Schwertmann (eds.) Iron in soils and clay minerals. D. Reidel Publishing Company, Dordrecht, the Netherlands.

Schulze, D.G., and J.B. Dixon. 1979. High gradient magnetic separation of iron oxides and magnetic minerals from soil clays. Soil Sci. Soc. Am. J. 43:793–799.

Schuppli, P.A., G.J. Ross, and J. McKeague. 1983. The effective removal of suspended materials from pyrophosphate extracts of soils from tropical and temperate regions. Soil Sci. Soc. Am. J. 47:1026–1032.

Schwarz, T. 1994. Ferricrete formation and relief inversion: An example from Central Sudan. Catena 21:257–268.

Schwertmann, U. 1964. Differenzierung der Eisenoxide des Bodens durch photochemische Extraktion mit saurer Ammoniumoxalat-Lösung. Z. Pflanzenernähr. Bodenkd. 105:194–202.

Schwertmann, U. 1966. Inhibitory effect of soil organic matter on the crystallization of amorphous ferric hydroxide. Nature (London) 212:645–646.

Schwertmann, U. 1988. Occurrence and formation of iron oxides in various pedoenvironments, p. 267–308. *In* J.W. Stucki, B.A. Goodman, and U. Schwertmann (eds.) Iron in soils and clay minerals. D. Reidel Publishing Company, Dordrecht, the Netherlands.

Schwertmann, U. 1991. Solubility and dissolution of iron oxides. Plant Soil 130:1–25.

Schwertmann, U. 1993. Relation between iron oxides, soil color, and soil formation, p. 51–69. *In* J.M. Bigham and E.J. Ciolkosz (eds.) Soil color. SSSA, Madison, WI.

Schwertmann, U., J.M. Bigham, and E. Murad. 1995a. The first occurrence of schwertmannite in a natural stream environment. Eur. J. Miner. 7:547–552.

Schwertmann, U., and L. Carlson. 1994. Aluminum influence on iron oxides. XVII. Unit-cell parameters and aluminum substitution of natural goethites. Soil Sci. Soc. Am. J. 58:256–261.

Schwertmann, U., and D.S. Fanning. 1976. Iron–manganese concretions in hydrosequences of soils in loess in Bavaria. Soil Sci. Soc. Am. J. 40:731–738.

Schwertmann, U., and H. Fechter. 1984. The influence of aluminium on iron oxides. XI. Aluminium substituted maghemite in soils and its formation. Soil Sci. Soc. Am. J. 48:1462–1463.

Schwertmann, U., and H. Fechter. 1994. The formation of green rust and its transformation to lepidocrocite. Clay Miner. 29:87–92.

Schwertmann, U., and W.R. Fischer. 1973. Natural "amorphous" ferric hydroxide. Geoderma 10:237–247.

Schwertmann, U., and R.W. Fitzpatrick. 1992. Iron minerals in surface environments, p. 7–30. *In* H.C.W. Skinner and R.W. Fitzpatrick (eds.) Biomineralization—Processes of iron and manganese. Catena Verlag, Cremlingen, Germany.

Schwertmann, U., R.W. Fitzpatrick, R.M. Taylor, and D.G. Lewis. 1979. The influence of aluminium on iron oxides. Part II. Preparation and properties of Al substituted hematites. Clays Clay Miner. 27:105–112.

Schwertmann, U., J. Friedl, G. Pfab, and A.U. Gehring. 1995b. Iron substitution in soil and synthetic anatase. Clays Clay Miner. 43:599–606.

Schwertmann, U., and N. Kämpf. 1983. Oxidos de ferro jovens em ambientes pedogeneticos brasileiros. Rev. Bras. Cienc. Solo 7:251–255.

Schwertmann, U., and N. Kämpf. 1985. Properties of goethite and hematite in kaolinitic soils of southern and central Brazil. Soil Sci. 139:344–350.

Schwertmann, U., and M. Latham. 1986. Properties of iron oxides in some New Caledonian Oxisols. Geoderma 39:105–123.

Schwertmann, U., and G. Pfab. 1994. Structural vanadium in synthetic goethite. Geochim. Cosmochim. Acta 58:4349–4352.

Schwertmann, U., and G. Pfab. 1996. Structural vanadium and chromium in lateritic iron oxides: Genetic implications. Geochim. Cosmochim. Acta 60:4279–4283.

Schwertmann, U., D.G. Schulze, and E. Murad. 1982. Identification of ferrihydrite in soils by dissolution kinetics, differential x-ray diffraction and Mössbauer spectroscopy. Soil Sci. Soc. Am. J. 46:869–875.

Schwertmann, U., and R.M. Taylor. 1989. Iron oxides, p. 379–438. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Scott, M.J., and J.J. Morgan. 1995. Reactions at oxide surfaces. 1. Oxidation of As(III) by synthetic birnessite. Environ. Sci. Technol. 29:1898–1905.

Senkayi, A.L., J.B. Dixon, and L.R. Hossner. 1986. Todorokite, goethite, and hematite: Alteration products of siderite in east Texas lignite overburden. Soil Sci. 142:36–42.

Shafdan, H., J.B. Dixon, and F.G. Calhoun. 1985. Iron oxide properties versus strength of ferruginous crust and iron glaebules in soils. Soil Sci. 140:317–325.

Shah Singh, S., and H. Kodama. 1994. Effect of the presence of aluminum ions in iron solutions on the formation of iron oxyhydroxides (FeOOH) at room temperature under acidic environment. Clays Clay Miner. 42:606–613.

Sidhu, P.S., J.L. Sehgal, M.K. Sinha, and N.S. Randhawa. 1977. Composition and mineralogy of iron–manganese concretions from some soils of the Indo-Gangetic plain in northwest India. Geoderma 18:241–249.

Singer, A., and P.M. Huang. 1990. Effects of humic acids on the crystallization of aluminum hydroxides. Clays Clay Miner. 38:47–52.

Singh, B., and R.J. Gilkes. 1992. Properties and distribution of iron oxides and their association with minor elements in the soils of south-western Australia. J. Soil Sci. 43:77–98.

Skinner, H.C.W., and R.W. Fitzpatrick (eds.). 1992. Biomineralization. Processes of iron and manganese. Catena Suppl. 21. Catena Verlag, Cremlingen, Germany.

Soil Survey Staff. 1975. Soil taxonomy—A basic system of soil classification for making and interpreting soil surveys. Agricultural Handbook No. 436. U.S. Government Printing Office, Washington, DC.

Speer, J.A. 1982. Zircon, p. 67–112. *In* P.H. Ribbe (ed.) Reviews in mineralogy. Vol. 5. Orthosilicates. Mineralogical Society of America, Washington, DC.

Staller, J.E. 2002. A multidisciplinary approach to understanding the initial introduction of maize into coastal Ecuador. J. Archaeol. Sci. 29:33–50.

Stanjek, H. 1987. The formation of maghemite and hematite from lepidocrocite and goethite in a Cambisol from Corsica, France. Z. Pflanzenernähr. Bodenkd. 150:314–318.

Stone, A.T., and J.J. Morgan. 1984a. Reduction and dissolution of manganese(III) and manganese(IV) oxides by organics. 1. Reaction with hydroquinone. Environ. Sci. Technol. 18:450–456.

Stone, A.T., and J.J. Morgan. 1984b. Reduction and dissolution of manganese(III) and manganese(IV) oxides by organics. 2. Survey of the reactivity of organics. Environ. Sci. Technol. 18:617–624.

Sudom, M.D., and R.J. St. Arnaud. 1971. Use of quartz, zirconium and titanium as indices in pedological studies. Can. J. Soil Sci. 51:385–396.

Süsser, P., and U. Schwertmann. 1991. Proton buffering in mineral horizons of some acid forest soils. Geoderma 49:63–76.

Tait, J.M., A. Violante, and P. Violante. 1983. Coprecipitation of gibbsite and bayerite with nordstrandite. Clay Miner. 18:95–99.

Tanaka, K., and J. White. 1982. Characterization of species adsorbed on oxidized and reduced anatase. J. Phys. Chem. 86:4708–4714.

Taylor, R.M. 1968. The association of manganese and cobalt in soils—Further observations. J. Soil Sci. 19:77–80.

Taylor, R.M. 1987. Non-silicates oxides and hydroxides, p. 129–201. *In* A.C.D. Newman (ed.) Chemistry of clays and clay minerals. John Wiley & Sons, New York.

Taylor, R.M., and R.M. McKenzie. 1980. The influence of aluminium on iron oxides. VI. The formation of Fe(II)–Al(III) hydroxy-chlorides, -sulphates, and -carbonates as new members of the pyroaurite group and their significance in soils. Clays Clay Miner. 28:179–187.

Taylor, R.M., R.M. McKenzie, A.W. Fordham, and G.P. Gillman. 1983. Oxide minerals, p. 309–334. *In* Soils: An Australian viewpoint. Academic Press, London, UK.

Taylor, R.M., R.M. McKenzie, and K. Norrish. 1964. The mineralogy and chemistry of manganese in some Australian soils. Aust. J. Soil Res. 2:235–248.

Taylor, R.M., and U. Schwertmann. 1974. Maghemite in soils and its origin. I. Properties and observations on soil maghemites. Clay Miner. 10:289–298.

Tebo, B.M., J.R. Bargar, B.G. Clement, G.J. Dick, K.J. Murray, D. Parker, R. Verity, and S.M. Webb. 2004. Biogenic manganese oxides: Properties and mechanisms of formation. Annu. Rev. Earth Planet. Sci. 32:287–328.

Tejan-Kella, M.S., R.W. Fitzpatrick, and D.J. Chittleborough. 1991. Scanning electron microscope study of zircons and rutiles from a podzol chronosequence at Cooloola, Queensland, Australia. Catena 18:11–30.

Tettenhorst, R., and D.A. Hofmann. 1980. Crystal chemistry of boehmite. Clays Clay Miner. 28:373–380.

Thanabalasingam, P., and W.F. Pickering. 1986. Effect of pH on the interaction between As(III) or As(V) and manganese(IV) oxide. Water Air Soil Pollut. 29:205–216.

Thompson, I.A., D.M. Huber, C.A. Guest, and D.G. Schulze. 2005. Fungal manganese oxidation in a reduced soil. Environ. Microbiol. 7:1480–1487.

Thompson, I.A., D.M. Huber, and D.G. Schulze. 2006. Evidence of a multicopper oxidase in Mn oxidation by *Gaeumannomyces graminis* var. *tritici*. Phytopathology 96:130–136.

Tokashiki, Y., J.B. Dixon, and D.C. Golden. 1986. Manganese oxide analysis in soils by combined x-ray diffraction and selective dissolution methods. Soil Sci. Soc. Am. J. 50:1079–1084.

Torrent, J., R. Guzman, and M.A. Parra. 1982. Influence of relative humidity on the crystallization of Fe(III) oxides from ferrihydrite. Clays Clay Miner. 30:337–340.

Torrent, J., U. Schwertmann, and V. Barrón. 1987. The reductive dissolution of synthetic goethite and hematite in dithionite. Clay Miner. 22:329–337.

Torrent, J., U. Schwertmann, H. Fechter, and F. Alferez. 1983. Quantitative relationships between soil colour and hematite content. Soil Sci. 136:354–358.

Towe, K.M., and W.F. Bradley. 1967. Mineralogical constitution of colloidal "hydrous ferric oxides". J. Colloid Interface Sci. 24:384–392.

Trolard F., G. Bourrié, M. Abdelmoula, P. Refait, and F. Feder. 2007. Fougerite, a new mineral of the pyroaurite–iowaite group: Description and crystal structure. Clays Clay Miner. 55:323–334.

Trolard, F., and Y. Tardy. 1987. The stabilities of gibbsite, boehmite, aluminous goethites and aluminous hematites in bauxites, ferricretes, and laterites as a function of water activity, temperature, and particle size. Geochim. Cosmochim. Acta 51:945–957.

Turner, S., and P.R. Buseck. 1979. Manganese oxide tunnel structures and their intergrowths. Science 203:456–458.

Turner, S., and P.R. Buseck. 1981. Todorokites: A new family of naturally occurring manganese oxides. Science 212:1024–1027.

Turner, S., and P.R. Buseck. 1983. Defects in nsutite (γ-MnO_2) and dry-cell battery efficiency. Nature 304:143–146.

Turner, S., and J.E. Post. 1988. Refinement of the substructure and superstructure of romanèchite. Am. Mineral. 73:1155–1161.

Uzochukwu, G.A., and J.B. Dixon. 1986. Manganese oxide minerals in nodules of two soils of Texas and Alabama. Soil Sci. Soc. Am. J. 50:1358–1363.

van Bodegom, P.M., J. van Reeven, and H.A.C.D. van der Gon. 2003. Prediction of reducible soil iron content from iron extraction data. Biogeochemistry 64:231–245.

van Wambeke, A. 1992. Soils of the tropics. McGraw-Hill, New York.

Varentsov, I.M. 1982. Groote Eylandt manganese oxide deposits, Australia. Chem. Erde 41:157–173.

Vgenopoulos, A.G. 1984. Genesis of boehmite resp. diaspore in bauxite in dependence of redox equilibrium. Chem. Erde 43:149–159.

Violante, A., and P.M. Huang. 1985. Influence of inorganic and organic ligands on the formation of aluminum hydroxides and oxyhydroxides. Clays Clay Miner. 33:181–192.

Violante, A., and P.M. Huang. 1994. Identification of pseudoboehmite in mixtures with phyllosilicates. Clay Miner. 29:351–359.

Violante, A., and M.L. Jackson. 1979. Crystallization of nordstrandite in citrate systems and in the presence of montmorillonite, p. 517–525. *In* M.M. Mortland and V.C. Farmer (eds.) Proc. Intl. Clay Conf. 1978 (Oxford), Elsevier Scientific Publishing Company, Amsterdam, the Netherlands.

Violante, A., and M.L. Jackson. 1981. Clay influence on the crystallization of aluminum hydroxide polymorphs in the presence of citrate, sulfate or chloride. Geoderma 25:199–214.

Violante, A., and P. Violante. 1980. Influence of pH, concentration and chelating power of organic anions on the synthesis of aluminum hydroxides and oxyhydroxides. Clays Clay Miner. 28:425–434.

Violante, P., A. Violante, and J.M. Tait. 1982. Morphology of nordstrandite. Clays Clay Miner. 30:431–437.

Wall, J.R.D., E.B. Wolfenden, E.H. Beard, and T. Deans. 1962. Nordstrandite in soil from West Sarawak, Borneo. Nature 196:264–265.

Wang, H.D., G.N. White, F.T. Turner, and J.B. Dixon. 1993. Ferrihydrite, lepidocrocite, and goethite in coatings from east Texas vertic soils. Soil Sci. Soc. Am. J. 57:1381–1386.

Waychunas, G.A. 1991. Crystal chemistry of oxides and oxyhydroxides, p. 11–68. *In* D.H. Lindsley (ed.) Reviews in mineralogy. Vol. 25. Oxide minerals: petrologic and magnetic significance. Mineralogical Society of America, Washington, DC.

Wechsler, B.A., D.H. Lindsley, and C.T. Prewit. 1984. Crystal structure and cation distribution in titanomagnetite ($Fe_{3-x}Ti_xO_4$). Am. Mineral. 69:754–770.

Weiss, J.V., D. Emerson, S.M. Backer, and J.P. Megonigal. 2003. Enumeration of Fe(II)-oxidizing and Fe(III)-reducing bacteria in the root zone of wetland plants: Implications for a rhizosphere iron cycle. Biogeochemistry 64:77–96.

Wells, M.A., R.J. Gilkes, and R.R. Anand. 1989. The formation of corundum and aluminous hematite by thermal dehydroxylation of aluminous goethite. Clay Miner. 24:513–530.

White, G.N., and J.B. Dixon. 1996. Iron and manganese distribution in nodules from a young Texas Vertisol. Soil Sci. Soc. Am. J. 60:1254–1262.

Wilding, L.P., and L.R. Drees. 1968. Distribution and implications of sponge spicules in surficial deposits in Ohio. Ohio J. Sci. 68:92–99.

Wilding, L.P., and L.R. Drees. 1971. Biogenic opal in Ohio soils. Soil Sci. Soc. Am. Proc. 35:1004–1010.

Wilding, L.P., and L.R. Drees. 1974. Contributions of forest opal and associated crystalline phases to fine silt and clay fractions of soils. Clays Clay Miner. 22:295–306.

Will, G., M. Bellotto, W. Parrish, and M. Hart. 1988. Crystal structures of quartz and magnesium germanate by profile analysis of synchrotron-radiation high-resolution powder data. J. Appl. Cryst. 21:182–191.

Williams, L.A., and D.A. Crerar. 1985. Silica diagenesis, II. General mechanisms. J. Sediment. Petrol. 55:312–321.

Williams, L.A., G.A. Parks, and D.A. Crerar. 1985. Silica diagenesis, I. Solubility controls. J. Sediment. Petrol. 55:301–311.

Yamada, N., M. Ohmasa, and S. Horiuchi. 1986. Textures in natural pyrolusite, β-MnO_2, examined by 1 MV HRTEM. Acta Crystallogr. B 42:58–61.

Zeese, R., U. Schwertmann, G.F. Tietz, and U. Jux. 1994. Mineralogy and stratigraphy of three deep lateritic profiles of the Jos plateau (central Nigeria). Catena 21:195–214.

Zwicker, W.K., W.O.J.G. Meijer, and H.W. Jaffe. 1962. Nsutite—A widespread manganese oxide mineral. Am. Mineral. 47:246–266.

Poorly Crystalline Aluminosilicate Clay Minerals

James Harsh
Washington State University

This chapter covers the poorly crystalline aluminosilicate materials, commonly known as allophane and imogolite. With dimensions generally less than 50 nm, allophane and imogolite are naturally occurring nanoparticles. Their solubility, high specific surface area, variable charge, and unique physical behavior impart special properties to soils that contain them. Although often associated with soils formed from volcanic material (Andisols), allophane and imogolite are found within a wide range of soil orders and derived from a variety of parent materials. Ferrihydrite, a poorly crystalline iron oxide that often occurs with allophane and imogolite, is covered elsewhere in this handbook. A timely review by Parfitt (2009) considered the role of allophane and imogolite in biogeochemical processes, Dahlgren et al. (2004) reviewed volcanic soils, and Floody et al. (2009) discussed industrial applications of allophane and other natural nanoclays.

23.1 Structure of Poorly Crystalline Materials

23.1.1 Definition of Short-Range Ordered Materials

Crystalline materials exhibit long-range order in their atomic structure. Structural features repeat over scales of at least micrometers in a single crystal or over several particles in a powder. Such minerals display narrow x-ray diffraction peaks and their crystal habit can be directly observed at the scale of an optical microscope or by the naked eye in the case of large crystals. Amorphous materials, on the other hand, exhibit no order, even in the local

environment of the atoms. The atoms are arranged in a variety of states and structures where bond lengths, coordination, and geometry vary from site to site. Poorly crystalline materials fall in between these two extremes, showing short-range, medium-range, or limited long-range order. Short-range order is observable by techniques that probe the local environment of each atom, such as nuclear magnetic resonance (NMR) or x-ray photoelectron spectroscopy. X-ray diffraction peaks are indicative of, at least, medium-range order, whereas long-range order is detectable by electron microscopy. Poorly crystalline materials lack medium-range repetition of structural units in at least one of the three spatial dimensions; that is, well defined electron or x-ray diffraction patterns (XRD) are not produced. Allophane and imogolite practically represent the range from amorphous to crystalline materials (Fyfe et al., 1987). In the interest of brevity, allophane, imogolite, and poorly crystalline iron hydrous oxides will be referred to collectively as short-range ordered (SRO) materials.

In spite of their wide occurrence as intermediates in soil formation, allophane and imogolite were, until the second-half of the Twentieth Century, missed or ignored because of their virtual absence from XRD. When they were found to be important constituents of soils derived from volcanic ash (Taylor, 1933; Yoshinaga and Aomine, 1962; Abidin et al., 2007), researchers turned to a variety of methods more conducive to their identification, characterization, and quantification. These include selective dissolution by oxalic acid to quantify Si and Al, thermoanalytic methods to identify and quantify specific minerals, structural characterization by vibrational spectroscopy (IR and FTIR) (Farmer et al., 1977; Dahlgren, 1994), and electron microscopy

(Wilson, 1987). Short-range structural information has been obtained with solid-state NMR and x-ray absorption spectroscopy (XAS). Even the diffuse peaks in XRD patterns are useful to identify and quantify poorly crystalline materials (Cruz and Real, 1991). More recently, computer modeling has been employed to test proposed structures against spectroscopic data and to determine which structures give an energy minimum (Guimaraes et al., 2007; Creton et al., 2008; Zhao et al., 2009). The application of such methods to determine imogolite and allophane structure is discussed in the following sections.

23.1.2 Structure of Imogolite

Imogolite is a mineral with relatively constant chemical composition ($SiAl_2O_5 \cdot 2H_2O$), only six-fold coordination for Al, as many as seven x-ray diffraction peaks and two very broad "bands" (Table 23.1), and a distinctive tubular morphology observable under the electron microscope (Figure 23.1). Thus, imogolite has the major features of a crystalline mineral, but some consider it paracrystalline because each tube displays long-range order primarily in one dimension—along the length of the tube (Greenland, 1982). Diffraction patterns and electron micrographs, however, show that bundles of imogolite are aligned in an array with a repeat distance of around 2 nm, consistent with the diameter of the tubes (Figure 23.1; Cradwick et al., 1972). Thus, imogolite displays an added dimension of repeating units similar to layered aluminosilicates with regular stacking.

The Al in imogolite occurs in a gibbsite-like sheet and the Si is coordinated to three oxygen atoms and one hydroxyl group. The oxygen atoms are shared with the Al octahedra and the apical hydroxyls point toward the inside of the imogolite tube (Figure 23.2; Cradwick et al., 1972; Zhao et al., 2009). The separation distances between close-packed tubes are typically around 2.3 nm for natural and 2.7–3.2 nm for synthetic imogolite (Mukherjee et al., 2005) and the repeat distance along the gibbsite sheet (c-axis) is 0.84 nm (Wada and Yoshinaga, 1969).

FIGURE 23.1 Electron micrographs of synthetic imogolite with selected area electron diffraction pattern (top left) (Yang et al., 2008), natural imogolite from pumice (bottom left), and allophane from pumice (center). (Reproduced from Parfitt, R.L. 2009. Allophane and imogolite: Role in soil biogeochemical processes. Clay Miner. 44:135–155. With the permission of the Mineralogical Society of Great Britain & Ireland.) The cartoon at the top right shows the orientation of imogolite nanotube bundles. (From Bonelli, B., I. Bottero, N. Ballarini, S. Passeri, F. Cavani, and E. Garrone. 2009. IR spectroscopic and catalytic characterization of the acidity of imogolite-based systems. J. Catal. 264:15–30.)

TABLE 23.1 Powder X-Ray Diffraction Peaks for Proto-Imogolite Allophane and Imogolite

Proto-Imogolite Allophane		Imogolite	
d (nm)	I	d (nm)	I
1.2vb[a]	70	1.6vb	100
0.43vb	10	0.79	70
0.34vb[b]	100	0.56	35
0.22vb[b]	50	0.44	10
0.19[a]	10	0.41	10
0.17	10	0.37	20
0.14[b]	20	0.33vb	65
		0.31	5
		0.26	5
		0.225vb	25

[a] Also present in halloysite-like allophane.
[b] Present in all three allophanes.

Unraveling the structure of imogolite relies on spectroscopy and modeling. Infrared spectra give absorption bands indicative of both the gibbsite-like sheet and the orthosilicate anion. The Si–O stretching vibration at 960 cm^{-1} is consistent with unpolymerized orthosilicate groups and the OH-stretching bands are consistent with Al–OH and Si–OH structures described by Cradwick et al. (1972). In addition, the band at 348 cm^{-1} indicates that one hydroxyl of the octahedral Al sheet is replaced by the silica oxyanion. This band distinguishes imogolite and proto-imogolite allophane from other SRO aluminosilicates (Farmer et al., 1977). Molecular dynamics simulations that generate XRD patterns and vibrational spectra based on the Cradwick model compared favorably with the experimental ones (Creton et al., 2008).

Solid-state NMR spectra also support the above model for imogolite (Goodman et al., 1985; Wilson et al., 1986; He et al., 1995).
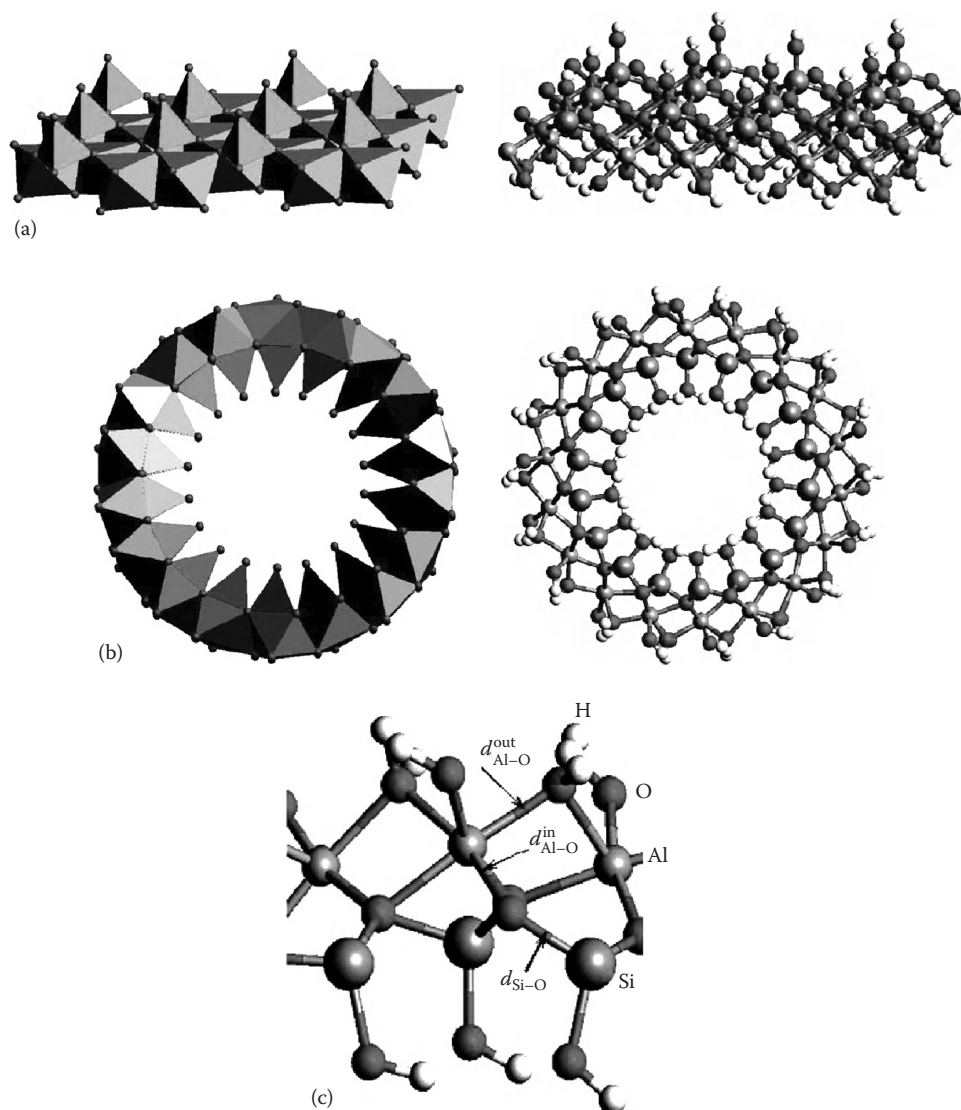
FIGURE 23.2 Polyhedral representations (left panel) and atomic structures (right panel) of (a) imogolite sheet and (b) imogolite nanotube with $N_\mu = 9$. (c) The local map of atomic structure. Si, Al, O, and H atoms and some representative bonds are indicated. (From Zhao, M., Y. Xia, and L. Mei. 2009. Energetic minimum structures of imogolite nanotubes: A first-principles prediction. *J. Phys. Chem. C* 113:14834–14837.)

The ^{27}Al-NMR spectra for various samples of synthetic and natural imogolite show little or no evidence of tetrahedrally coordinated Al (Figure 23.3). The ^{29}Si resonance line at -78 ppm of imogolite is consistent with Si bound to three Al–O and one OH group and serves to identify imogolite or proto-imogolite allophane in whole soils (Wilson et al., 1986).

The difference in size between natural and synthetic imogolite—2.3 vs. 2.7–3.2 nm—can be explained by density functional theory first-principle calculations that show two energy minima for the imogolite structure (Figure 23.4). The global minimum occurs with a gibbsite sheet consisting of 9 gibbsite units around the imogolite circumference and a local minimum with 12 gibbsite units (Zhao et al., 2009). These minima are consistent with the external diameters, circumferences, separation distances, and axial dimensions of imogolite determined experimentally by XRD and electron diffraction.

23.1.3 Structure of Allophane

Allophane refers to a group of SRO aluminosilicate clay minerals with no long-range order, only two diffuse XRD bands, and variable composition, generally ranging from 2:1 to 1:1 Al/Si molar ratio (Harsh et al., 2002). Endmember allophane would be completely disordered, but probably does not exist in nature (Fyfe et al., 1987). The further structural and chemical characterization of allophane found in natural environments has led to the identification of three major types of allophane. These have been tentatively named proto-imogolite allophane, halloysite-like (or defect kaolin) allophane, and hydrous feldspathoid allophane.

Proto-imogolite allophane (Al-rich allophane) has an Al/Si ratio close to 2, a gibbsite-like sheet of octahedrally coordinated Al, and orthosilicate groups sharing three oxygens with Al. Thus, proto-imogolite allophane has the same short-range order
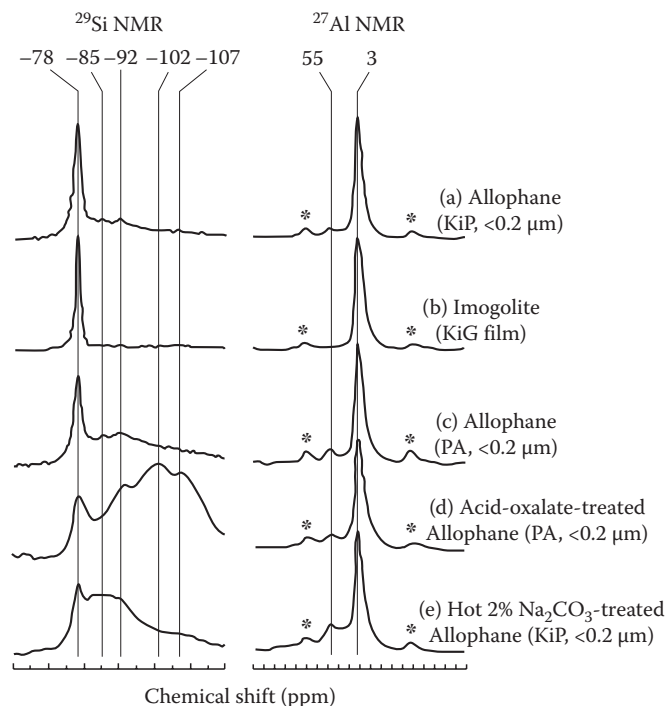
FIGURE 23.3 ^{29}Si and ^{27}Al MAS NMR spectra of (a) Al-rich KiP allophane (<0.2 μm), (b) imogolite (KiG) film, (c) Si-rich PA allophane (<0.2 μm), (d) acid-oxalate-treated Si-rich PA allophane, and (e) hot 2% Na_2CO_3-treated Al-rich KiP allophane. *, Spinning side band (SSB). (From Hiradate, S., and S.-I. Wada. 2005. Weathering process of volcanic glass to allophane determined by 27 Al and 29 Si solid-state NMR. Clays Clay Miner. 53:401–408. Figure 3. Reproduced with kind permission of The Clay Minerals Society, publisher of *Clays and Clay Minerals*.)

as imogolite, but does not exhibit the tubular morphology. Instead, it first forms fragments of the imogolite structure, then, depending on solution conditions, forms imogolite or spherical allophane particles about 3.5 nm in diameter (Farmer and Russell, 1990). The NMR and FTIR spectra are nearly identical to those for imogolite (Figure 23.3). The XRD "peaks" are very broad and the pattern lacks several lines indicative of imogolite (Table 23.1).

According to one view, allophane with an Al/Si ratio near 1:1 probably has a structure that is closer to kaolinite or halloysite with defects in the tetrahedral sheet. Infrared spectra of Si-rich allophanes indicate the presence of polymerized silica tetrahedra and octahedrally coordinated Al (Parfitt et al., 1980; Farmer and Russell, 1990; Parfitt, 1990). Both IR and NMR spectra suggest that Si-rich allophanes often contain both the halloysite and imogolite structural units. The halloysite-like feature in the NMR spectrum is characterized by a broad ^{29}Si resonance around −86 ppm. Varying amounts of the two types of allophane could account for a large variation in the Al/Si ratios (from 2.5 to 1) when found in weathered pumice. Recent solid-state NMR (MacKenzie et al., 1991) and x-ray photoelectron spectroscopy (He et al., 1995) studies of the silica-rich allophane support the presence of both imogolite-like and defect kaolin-like structures.

Recent Al-K edge x-ray absorption near edge structure (XANES) analysis of Al-rich allophanes and imogolite, raises
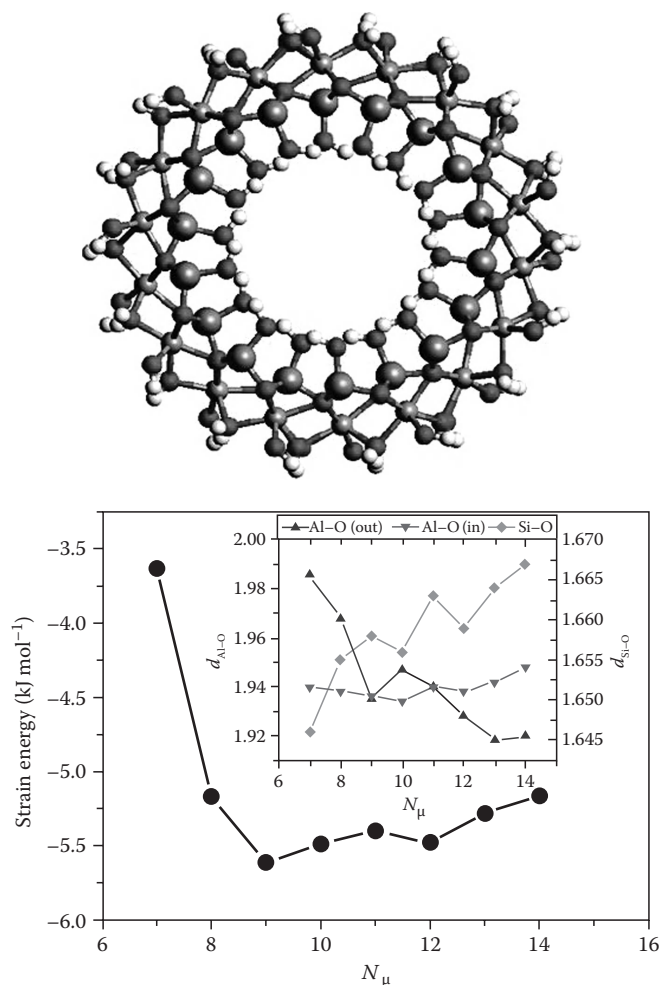
FIGURE 23.4 Variation of the strain energy of imogolite nanotubes (cross section shown above) relative to imogolite sheet as a function of $N_μ$ (number of gibbsite units in the circumference). The error bars (<0.003 kJ mol^{-1}) are smaller than symbols. The bond length evolution of Al–O and Si–O bonds is plotted in the inset of this figure. (From Zhao, M., Y. Xia, and L. Mei. 2009. Energetic minimum structures of imogolite nanotubes: A first-principles prediction. J. Phys. Chem. C 113:14834–14837.)

questions about the above models. Specifically, Al-K-XANES shows only one octahedral Al site in both imogolite and allophane, regardless of Al/Si ratio (Ildefonse et al., 1994). In contrast, gibbsite and 1:1 clay minerals have two octahedral Al sites, while 2:1 phyllosilicates have one. This implies allophane and imogolite have a dioctahedral sheet that resembles that of pyrophyllite or mica rather than kaolinite.

Another Si-rich allophane with a significant, if not dominant, amount of tetrahedrally coordinated Al resembles a hydrous feldspathoid and contains no imogolite units. The basic structure appears to be that of a framework silicate with 1:3 Al for Si substitution. Octahedral Al neutralizes some of the negative charge generated by the substitution and may be associated with the inner surfaces of spherical particles. Particles isolated from a stream deposit in New Zealand are less than 3 nm in diameter and form in CO_2-charged water that increases in pH as it

degases (Wells et al., 1977; Childs et al., 1990). Similar particles can be synthesized in neutral to alkaline solutions and IR and NMR spectra are consistent with a hydrous feldspathoid structure (Farmer et al., 1979a; Wada and Wada, 1981). No –78 ppm resonance is observed in the ^{29}Si NMR spectra, as in imogolite, and the changes in NMR spectra with heating confirm that the hydrous feldspathoid allophane differs in structure from both proto-imogolite and kaolin defect allophane (MacKenzie et al., 1991; Ildefonse et al., 1994). Ildefonse et al. (1994) used ^{27}Al NMR and Al-K XANES to characterize Si-rich allophanes from soils and found them to be similar to those found in stream deposits, suggesting that Si activity in solution determines their formation.

23.2 Identification and Synthesis of Allophane and Imogolite

23.2.1 Imogolite Identification and Synthesis

Imogolite can be separated from soils and pumice deposits after removing material that strongly interacts with its surface. Pretreatment usually includes a dithionite–citrate–bicarbonate extraction to remove iron oxides and hydrogen peroxide to remove organic matter. Neither significantly dissolves nor alters imogolite or allophane. Following pretreatment, ultrasonic dispersion of the suspended pumice or soil at low pH (e.g., pH 4 HCl) separates imogolite from phyllosilicate clay minerals. As discussed in Section 23.4.1, imogolite surfaces are positively charged in acid solutions and form stable suspensions, while the 2:1 phyllosilicates flocculate at low pH. Further concentration of SRO materials is accomplished by separating the <0.08 μm fraction because average allophane and imogolite particle sizes are much less than those of the 1:1 phyllosilicates. Of course, no purification procedure is ideal; unwanted minerals and some alteration of the target material will always be present.

While spectroscopic methods elucidate imogolite structure and often aid in quantitative methods, electron microscopy is the most reliable indicator of its presence in soils or successful synthesis in the laboratory. Both synthetic imogolite and that isolated from a pumice bed show long, bundled tubes (Figure 23.1), whereas, imogolite isolated from a forest soil is somewhat more fragmented (Su and Harsh, 1996).

Refluxing an acidic solution with millimolar concentrations of aluminum perchlorate and tetraethyl orthosilicate for 5 days at 95°C produces imogolite nearly identical to the natural material isolated from soils or pumice except for the larger diameter mentioned above (Farmer et al., 1983). The larger diameter may be due, in part, to the high-temperature synthesis, because imogolite prepared at room temperature has a diameter of about 23 Å (Wada, 1987). Levard et al. (2009) synthesized imogolite from decimolar concentrations of Al and Si, perhaps opening the door to manufacture of large quantities of material.

23.2.2 Allophane Identification and Synthesis

Separation of allophane from natural deposits, including soils, is similar to that of imogolite. Allophane is positively charged in a low pH solution and will form a stable suspension along with imogolite when ultrasonically dispersed. The same pretreatment to remove iron oxides and organic material applies. With allophane, however, separation from imogolite is possible in principle because it is negatively charged in alkaline solutions where imogolite flocculates.

After separation, one can discern allophane in electron micrographs, but it is far less distinctive than imogolite (Figure 23.4). It generally occurs as an amorphous mass of material coating other particles. Often the spheroidal morphology is evident in the Al-rich allophane. The spheroids of Al-rich allophane form aggregates that do not disperse easily. Damage from the electron beam is rapid and care must be taken to obtain good micrographs.

Like imogolite, allophane is synthesized from millimolar solutions of Al and monomeric Si(OH)$_4$. Allophane can be precipitated rapidly from solutions at room temperature. Methods, including heated synthesis, have been developed to precipitate allophane from concentrated solutions—10–100 mM (Ohashi et al., 1998; Montarges-Pelletier et al., 2005). The Al/Si ratio and pH of the matrix solution determine the nature of the products. Al-rich allophane forms at room temperature from an acidic Al and orthosilicate solution with Al/Si ratio near 2:1. The presence of Ca and Mg ions favors allophane formation over imogolite, whereas Na and K do not (Abidin et al., 2007). Because the acidic environment keeps Al in octahedral coordination, decreasing the Al/Si ratio while keeping the pH below neutral probably favors the defect kaolinite structure. Between pH 6 and 7, the hydrous feldspathoid structure predominates with the proportion of tetrahedral Al increasing as the Al/Si ratio decreases (Farmer and Russell, 1990). This structure exists in natural alkaline environments, including the Silica Springs deposit in New Zealand (Wells et al., 1977) and many soils (Childs et al., 1990; Ildefonse et al., 1994). Farmer and Russell (1990) reviewed the synthesis and occurrence of imogolite and the three types of allophane described here.

23.3 Occurrence of Imogolite and Allophane in Natural Environments

Allophane and imogolite tend to be associated with volcanic ash deposits and soils derived from volcanic debris, because the rapid release of Al and Si from materials such as volcanic glass results in the precipitation of SRO aluminosilicates. The Kitakami pumice beds in Japan have long been a source of relatively pure separates of both allophane and imogolite and are among the most studied of the naturally occurring SRO materials, including the KiG imogolite in Figure 23.3 (Miyauchi and Aomine, 1966; Wada and Matsubara, 1968; Yoshinaga, 1968; Yoshinaga et al., 1973). Imogolite was first identified in Ando (dark-colored) soils of Japan (Yoshinaga and Aomine, 1962) and allophane was found to give them many of their unique properties. Since then, SRO aluminosilicates have been identified in soils derived from many parent materials, including sandstone, gneiss, granite, and basalt (Harsh et al., 2002). In basalt, ferrihydrite, a poorly crystalline

iron oxide, often occurs in addition to allophane and imogolite. Nevertheless, the largest concentrations of SRO aluminosilicates occur in deposits of volcanic tephra and soils developed therein. Soils derived from volcanic tephra occur all over the world with well studied examples from New Zealand (Taylor, 1933; Wells et al., 1977; Parfitt, 1990), Chile (Vicente and Besoain, 1961; Besoain, 1969; Nissen and Kuehne, 1976; Besoain and Gonzalez, 1977; Floody et al., 2009), and the Cascade Range of North America (Ugolini and Dahlgren, 1991). A large deposit of allophane—more than 4000 km^2 and 16 m thick—was recently discovered in Ecuador and may be a source for commercial allophane (Kaufhold et al., 2009, 2010). McDowell and Hamilton (2009) report that SRO aluminosilicates are probably present on Mars and interfere with quantification of phyllosilicates on the surface.

The key factor in the formation of allophane and imogolite in soil is sufficient Al and Si in solution. Thus, conditions that lead to rapid weathering of primary minerals, such as high rainfall and low pH, favor their formation. Both imogolite and Al-rich allophane tend to dominate at low pH where Al/Si is high, whereas halloysite and/or Si-rich allophanes form when soluble Si is high ($>10^{-3.45}$). As a result, low rainfall or xeric-moisture regimes, concentrating soluble silica, have been observed to favor halloysite formation at the expense of allophane and imogolite (Parfitt and Wilson, 1985; Singleton et al., 1989; Takahashi et al., 1993). Noncrystalline minerals are found in Andisols, Inceptisols, Entisols, Spodosols, and Ultisols. Generally, allophane and imogolite occur in horizons where organic matter is low so that Al exists in inorganic complexes. Al-humus complexes, which give soils many of the same properties attributed to allophanic soils, often dominate organic and A horizons, particularly in forested soils (Dahlgren et al., 2004). Complexation of Al by organic acids and humic materials inhibits SRO aluminosilicate formation indicating that free Al^{3+} or Al-hydroxy complexes may be required.

Short-range ordered materials are generally less stable than their crystalline counterparts, particularly kaolinite. Their rapid formation results because the interfacial energy difference between mineral surfaces and an aqueous solution slows nucleation. Poorly crystalline surfaces have lower surface tensions and nucleate more easily, that is, at a lower saturation index than a crystalline mineral of similar composition. As a result, allophane and imogolite are often found as precursors to kaolinite in weathering soil profiles and may serve as templates for its heterogeneous nucleation (Steefel et al., 1990).

Quantification of SRO aluminosilicates in soils is achieved by selective chemical extraction and solution analysis of Al and Si (Dahlgren, 1994). Commonly, one extracts the soil first with Na-pyrophosphate, which effectively extracts Al and Fe from soil organic complexes, but dissolves little from the inorganic fraction. Then, an acid ammonium oxalate solution (0.2 M, pH 3) selectively dissolves allophane, imogolite, and poorly crystalline iron oxides such as ferrihydrite. Subtracting the organically complexed from the oxalate-extracted, gives the SRO-associated Al and Fe. Dividing this SRO Al value by the oxalate-extractable Si provides the average SRO Al/Si ratio (Parfitt and Wilson, 1985). A ratio near 2:1 suggests a soil dominated by imogolite

and proto-imogolite allophane whereas a smaller value indicates Si-rich allophane is present. Embryonic halloysite may also partially dissolve in the oxalate solution (Wada and Kakuto, 1985). Ratios greater than 2:1 may occur when other labile sources of Al are present, such as hydroxy-Al interlayers in clay minerals.

23.4 Surface Charge Characteristics of Short-Range Ordered Aluminosilicates and Variable-Charge Soils

23.4.1 Surface Charge Determination

23.4.1.1 Operational Definitions

Two major sources of surface charge on SRO aluminosilicates will be considered here, that is, permanent structural charge (σ_o) from isomorphic substitution of Al in Si tetrahedra and variable surface charge due to ion association at surface hydroxyl groups (Sposito, 2004). Ion associations with surface hydroxyls include complexed protons (σ_H), inner-sphere complexes with ions other than H^+ (σ_{IS}), and outer-sphere complexes (σ_{OS}). A charge balance equation can be written as follows:

$$\sigma_P = \sigma_o + \sigma_H + \sigma_{IS} + \sigma_{OS} = -\sigma_D \qquad (23.1)$$

where

  σ_P represents the particle charge or total charge from all sources

  σ_D is the diffuse layer charge that consists of the net charge from the swarm of hydrated ions around the particle

The proton charge on variable-charge surfaces is determined by potentiometric titration with H and OH, which adsorb to surface silanol ($\equiv Si-OH$) or aluminol ($\equiv Al-OH$) groups. Relative proton adsorption is easy to determine from the difference between added and remaining protons in solution, but the absolute proton charge on the surface is not evident because the initial charge, before titration, is not known. The point of zero net proton charge (PZNPC) is often estimated by the point of zero salt effect (PZSE)—the pH at the crossover point of three or more titrations at different background salt concentrations. If the background salt does not form inner-sphere complexes, only the proton charge is determined in the titration. The proton charge is more reliably determined by plotting the net adsorbed ion charge ($\sigma_{IS} + \sigma_{OS} + \sigma_D$) against proton charge (σ_H) (Chorover and Sposito, 1995). The slope should be −1 and the x- and y-intercepts equal to σ_o. The PZNPC is then the pH where ($\sigma_{IS} + \sigma_{OS} + \sigma_D$) = −$\sigma_o$.

Inner-sphere, outer-sphere, and diffuse layer cations and anions balance the total negative and positive charge on the surface. Thus, the point of zero net charge (PZNC) can be determined by ion adsorption across a range of pH in an electrolyte, such as NaCl, whose ions do not form inner-sphere complexes with the aluminol or silanol groups. Finally, electrophoretic mobility provides information about the tendency for ions to

form outer-sphere complexes with allophane and imogolite. Their mobility when suspended in an aqueous solution subjected to an electric field depends on the ionic strength of the solution and the nature of the surface complexes. The isoelectric point is the pH where the particles are stationary in the field and is an estimate of the PZC—the point where $\sigma_D = 0$.

With all Al in imogolite and proto-imogolite allophane in octahedral coordination, no structural charge arises from isomorphic substitution. Gustafsson (2001), however, proposed that structural charge could arise from differences in Al–O bond valences leading to a weak positive charge on outer tube walls and negative charge in tubular pores. This hypothesis was later supported by quantum mechanical calculations of charge distribution in imogolite (Guimaraes et al., 2007). The Si-rich allophane, on the other hand, has structural charge arising from Al for Si substitution. However, Su et al. (1992) found in synthetic allophanes of this type that structural charge from tetrahedral Al was not completely balanced by exchangeable cations. In this case, it was hypothesized that Al present as hydroxy complexes or polymers balanced much of the negative σ_o.

The charge characteristics of imogolite and some synthetic allophanes as determined by these methods are summarized in Figure 23.5 and Table 23.2. Because the PZNC and PZSE values of imogolite and Al-rich allophane are all greater than 6, these
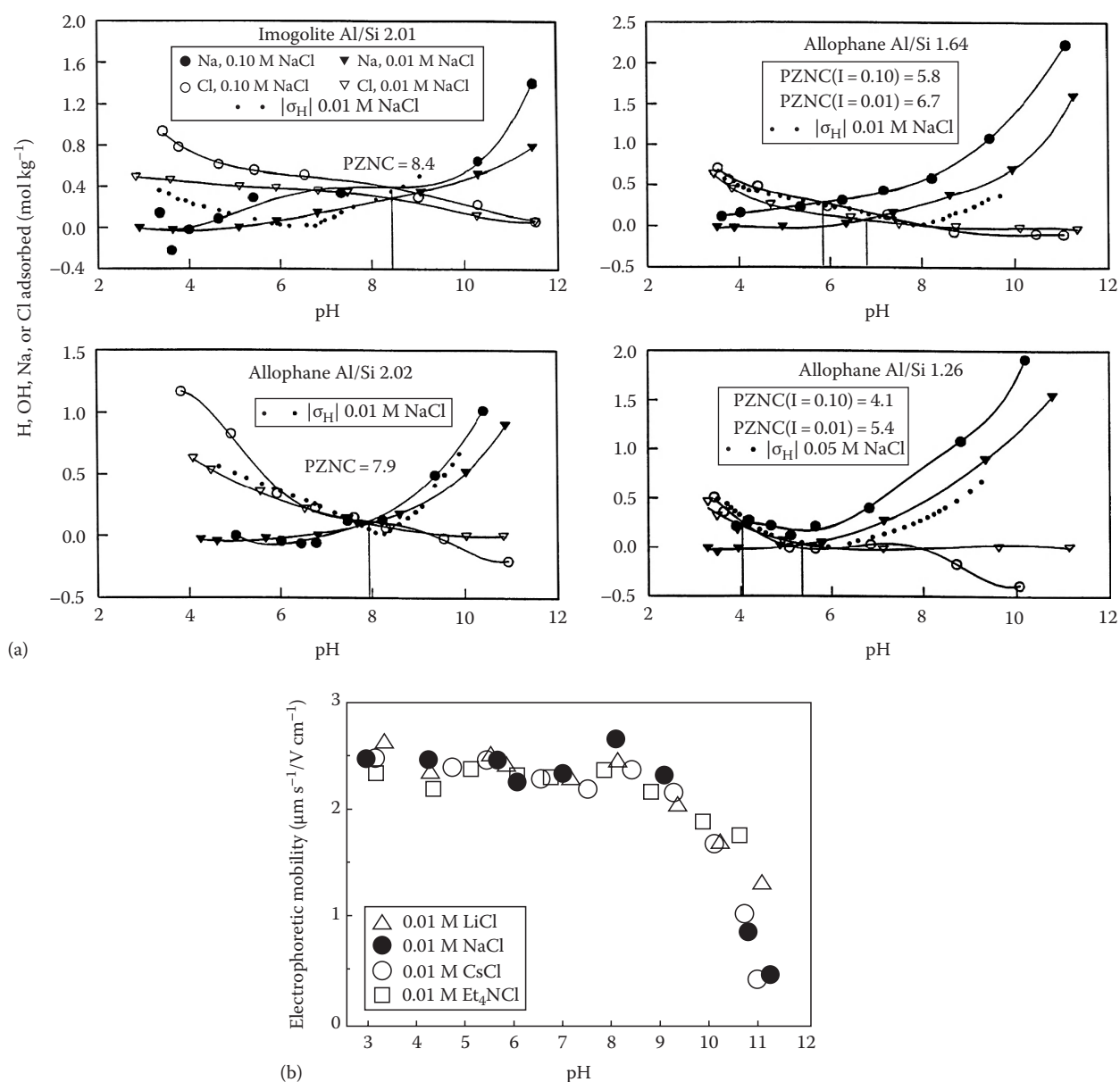
FIGURE 23.5 Charge properties of synthetic imogolite and allophane determined by (a) potentiometric titration and ion adsorption (From Su, C.M. et al., *Clays Clay Miner.*, 40, 280, 1992, Figure 1. Reproduced with kind permission of The Clay Minerals Society, publisher of *Clays and Clay Minerals*.) (b) electrophoretic mobility (From Harsh, J.B., S.J. Traina, J. Boyle, and Y. Yang. 1992. Adsorption of cations on imogolite and their effect on surface-charge characteristics. Clays Clay Miner. 40:700–706, Figure 4. Reproduced with kind permission of The Clay Minerals Society, publisher of *Clays and Clay Minerals*.)

		PZNC in NaCl			
Material	Al/Si	0.10 M	0.01 M	PZSE	PZC
Imogolite	2.0	8.4	8.4	6.5	>10
Allophane	2.0	7.9	7.9	8.3	10
Allophane	1.6	5.8	6.7	7.7	9
Allophane	1.2	4.1	5.4	5.9	7.6

Source: Su, C.M., J.B. Harsh, and P.M. Bertsch. 1992. Sodium and chloride sorption by imogolite and allophanes. Clays Clay Miner. 40:280–286.

materials impart positive charge in acid soils. The Si-rich allophanes may be either positively or negatively charged depending on soil pH. The Si-rich allophanes also have a PZSE greater than the PZNC, another indicator that permanent negative charge exists in these materials as a result of Al for Si substitution. Imogolite has a PZSE greater than its PZNC, constant adsorption of Cl over a large pH range, and an isoelectric point above pH 10 (Harsh et al., 1992; Tsuchida et al., 2005), all of which might be due in part to cation distribution inside imogolite tubes as well as structural charge distribution from bond valence differences as discussed above (Gustafsson, 2001).

23.4.2 Interaction with Anions and Cations

Inner-sphere complexation of anions on allophane and imogolite is common. Phosphate, fluoride, citrate, borate, arsenate, and selenite are known to form inner-sphere complexes with SRO aluminosilicates. This reaction contributes to σ_{IS} and can shift the points of zero charge (PZCs) of the material. As a result, more negative charge is likely in the presence of these anions, particularly in acid soils. Caution should be used in interpreting any strong sorption as surface complexation. Veith and Sposito (1977) and Su and Harsh (1993) showed that phosphate and fluoride, respectively, may react with SRO aluminosilicates to form new solid phases. Boron may substitute into tetrahedral sites in coprecipitation with allophane (Su and Suarez, 1997).

In soils containing allophane and imogolite, strong interaction with phosphate can lead to deficiency of this macronutrient in crops. Calcium silicate can be added to such soils to compete for sorption sites and enhance phosphate availability. The ability of SRO materials and allophanic soils to adsorb anions by ligand or anion exchange has led to suggestions that they be used to remove contaminants such as phosphate, selenium, technetium, glyphosate, chromium, arsenic, and iodine from wastewaters (Wells and Parfitt, 1987; Gu and Schulze, 1991; Babel and Opiso, 2007; Gimsing et al., 2007; Hopp et al., 2008; Opiso et al., 2009; Ballantine and Tanner, 2010).

Many trace metals and metalloids form surface complexes with aluminol and silanol groups as evidenced by extensive studies on silica and aluminum oxides; however, fewer studies have been performed on SRO aluminosilicates or soils containing them. One study of metal adsorption to a soil dominated

by allophane and imogolite in the clay fraction showed the following order of decreasing affinity, which is quite similar to the selectivity of aluminol groups for divalent metals (Abd-Elfattah and Wada, 1981):

$$Pb > Cu > Zn > Co \approx Cd > Mg \approx Ca$$
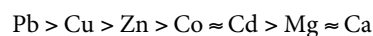
As on Al oxides, Pb, Cu, Zn, and Co appear to form inner-sphere surface complexes, whereas Cd, Ca, and Mg are held by electrostatic forces (Clark and McBride, 1984; Yuan et al., 2002). Alkaline earth and alkali metals, halide anions, NO_3, SO_4, and ClO_4 will generally exist as outer-sphere complexes or in the diffuse layer of allophane and imogolite. Adsorption of exchangeable cations on allophanic soils appears to be similar to that on soils dominated by smectites (Nakahara and Wada, 1994). Highly selective K exchange on Andisols has been reported (Espino-Mesa and Hernández-Moreno, 1994), but this could result from trace amounts of illitic minerals or alunite formation in addition to reactions with allophane or imogolite. Allophanic soils do seem capable of adsorbing large amounts of nitrate below the root zone, presumably through interaction with SRO aluminosilicates (Maeda et al., 2008). Circumstantial evidence exists for specific adsorption of Cl on imogolite, as discussed above, but this issue has not been resolved and is not consistent with electrophoretic measurements (Figure 23.5).

The interaction of allophane and allophanic soils with organic matter (discussed in the following section) also affects the surface charge and sorption properties of the materials. Organic matter can be expected to increase the negative charge on minerals dominated by aluminol surface groups because of the carboxylate groups of humic materials. Treatment of allophanic materials and soils with hydrogen peroxide to remove organic matter increases the PZC (Escudey et al., 1986) and decreases phosphate sorption (Mora and Canales, 1995), respectively. Conversely, adding iron oxides increases the PZC of allophane, rendering the surface more positively charged (Escudey and Galindo, 1983). Iron oxides may react selectively with silanol groups, which have an acid PZNPC. The PZNPCs of Fe and Al oxides, on the other hand, are greater than 6.

23.5 Interaction of Allophane and Imogolite with Other Soil Constituents

23.5.1 Organic Matter and Iron Oxides

The term "Andisol" comes from the Japanese words *ando* and *sol*, meaning dark-colored soil. The dark color comes from the fact that soils formed from volcanic debris are often high in organic matter, especially in surface horizons. Noncrystalline materials may play a role in organic matter retention through one or more mechanisms. First, humic substances adsorb strongly to imogolite and allophane (Parfitt et al., 1977; Parfitt, 2009) and sorption is possibly responsible for inhibiting degradation (Basile-Doelsch et al., 2005; Schneider et al., 2010). Second, reactive Al released

from SRO materials complexes with humic and fulvic acids and leads to precipitation and/or inhibition of degradation (Boudot, 1992; Wada, 1995; Matus et al., 2008). Finally, SRO materials may enhance the "humification" of soil organic carbon by facilitating bonding reactions or molecular assemblages (Gonzalez-Perez et al., 2007; Fukushima et al., 2009). In the presence of hydrogen peroxide, allophane associated with Fe may catalyze the degradation of organic materials, including xenobiotics, by Fenton reactions (Garrido-Ramirez et al., 2010).

Reaction with organic matter may give soils containing SRO aluminosilicates an important role in carbon sequestration. If these materials are instrumental in increasing the residence time of organic matter in soils, then preserving the soils containing them may be beneficial in order to reduce emission of greenhouse gases such as CO_2. Dahlgren et al. (2004) stated that 0.8% of the world's soils are allophanic and that they contain 5% of the global soil carbon. In his excellent review of the biogeochemistry of allophane and imogolite, Parfitt (2009) points out that the stabilization of organic matter is partially maintained as forests are converted to pasture, but does not apply to fresh organic matter or plant-derived material. The recalcitrant carbon originates from microbial byproducts, is stabilized slowly, and persists for thousands of years.

23.5.2 Water

Both a high specific surface area and a high value of the relative microporosity characterize allophane. The latter probably
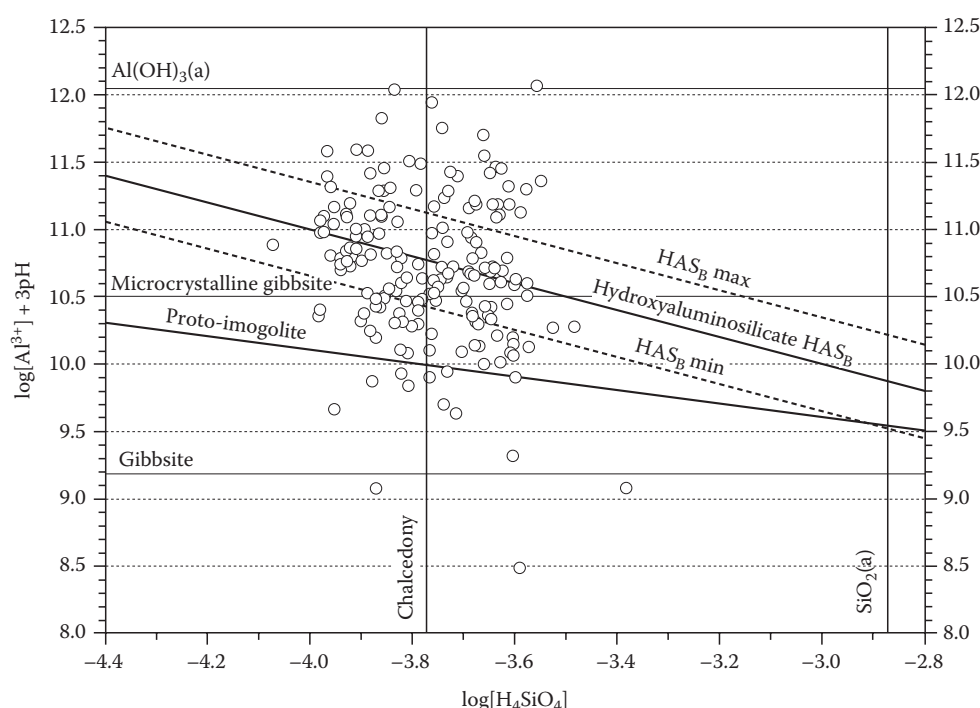
FIGURE 23.6 Stability relations for proto-imogolite (Lumsdon and Farmer, 1995), Si-rich allophane (HAS_B), gibbsite, and microcrystalline gibbsite. (From Dobrzynski, D. 2007. Chemistry of neutral and alkaline waters with low Al^{3+} activity against hydroxyaluminosilicate HASB solubility. The evidence from ground and surface waters of the Sudetes Mts. (SW Poland). Aquat. Geochem. 13:197–210.)

(Two-ion parameter $log[Al^{3+}] + 3pH$ versus $log[H_4SiO_4]$ in waters with $log[Al^{3+}] < -10$ (size—188 samples). Solubility lines for amorphous $Al(OH)_3$ and SiO_2 forms plotted at 7°C, after solubility constant for gibbsite ($log\ K^{25} = 8.11$), microcrystalline gibbsite ($log\ K^{25} = 9.35$), and $Al(OH)_3(a)$ ($log\ K^{25} = 10.8$) calculated using the van't Hoff equation. Solubility constant $log\ K^7$ for chalcedony and amorphous SiO_2 calculated using the analytical expression after Nordstrom et al. (1990).

Proto-imogolite solubility line at 7°C plotted for the reaction:

$$0.5Al_2SiO_3(OH)_4(s) + 3H^+ = Al^{3+} + 0.5H_4SiO_4 + 1.5H_2O$$
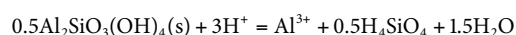
($log\ K^{25} = 7.02$, $\Delta H = -96.8\ kJ\ mol^{-1}$) after Lumsdon and Farmer (1995).

HASB solubility line for the reaction:

$$Al_2Si_2O_5(OH)_4 + 6H^+ = 2Al^{3+} + 2H_4SiO_4 + H_2O\ (log\ K^7 = 14.0 \pm 0.7)$$

with uncertainty of solubility plotted.)

results largely from micropores formed from aggregates of particles because the porosity decreases with the shrinkage observed on drying. As a result, water retention in allophanic soils is very high at a given water potential compared to soils dominated by crystalline clays. This, in part, accounts for the high productivity often seen in forested soils high in SRO materials (Martini and Luzuriaga, 1989; Nizeyimana, 1997). Macroporosity is also high in most allophanic soils resulting in good infiltration and hydraulic conductivity. The latter parameter has been found to decrease significantly in soils at pH values where allophane is known to disperse easily (pH = 3 and 11) (Nakagawa and Ishiguro, 1994).

The aqueous solubility of imogolite and proto-imogolite allophane shows them to be less stable than kaolinite at soluble silica activities expected in natural soils (Farmer et al., 1979b; Percival, 1991; Su and Harsh, 1994; Gustafsson et al., 1998; Su and Harsh, 1998; Wada and Kakuto, 1999). However, imogolite and proto-imogolite allophane are more stable than halloysite at low Si activity (Su and Harsh, 1994). Thermodynamically, halloysite is stable relative to imogolite at a log $Si(OH)_4$ activity greater than about −3.5. Halloysite is more stable than allophane at a lower Si activity, but allophane formation in preference to halloysite is commonly observed in soils at log $(H_4SiO_4) > −3.45$ (Singleton et al., 1989). This incongruity may indicate that soil solution analyses performed at a later date may not be relevant to the time of mineral formation. Kinetic considerations may favor the SRO materials relative to halloysite; this certainly is true relative to kaolinite. The solubility constants determined for imogolite and allophane may be incorrect or the solid phases controlling the solubility of Al, such as gibbsite or Al complexed with organic matter, may not be well understood (Gustafsson et al., 2001). Current solubility data seem to support different solubility constants for synthetic and soil imogolite—log K_{syn} = 6.05 and log K_{soil} = 6.60 at 25°C for the reaction:

$$0.5(HO)_3Al_2O_3SiOH(s) + 3H^+ = Al^{3+} + 0.5H_4SiO_4 + 1.5H_2O$$

(23.2)
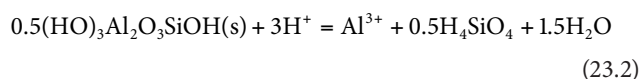
The solubility of minerals in soils is complicated by nonequilibrium conditions, difficulty in speciating soluble ions, particularly Al, and determination of controlling phases (Gustafsson et al., 2001). Nevertheless, the chemistry of soil solutions and natural waters may often be determined by these rapidly formed, but persistent, poorly crystalline materials (Dobrzynski, 2007) (Figure 23.6).

References

Abd-Elfattah, A., and Wada, K. (1981). Adsorption of lead, copper, zinc, cobalt, and cadmium by soils that differ in cation-inchange materials. J. Soil Sci. 32:271–283.

Abidin, Z., N. Matsue, and T. Henmi. 2007. Differential formation of allophane and imogolite: Experimental and molecular orbital study. J. Comput. Aided Mater. Des. 14:5–18.

Babel, S., and E.M. Opiso. 2007. Removal of Cr from synthetic wastewater by sorption into volcanic ash soil. Int. J. Environ. Sci. Technol. 4:99–107.

Ballantine, D.J., and C.C. Tanner. 2010. Substrate and filter materials to enhance phosphorus removal in constructed wetlands treating diffuse farm runoff: A review. N.Z. J. Agric. Res. 53:71–95.

Basile-Doelsch, I., R. Amundson, W.E.E. Stone, C.A. Masiello, J.Y. Bottero, F. Colin, F. Masin, D. Borschneck, and J.D. Meunier. 2005. Mineralogical control of organic carbon dynamics in a volcanic ash soil on La Reunion. Eur. J. Soil Sci. 56:689–703.

Besoain, E. 1969. Imogolite [allophane] in volcanic soils in Chile. Geoderma 2:151–169.

Besoain, M.E., and M.S.P. Gonzalez. 1977. Mineralogy, genesis and classification of volcanic ash soil derivatives of the central-south region of Chile. Cienc. Invest. Agrar. 4:109–130.

Bonelli, B., I. Bottero, N. Ballarini, S. Passeri, F. Cavani, and E. Garrone. 2009. IR spectroscopic and catalytic characterization of the acidity of imogolite-based systems. J. Catal. 264:15–30.

Boudot, J.P. 1992. Relative efficiency of complexed aluminum, noncrystalline Al hydroxide, allophane and imogolite in retarding the biodegradation of citric acid. Geoderma 52:29–39.

Childs, C.W., R.L. Parfitt, and R.H. Newman. 1990. Structural studies of Silica Springs allophane. Clay Miner. 25:329–341.

Chorover, J., and G. Sposito. 1995. Surface-charge characteristics of kaolinitic tropical soils. Geochim. Cosmochim. Acta 59:875–884.

Clark, C.J., and McBride, M.B. (1984). Chemisorption of Cu(II) and Co(II) on allophane and imogolite. Clays Clay Miner. 32:300–310.

Cradwick, P.D.G., V.C. Farmer, J.D. Russell, C.R. Masson, K. Wada, and N. Yoshinaga. 1972. Imogolite, a hydrated aluminum silicate of tubular structure. Nat. Phys. Sci. 240:187–189.

Creton, B., D. Bougeard, K.S. Smirnov, J. Guilment, and O. Poncelet. 2008. Molecular dynamics study of hydrated imogolite. 1. Vibrational dynamics of the nanotube. J. Phys. Chem. C 112:10013–10020.

Cruz, M.D.R., and L.M. Real. 1991. Practical determination of allophane and synthetic alumina and iron-oxide gels by x-ray-diffraction. Clay Miner. 26:377–387.

Dahlgren, R.A. 1994. Quantification of allophane and imogolite, p. 430–451. *In* J. Amonette and L.W. Zelazny (eds.) Quantitative methods in soil mineralogy. SSSA, Madison, WI.

Dahlgren, R.A., M. Saigusa, and F.C. Ugolini. 2004. The nature, properties and management of volcanic soils. Adv. Agron. 82:113–182.

Dobrzynski, D. 2007. Chemistry of neutral and alkaline waters with low Al^{3+} activity against hydroxyaluminosilicate HASB solubility. The evidence from ground and surface waters of the Sudetes Mts. (SW Poland). Aquat. Geochem. 13:197–210.

Escudey, M., and G. Galindo. 1983. Effect of iron-oxide coatings on electrophoretic mobility and dispersion of allophane. J. Colloid Interface Sci. 93:78–83.

Escudey, M., G. Galindo, and J. Ervin. 1986. Effect of iron oxide dissolution treatment on the isoelectric point of allophanic soils. Clays Clay Miner. 34:108–110.

Espino-Mesa, M., and J.M. Hernández-Moreno. 1994. Potassium selectivity in Andic soils in relation to induced acidity, sulphate status and layer silicates. Geoderma 61:191–201.

Farmer, V.C., M.J. Adams, A.R. Fraser, and F. Palmieri. 1983. Synthetic imogolite: Properties, synthesis, and possible applications. Clay Miner. 18:459–472.

Farmer, V.C., A.R. Fraser, J.D. Russell, and N. Yoshinaga. 1977. Recognition of imogolite structures in allophanic clays by infrared spectroscopy. Clay Miner. 12:55–57.

Farmer, V.C., and J.D. Russell. 1990. Structures and genesis of allophanes and imogolite and their distribution in non-volcanic soils, p. 165–178. *In* M.F. De Boodt, M.H.B. Hayes and A. Herbillon (eds.) Soil colloids and their associations in aggregates. Plenum Press, New York.

Farmer, V.C., J.D. Russell, and M.L. Berrow. 1979a. Characterization of the chemical structure of natural and synthetic aluminosilicate gels and sols by infrared spectroscopy. Geochim. Cosmochim. Acta 43:1417–1420.

Farmer, V.C., B.F.L. Smith, and J.M. Tait. 1979b. The stability, free energy and the heat of formation of imogolite. Clay Miner. 14:103–107.

Floody, M.C., B.K.G. Theng, P. Reyes, and M.L. Mora. 2009. Natural nanoclays: Applications and future trends—A Chilean perspective. Clay Miner. 44:161–176.

Fukushima, M., A. Miura, M. Sasaki, and K. Izumo. 2009. Effect of an allophanic soil on humification reactions between catechol and glycine: Spectroscopic investigations of reaction products. J. Mol. Struct. 917:142–147.

Fyfe, C., L. Evans, W. Chesworth, J. Graham, and M. McBride. 1987. Operation definition of imogolite/allophane phases by multitechnique analysis, p. 829–836. *In* C.R. Rodriquez and Y. Tardy (eds.) Geochemistry and mineral formation in the earth surface. Cons. Super. Invest. Cient, Spain. CNRS, France.

Garrido-Ramirez, E.G., B.K.G. Theng, and M.L. Mora. 2010. Clays and oxide minerals as catalysts and nanocatalysts in Fenton-like reactions—A review. Appl. Clay Sci. 47:182–192.

Gimsing, A.L., C. Szilas, and O.K. Borggaard. 2007. Sorption of glyphosate and phosphate by variable-charge tropical soils from Tanzania. Geoderma 138:127–132.

Gonzalez-Perez, J.A., C.D. Arbelo, F.J. Gonzalez-Vila, A.R. Rodriguez, G. Almendros, C.M. Armas, and O. Polvillo. 2007. Molecular features of organic matter in diagnostic horizons from andosols as seen by analytical pyrolysis. J. Anal. Appl. Pyrol. 80:369–382.

Goodman, B.A., J.D. Russell, B. Montez, E. Oldfield, and R.J. Kirkpatrick. 1985. Structural studies of imogolite and allophanes by aluminum-27 and silicon-29 nuclear magnetic resonance spectroscopy. Phys. Chem. Miner. 12:342–346.

Greenland, D.H. 1982. Chemistry and the soil environment—Surfaces and sorption processes, p. 99–110. *In* K.J. Laider (ed.) IUPAC frontiers in chemistry. IRRI, Manila, Philippines.

Gu, B., and R.K. Schulze. 1991. Anion retention in soil: Possible application to reduce migration of buried technetium and iodine. NUREG/CR-5464. U.S. Nuclear Regulatory Commission, Washington, DC.

Guimaraes, L., A.N. Enyashin, J. Frenzel, T. Heine, H.A. Duarte, and G. Seifert. 2007. Imogolite nanotubes: Stability, electronic, and mechanical properties. ACS Nano 1:362–368.

Gustafsson, J.P. 2001. The surface chemistry of imogolite. Clays Clay Miner. 49:73–80.

Gustafsson, J.P., D. Berggren, M. Simonsson, M. Zysset, and J. Mulder. 2001. Aluminium solubility mechanisms in moderately acid Bs horizons of podzolized soils. Eur. J. Soil Sci. 52:655–665.

Gustafsson, J.P., D.G. Lumsdon, and M. Simonsson. 1998. Aluminium solubility characteristics of spodic B horizons containing imogolite-type materials. Clay Miner. 33:77–86.

Harsh, J.B., J. Chorover, and E. Nizeyimana. 2002. Allophane and imogolite, p. 291–322. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA Book Series No. 7. SSSA, Madison, WI.

Harsh, J.B., S.J. Traina, J. Boyle, and Y. Yang. 1992. Adsorption of cations on imogolite and their effect on surface-charge characteristics. Clays Clay Miner. 40:700–706.

He, H., T.L. Barr, and J. Klinowski. 1995. ESCA and solid-state NMR studies of allophane. Clay Miner. 30:201–209.

Hiradate, S., and S.-I. Wada. 2005. Weathering process of volcanic glass to allophane determined by 27Al and 29Si solid-state NMR. Clays Clay Miner. 53:401–408.

Hopp, L., P.S. Nico, M.A. Marcus, and S. Peiffer. 2008. Arsenic and chromium partitioning in a podzolic soil contaminated by chromated copper arsenate. Environ. Sci. Technol. 42:6481–6486.

Ildefonse, P., R.J. Kirkpatrick, B. Montez, G. Calas, A.M. Flank, and P. Lagarde. 1994. 27Al MAS NMR aluminum x-ray absorption near edge structure study of imogolite and allophanes. Clays Clay Miner. 42:276–287.

Kaufhold, S., A. Kaufhold, R. Jahn, S. Brito, R. Dohrmann, R. Hoffmann, H. Gliemann, P. Weidler, and M. Frechen. 2009. A new massive deposit of allophane raw material in Ecuador. Clays Clay Miner. 57:72–81.

Kaufhold, S., K. Ufer, A. Kaufhold, J.W. Stucki, A.S. Anastácio, R. Jahn, and R. Dohrmann. in press. Quantification of allophane clay from Ecuador. Clays Clay Miner. 58:707–716.

Levard, C., A. Masion, J. Rose, E. Doelsch, D. Borschneck, C. Dominici, F. Ziarelli, and J.-Y. Bottero. 2009. Synthesis of imogolite fibers from decimolar concentration at low temperature and ambient pressure: A promising route for inexpensive nanotubes. J. Am. Chem. Soc. 131:17080–17081.

Lumsdon, D.G., and V.C. Farmer. 1995. Solubility characteristics of proto-imogolite sols—How silicic-acid can de-toxify aluminum solutions. Eur. J. Soil Sci. 46:179–186.

MacKenzie, K.J.D., M.E. Bowden, and R.H. Meinhold. 1991. The structure and thermal transformations of allophanes studied by silicon-29 and aluminum-27 high-resolution solid-state NMR. Clays Clay Miner. 39:337–346.

Maeda, M., H. Ihara, and T. Ota. 2008. Deep-soil adsorption of nitrate in a Japanese andisol in response to different nitrogen sources. Soil Sci. Soc. Am. J. 72:702–710.

Martini, J.A., and C. Luzuriaga. 1989. Classification and productivity of six Costa Rican andepts. Soil Sci. 147:326–338.

Matus, F., E. Garrido, N. Sepulveda, I. Carcamo, M. Panichini, and E. Zagal. 2008. Relationship between extractable Al and organic C in volcanic soils of Chile. Geoderma 148:180–188.

McDowell, M.L., and V.E. Hamilton. 2009. Seeking phyllosilicates in thermal infrared data: A laboratory and Martian data case study. J. Geophys. Res. 114:E06007/1–E06007/21.

Miyauchi, N., and S. Aomine. 1966. Mineralogy of gel-like substance in the pumice bed in Kanuma and Kitakami districts. Soil Sci. Plant Nutr. (Tokyo) 12:187–190.

Montarges-Pelletier, E., S. Bogenez, M. Pelletier, A. Razafitianamaharavo, J. Ghanbaja, B. Lartiges, and L. Michot. 2005. Synthetic allophane-like particles: Textural properties. Colloids Surf., A 255:1–10.

Mora, M.L., and J. Canales. 1995. Humin-clay interactions on surface reactivity in chilean Andisols. Commun. Soil Sci. Plant Anal. 26:2819–2828.

Mukherjee, S., V.A. Bartlow, and S. Nair. 2005. Phenomenology of the growth of single-walled aluminosilicate and aluminogermanate nanotubes of precise dimensions. Chem. Mater. 17:4900–4909.

Nakagawa, T., and M. Ishiguro. 1994. Hydraulic conductivity of an allophanic andisol as affected by solution pH. J. Environ. Qual. 23:208–210.

Nakahara, O., and S.I. Wada. 1994. Ca^{2+} and Mg^{2+} adsorption by an allophanic and a humic andisol. Geoderma 61:203–212.

Nissen M.J., and A.G. Kuehne. 1976. Mineralogical study of three forest soils from Valdivia Province, Chile. Agro Sur 4:1–11.

Nizeyimana, E. 1997. A toposequence of soils derived from volcanic materials in Rwanda: Morphological, chemical, and physical properties. Soil Sci. 162:350–360.

Nordstrom, D.K., Plummer, L.N., Langmuir, D., Busenberg, E., May, H.M., Jones, B.F., and Parkhurst, D.L. (1990). Revised chemical-equilibrium data for major water-mineral reactions and their limitations. ACS Symposium Series 416:398–413.

Ohashi, F., M. Maeda, M. Suzuki, S. Tomura, S. Wada, and Y. Kakuto. 1998. Synthesis of spherical hollow amorphous aluminosilicate using rapid mixing method. Natl Ind. Res. Inst. Nagoya 46:397–404.

Opiso, E., T. Sato, and T. Yoneda. 2009. Adsorption and co-precipitation behavior of arsenate, chromate, selenate and boric acid with synthetic allophane-like materials. J. Hazard. Mater. 170:79–86.

Parfitt, R.L. 1990. Allophane in New Zealand—A review. Aust. J. Soil Res. 28:343–360.

Parfitt, R.L. 2009. Allophane and imogolite: Role in soil biogeochemical processes. Clay Miner. 44:135–155.

Parfitt, R.L., A.R. Fraser, and V.C. Farmer. 1977. Adsorption on hydrous oxides 3. Fulvic-acid and humic-acid on goethite, gibbsite and imogolite. J. Soil Sci. 28:289–296.

Parfitt, R.L., R.J. Furkert, and T. Henmi. 1980. Identification and structure of two types of allophane from volcanic ash soils and tephra. Clays Clay Miner. 28:328–334.

Parfitt, R.L., and A.D. Wilson. 1985. Estimation of allophane and halloysite in three sequences of volcanic soils, New Zealand, p. 1–8. *In* E.F. Caldas and D.H. Yaalon (eds.) Volcanic soils: Weathering and landscape relationships of soils on Tephra and Basalt. Catena Suppl. 7. Catena-Verlag, Cremlin-Destedt, Germany.

Percival, H. 1991. Soil water chemistry and mineral stability in soils. A physical chemistry approach. Chem. N.Z. 55: 66–70, 74.

Schneider, M.P.W., T. Scheel, R. Mikutta, P. van Hees, K. Kaiser, and K. Kalbitz. 2010. Sorptive stabilization of organic matter by amorphous Al hydroxide. Geochim. Cosmochim. Acta 74:1606–1619.

Singleton, P.L., M. McLeod, and H.J. Percival. 1989. Allophane and halloysite content and soil solution silicon in soils from rhyolitic volcanic material. N.Z. Aust. J. Soil Res. 27:67–77.

Sposito, G. 2004. The surface chemistry of natural particles. Oxford University Press, New York.

Steefel, C.I., P. van Capellen, K.L. Nagy, and A.C. Lasaga. 1990. Modeling water-rock interaction in the surficial environment: The role of precursors, nucleation, and Ostwald ripening, p. 322–325. *In* Geochemistry of the earth's surface and of mineral formation. 2nd Int. Symp. Aix-en-Provence, France.

Su, C.M., and J.B. Harsh. 1993. The electrophoretic mobility of imogolite and allophane in the presence of inorganic anions and citrate. Clays Clay Miner. 41:461–471.

Su, C.M., and J.B. Harsh. 1994. Gibbs free energies of formation at 298 K for imogolite and gibbsite from solubility measurements. Geochim. Cosmochim. Acta 58:1667–1677.

Su, C.M., and J.B. Harsh. 1996. Alteration of imogolite, allophane, and acidic soil clays by chemical extractants. Soil Sci. Soc. Am. J. 60:77–85.

Su, C.M., and J.B. Harsh. 1998. Dissolution of allophane as a thermodynamically unstable solid in the presence of boehmite at elevated temperatures and equilibrium vapor pressures. Soil Sci. 163:299–312.

Su, C.M., J.B. Harsh, and P.M. Bertsch. 1992. Sodium and chloride sorption by imogolite and allophanes. Clays Clay Miner. 40:280–286.

Su, C.M., and D.L. Suarez. 1997. Boron sorption and release by allophane. Soil Sci. Soc. Am. J. 61:69–77.

Takahashi, T., R. Dahlgren, and P. Vansusteren. 1993. Clay mineralogy and chemistry of soils formed in volcanic materials in xeric moisture regime of northern California. Geoderma 59:131–150.

Taylor, N.H. 1933. Soil processes in volcanic ash-beds. N.Z. J. Sci. Technol. 14:338–352.

Tsuchida, H., S. Ooi, K. Nakaishi, and Y. Adachi. 2005. Effects of pH and ionic strength on electrokinetic properties of imogolite. Colloids Surf., A 265:131–134.

Ugolini, F.C., and R.A. Dahlgren. 1991. Weathering environments and occurrence of imogolite allophane in selected Andisols and spodosols. Soil Sci. Soc. Am. J. 55:1166–1171.

Veith, J.A., and G. Sposito. 1977. Reactions of aluminosilicates, aluminum hydrous oxides, and aluminum oxide with o-phosphate: The formation of x-ray amorphous analogs of variscite and montebrasite. Soil Sci. Soc. Am. J. 41:870–876.

Vicente, J.G., and E. Besoain. 1961. Mineralogy of clays from some volcanic soils of Chile. An. Edafol. Agrobiol. 20:497–550.

Wada, S. 1987. Imogolite synthesis at 25-degrees-C. Clays Clay Miner. 35:379–384.

Wada, K. 1995. Role of aluminum and iron in the accumulation of organic matter in soils with variable charge, p. 47–58. *In* P.M. Huang et al. (eds.) Environmental impact of soil component interactions. CRC Press, Boca Raton, FL.

Wada, K., and Y. Kakuto. 1985. Embryonic halloysites in Ecuadorian soils derived from volcanic ash. Soil Sci. Soc. Am. J. 49:1309–1318.

Wada, S., and Y. Kakuto. 1999. Solubility and standard gibbs free energy of formation of natural imogolite at 25degrees C and 1 atm. Soil Sci. Plant Nutr. 45:947–953.

Wada, K., and I. Matsubara. 1968. Differential formation of allophane, imogolite, and gibbsite in the Kitakami pumice bed. Trans. 9th Int. Congr. Soil Sci. 3:123–131.

Wada, S.I., and K. Wada. 1981. Reactions between aluminate ions and orthosilicic acid in dilute, alkaline to neutral solutions. Soil Sci. 132:267–273.

Wada, K., and N. Yoshinaga. 1969. Structure of imogolite. Am. Mineral. 54:50–71.

Wan, G., Peraval, H.J., Theng, B.K.G., and Parfitt, R.L. (2002). Sorption of copper and cadmium by allophane–humic complenes. Dev. Soil Sci. 28A:37–47.

Wells, N., C.W. Childs, and C.J. Downes. 1977. Silica Springs, Tongariro National Park, New Zealand—Analyses of the spring water and characterization of the alumino-silicate deposit. Geochim. Cosmochim. Acta 41:1497–1506.

Wells, N., and R.L. Parfitt. 1987. Occurrences of short-range order clays and their use in pollution control. Proc. Pacific Rim Congress 87:469–473.

Wilson, M.J. 1987. A handbook of determinative methods in clay mineralogy. Blackie and Son, Ltd., Glasgow, U.K.

Wilson, M.A., S.A. McCarthy, and P.M. Fredericks. 1986. Structure of poorly-ordered aluminosilicates. Clay Miner. 21:879–897.

Yang, H., C. Wang, and Z. Su. 2008. Growth mechanism of synthetic imogolite nanotubes. Chem. Mater. 20:4484–4488.

Yoshinaga, N. 1968. Identification of imogolite in the filmy gel materials in the Imaichi and Shichihonzakura pumice beds. Soil Sci. Plant Nutr. (Tokyo) 14:238–246.

Yoshinaga, N., and S. Aomine. 1962. Imogolite in some Ando soils. Soil Sci. Plant Nutr. (Tokyo) 8:114–121.

Yoshinaga, N., M. Nakai, and M. Yamaguchi. 1973. Unusual accumulation of gibbsite and halloysite in the Kitakami pumice bed, with a note on their genesis. Clay Sci. 4: 155–165.

Yuan, G., Percival, H.J. Theng, B.K.G., and Parfitt, R.L. (2002). Sorption of Copper and Cadmium by allophane-humic complexes. Dev. Soil Sci. 28A:37–47.

Zhao, M., Y. Xia, and L. Mei. 2009. Energetic minimum structures of imogolite nanotubes: A first-principles prediction. J. Phys. Chem. C 113:14834–14837.

IV

Soil Biology and Biochemistry: Soil Biology in Its Second Golden Age

E.A. Paul
Colorado State University

Paolo Nannipieri
University of Florence

SOIL BIOLOGY, THE STUDY OF soil organisms and their processes, discusses the most abundant living organism on earth. The early characterization of visible fungi, such as that in the *Theatrum Fungorium*, published in 1675 and Leeuwenhoek's microscopic investigation of bacteria during the same century, represented the initial creation of this field from an observational viewpoint. The study of organisms and soil processes started with Darwin's explanation of the role of earthworms in soil formation. The studies in the mid-nineteenth century of nitrogen transformations by Warington, Lawes, and Gilbert (Paul, 2007) represented the beginnings of scientific soil process studies. However, the true recognition of this field occurred during the period of 1890–1910 when organisms carrying out the processes necessary for decomposition and nitrogen and sulfur cycles were described by Beijerinck and Winogradsky (Waksman, 1932). That period has been named the "golden age" of soil microbiology (Waksman, 1932). The advent of tracers in the mid-twentieth century allowed the measurement of the dynamics of the processes in situ. Molecular techniques that were adapted to soils came about 100 years after the first golden age with the initial paper of Torsvik (1980). Great advances in molecular techniques, for both the study of biological nuclear materials and for modern studies of organic matter structure, have occurred during the last 20 years. The soil biology chapters in this book describe these advances in the classification of the soil organisms and the measurement of their processes. The progress and advancement of our knowledge has been so profound that we can depict this section as the "second golden age" of soil biology.

The soil bacteria and archaea, ranging in size from 0.2 to a few microns in diameter, have populations between 10^9 and 10^{10} organisms per gram of soil. The smallest soil biota include the submicroscopic viruses and prions that require living cells for reproduction. They exist in soil in numbers equivalent to the bacteria. The larger, filamentous fungi often have a similar biomass to that of the bacteria, but in examples such as the *Armillaria*, associated with multiple tree roots they can exist as a single organism with an area of several hectares (Thorn and Lynch, 2007). The recently discovered archaea, with their special cell structures, in addition to the predatory protozoa, comprise the remainder of the soil microorganisms that carry out the majority of nutrient transformations in soil. The archaea and bacteria are characterized by a lack of a nucleus and other structures, such as mitochondria, yet these are the only organisms that contain the complex enzyme systems for N_2 fixation, nitrification, methane oxidation, and reduction, and the inorganic sulfur and metal transformations (Killham and Prosser, 2007). Although these are the earth's oldest living inhabitants, their small size and lack of morphological diversity had made their study difficult until the advent of the soil molecular studies (Torsvik, 1980). While less in number and biomass, the soil animals, or fauna, play significant roles in soil mixing, or bioengineering, and the consumption of microorganisms (Coleman and Crossley, 1996; Lavelle et al., 2004). The total soil biota comprise 2%–5% of the soil organic carbon content and 4%–8% of its

nitrogen (Bardgett et al., 2005). Because of the close interaction of the living cells with external enzymes and the soil mineral matrix, soil has been often described as a self-controlled living entity with many similar characteristics over many parts of the world (McGill, 2007). The factors controlling biochemical reactions and those affecting population and soil organic matter turnover are similar. Thus, while individual soils show the interaction of these factors with their environment (Morris and Blackwood, 2007), many advances will be made by studying the unifying principles (Fierer et al., 2009) that control soil biology on a global basis.

Soil biochemistry, a second component of soil biology, refers to both the physiology and chemistry within the living cells that catalyze life's functions, as well as the activity of external enzymes. Important biological molecules (i.e., nucleic acids and proteins, including extracellular enzymes) can be adsorbed by surface-reactive particles (i.e., clays), thus becoming protected against the degradation reactions of proteases and nucleases without losing their activity. The capacity of the immobilized enzyme to catalyze the target reactions, and the availability of the extracellular adsorbed DNA to be taken up by competent bacterial cells to be incorporated into the bacterial genome, is strongly affected by the soil matrix consisting of silts, clays, and sesquioxides (Stotzky, 1986; Nielsen et al., 2006; Pietramellara et al., 2009). This section of the handbook contains chapters on soil biota, the status and origin of soil enzymes, and the meaning of current enzyme assays. Organisms that inhabit soil carry out the nutrient transformations that make life possible (Coleman and Crossley, 1996). As decomposers of the products of photosynthesis and thus plant growth, the soil biota complete the carbon cycle. The current global change problems result from a disturbance of a very old steady state between photosynthesis and decomposition. Decomposition proceeds at a slightly smaller rate than photosynthesis (Stevenson, 1994) resulting in a buildup of the soil organic matter, which is so essential to sustainable agriculture, ecosystem functioning, and carbon sequestration in global change. Soil at $15–20 \times 10^{15}$ g C has twice as much C as the atmosphere. The recent utilization of photosynthesized C stored in coal and oil deposits over millions of years, the cultivation of soils, and the disturbance of ecosystems, such as occurs in fires, is upsetting this steady state and has caused our current climate change problems by returning, rather than fixing, more CO_2 into the atmosphere (Ojima, 1992). For a long time, it has been hypothesized that recalcitrance of organic substrates in soil was the main driver of soil organic matter survival in the long term. Clear analytical evidence now shows that inherently recalcitrant plant compounds such as *n*-alkanes and lignins turn over faster than bulk soil organic carbon (Marschner et al., 2008).

Nitrogen is as important to life and soil organic matter formation as is carbon. In addition, many of the same organisms that decompose carbon also mineralize nitrogen.

Nitrogen also undergoes numerous inorganic reactions, such as nitrification–denitrification and N_2 fixation, which are primarily carried out by bacteria and archaea that complete the complex nitrogen cycle and make plant growth possible

(Norton, 2008). Where N is present in excess quantities, these reactions can also cause water pollution and the release of greenhouse gases that are much more potent than carbon dioxide.

The great diversity of soil biota, in addition to the importance of their metabolism, their large numbers, and the possible significance of still-to-be-identified soil organisms, increase the value of the study of soil biology from a pure science–natural history standpoint. This diversity also contributes to ecological stability (Wall, 2004). Due to the wide variety in both number and type of decomposing organisms, the loss of one or two species may not affect this process. However, specialized processes, such as nitrification carried out by a few specialized bacteria, can be affected and would only be more sensitive to disturbance. The vast competitiveness and number of organisms assure the breakdown of most organics, whether natural or man-made, and also serve in self-cleansing the soil of potential disease-carrying organisms. Even so, soil is not infallible. Overloading with man-made chemicals and toxic organisms can lead to a breakdown of this self-cleansing characteristic.

The field of soil biology has both greatly benefited from and contributed to present breakthroughs in molecular biology. These techniques are now available to a wide range of ecologists, microbiologists, and soil scientists, and allow for the characterization of previously unknown organisms, such as the participation of archaea in nitrification (Nicol and Schleper, 2006). Process-specific probes allow for the determination of genes encoding for specific processes, such as those described in this book for the uptake of phosphorus by mycorrhiza. The study of proteomics allows for the measurement of the compounds actually involved in the reactions and will tie new studies with the rich classical literature on soil enzymes (Nannipieri and Paul, 2009). It is now possible to study microbial growth by tying stable-isotope probing or metagenomic measurements (which include the molecular structure) to more classical microbial growth and incubation studies, providing access to a great amount of information. Many factors affect gene expression in soil and the study of DNA by itself is not sufficient. Examples of the study of the presence and expression of genes, in addition to the characteristics of the expressed proteins, are now available (Nannipieri and Paul, 2009). While many reactions in soil have been characterized, many new reactions and organisms are being discovered; for example, the organisms that are involved in anaerobic metabolism of methane and sulfur (Section 26.4). The Anammox bacteria that oxidize ammonia to N_2 under anaerobic conditions in the presence of nitrite–nitrate are but one example. They have been found in biofilms of oxygen-limited reactors (Jetten, 2001), and it is reasonable to hypothesize their presence in soil under anaerobic conditions (Nannipieri and Paul, 2009).

It is the intent of the author's of this book to integrate new breakthroughs with the vast amount of information from classical literature such that a true foundation for even greater advances in knowledge is established. It provides numerous examples of the study of the presence and expression of genes, as well as the characteristics of expressed proteins. The description of viruses, based on molecular techniques, present an integrated analysis of their numbers and classification. Molecular methods for the study of archaea, bacteria, and fungi have richly contributed to our new knowledge of the vast array of organisms in soil. At least 90% of the bacteria and archaea active in soil are yet to be identified, and the fungi and nematodes are being reclassified. Process probes have revolutionized our understanding of the functioning of the important mycorrhizal organisms and their associations with plants. The development of new techniques, such as those for the rapid throughput of molecular information in pyrosequencing and the combination of tracers and molecular techniques as used in stable-isotope probing indicates that more advances are forthcoming. Soil biology can only progress based from an integrated knowledge of all organisms involved and an increased understanding of the methods of analysis, their strengths, and limitations (Andren et al., 2008). In spite of the diversity in both soil processes and soil organisms, there is now also a recognition of the great unity in the biotic and abiotic controls that affect both the organisms and their processes. This section has brought together both the recent knowledge and the fundamental unifying principles in a readable format with good reference to both classical and modern literature. It should comprise both an excellent introduction to the soil biota, as well as a good indication of potential progress in the future.

References

Andren, O., H. Kirchmann, T. Katterer, J. Magid, E.A. Paul, and D.C. Coleman. 2008. Visions of a more precise soil biology. Eur. J. Soil Sci. 59:380–390.

Bardgett, R.D., M.B. Usher, and D.W. Hopkins. 2005. Biological diversity and function in soils. Cambridge University Press, Cambridge, U.K.

Coleman, D.C., and D.A. Crossley Jr. 1996. Fundamentals of soil ecology. Academic Press, London, U.K.

Fierer, N., S. Grandy, J. Six, and E.A. Paul. 2009. Searching for unifying principles in soil ecology. Soil Biol. Biochem. 41:2249–2256.

Jetten, M.S.M. 2001. New pathway for ammonia conversion in soil and aquatic systems. Plant Soil 230:9–19.

Killham, K., and J. Prosser. 2007. Soil prokaryotes, p. 119–144. *In* E.A. Paul (ed.) Soil microbiology, ecology and biochemistry. Academic Press, San Diego, CA.

Lavelle, P., D.E. Bignell, M.C. Austen, V.K. Brown, V. Behan-Pelletier, J.R. Garey, P.S. Giller et al. 2004. Connecting soil and sediment biodiversity: The role of scale and implications for management, p. 193–224. *In* D.H. Wall (ed.) Sustaining biodiversity and ecosystem services in soils and sediments. Island Press, Washington, DC.

Marschner, B., S. Brodowski, A. Dreves, G. Gleixner, A. Gude, P.M. Grootes, U. Hamer et al. 2008. How relevant is recalcitrance for the stabilization of organic matter in soils? J. Plant Nutr. Soil Sci. 171:91–110.

McGill, W.B, 2007. The physiology and biochemistry of soil organisms, p. 211–256. *In*. E.A. Paul (ed.) Soil microbiology, ecology and biochemistry. Academic Press, San Diego, CA.

Morris, S.J., and C. Blackwood. 2007. The ecology of soil organisms, p. 195–230. *In*. E.A. Paul (ed.) Soil microbiology, ecology and biochemistry. Academic Press, San Diego, CA.

Nannipieri, P., and E.A. Paul. 2009. The chemical and functional characterization of soil organic N and its biotic components. Soil Biol. Biochem. 41:2357–2369.

Nicol, G.W., and C. Schleper. 2006. Ammonia-oxidising Crenarchaeota: Important players in the nitrogen cycle? Trends Microbiol. 14:207–212.

Nielsen, K.M., L. Calamai, and G. Pietramellara. 2006. Stabilization of extracellular DNA by transient binding to various soil surfaces, p. 141–158. *In* P. Nannipieri and K. Smalla (eds.) Nucleic acids and proteins in soil. Springer, Heidelberg, Germany.

Norton, J.M. 2008. Nitrification in agricultural soils, p. 173–199. *In* J.S. Schepers, W.B. Raun, R.F. Follett, R.H. Fox, and G.W. Randall (eds.) Nitrogen in agricultural systems. Agronomy Monograph 49. ASA, Madison, WI.

Ojima, D. 1992. Modelling the earth system. UCAR, Boulder, CO.

Paul, E.A. 2007. Soil microbiology, ecology and biochemistry in perspective, p. 3–24. *In*. E.A. Paul (ed.) Soil microbiology, ecology and biochemistry. Academic Press, San Diego, CA.

Pietramellara, G., J. Ascher, F. Borgogni, M.T. Ceccherini, G. Guerri, and P. Nannipieri. 2009. Extracellular DNA in soil and sediment: Fate and ecological relevance. Biol. Fertil. Soils 45:219–235.

Stevenson, F.J. 1994. Humus chemistry: Genesis, composition, reactions. 2nd Ed. John Wiley & Sons, New York.

Stotzky, G. 1986. Influence of soil mineral colloids on metabolic process, growth, adhesion, and ecology of microbes and viruses, p. 305–428. *In* P.M. Huang and M. Stotzky (eds.) Interactions of soil minerals with natural organics and microbes. SSSA, Madison, WI.

Thorn, R.G., and D.J. Lynch. 2007. Fungi and eukaryotic algae, p. 145–162. *In*. E.A. Paul (ed.) Soil microbiology, ecology and biochemistry. 3rd Ed. Academic Press, San Diego, CA.

Torsvik, V.I. 1980. Isolation of bacterial DNA from soil. Soil Biol. Biochem. 12:15–21.

Waksman, S.A. 1932. The principles of soil microbiology. Williams & Wilkins, Baltimore, MD.

Wall, D.H. 2004. Sustaining biodiversity and ecosystem services in soils and sediments. Island Press, Washington, DC.

24

Microbiota

Raffaella Balestrini
Università di Torino

Valeria Bianciotto
Università di Torino

Paola Bonfante
Università of Torino

Michael Schloter
Institute of Soil Ecology

Sharath Srinivasiah
University of Delaware

R. Greg Thorn
University of Western Ontario

Kurt E. Williamson
The College of William and Mary

K. Eric Wommack
University of Delaware

24.1 Viruses in Soil Ecosystems

Kurt E. Williamson

Sharath Srinivasiah

K. Eric Wommack

24.1.1 Introduction

Viruses are obligate genetic parasites that require the replication machinery of a cellular host in order to complete their own reproductive cycles. Outside of host cells, viruses exist as particles consisting of a nucleic acid genome surrounded by a protein coat, known as a *capsid*. Depending upon the specific virus, the genome may comprise DNA or RNA, which in turn may be double or single stranded. Viral capsids may be icosahedral, spherical, pleomorphic, filamentous, or rod shaped. Virus structure may also include additional features such as lipid envelopes, apical spikes, and tails (Figure 24.1). Soils provide an incredible range of niches housing a diverse array of organisms, leading to a similarly diverse representation of viruses that parasitize these organisms. Soils have been shown to contain plant viruses (Fillhart et al., 1998; Delogu et al., 2003), insect viruses (Fuxa, 2004; Christian et al., 2006), fungal viruses (Melzer and Bidochka, 1998), animal viruses (Duboise et al., 1979; Santamaria and Toranzos, 2003; Pourcher et al., 2007), and bacterial viruses (also known as bacteriophages or simply "phages") (Yin et al., 1997; Ashelford et al., 1999a, 2000, 2003; Keel et al., 2002; Williamson et al., 2003, 2005, 2007).

Historically, the study of viruses in soil has focused on economically significant pathogens of plants, animals, and insects. A large body of research is available concerning the fate and transport of land-applied human viral pathogens through wastewater or biosolids (Duboise et al., 1979; Santamaria and Toranzos, 2003). However, more recent data indicate that the majority of naturally occurring soil viruses are tailed phages with dsDNA genomes (Williamson et al., 2005). This is consistent with the overwhelming abundance of bacteria relative to other potential host organisms in soil ecosystems.
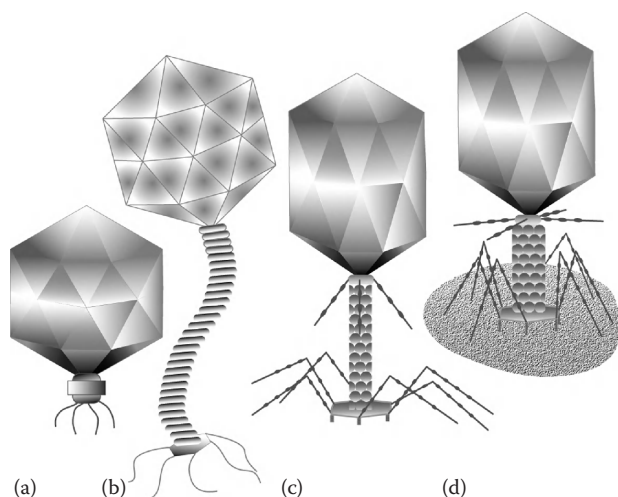
FIGURE 24.1 Common morphologies of soil viruses. (a) *Podoviridae*, icosahedral capsid symmetry with short, noncontractile tail; (b) *Siphoviridae*, icosahedral capsid symmetry with long flexible, non-contractile tail; (c) *Myoviridae*, prolate icosahedral capsid symmetry with long, contractile tail; (d) rendering of *Myoviridae* particle attached to host cell, tail contracted. Tail structures are used in host recognition, attachment, and infection.

24.1.2 Virus Life Cycle

In order to appreciate the influences that viruses can exert within the soil ecosystem, it is important to understand the modes through which viruses replicate. Most viruses replicate through one of two major life cycles: acute or persistent. Our current understanding of viral replication is based largely upon studies of bacteriophages; thus, while the terms "phage" and "virus" are often used interchangeably, it is important to remember this distinction.

The acute pathway, also known as lytic replication in phages, is carried out by *virulent* viruses. Acute viral infections are characterized by rapid production of intracellular virus particles followed by host cell lysis and release of progeny virions (Figure 24.2). Acute infections tend to quickly eliminate local populations of susceptible hosts, and extracellular virus particles must survive long enough to propagate infection if the virus strain is to remain successful. By contrast, persistent infections, also known as *lysogenic* replication among phages (Figure 24.2), are carried out by *temperate* viruses. Persistent infections typically involve the silencing of genes involved in lytic functions by a phage-encoded repressor protein. This cryptic state, termed *lysogeny*, can be maintained indefinitely. During cell division, the phage genome or *prophage* is regularly assorted to daughter cells either as an integrated part of the host chromosome (Ptashne, 1991) or as an extrachromosomal element (Bertani, 2004). In this fashion, the phage is maintained within the host population (Figure 24.2).

While lysogeny is highly stable, this relationship is not terminal. Environmental signals, particularly damage to host DNA, can initiate a cascade of cell responses that inactivate
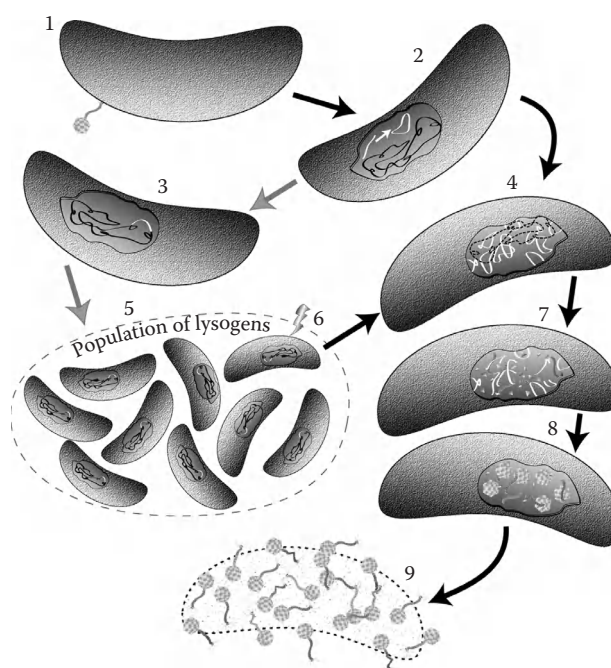
FIGURE 24.2 Virus replication: temperate versus lytic. (1) All viral replication begins with virus attachment to host cell and the injection of viral nucleic acid into the host cytoplasm. (2) Viral nucleic acids (in white) are processed to allow replication. For many dsDNA viruses, this step involves circularization of a linear genome. Based on signals received from the host cell, the virus may proceed through one of two replication pathways. *Lytic replication*: (4) Early viral genes are expressed. The host chromosome is restricted (dashed black lines) and typically used as building blocks for the replication of new viral genomes (white circles). (7) Late viral genes are expressed: viral structural proteins are synthesized and self-assemble into multisubunit structures such as the procapsid, tail, and baseplate. (8) Viral genomes are packaged into capsids and component subunits (capsids and tails) self-assemble into mature virions. (9) Viral-mediated host cell lysis releases newly produced viruses into the environment. *Lysogenic replication*. (3) Viral repressor protein prevents the expression of early viral genes by blocking transcription. Viral DNA (white) is integrated into the host chromosome (black) through a recombination process mediated by both viral and host proteins. (5) Stable integration enables transmission of the viral genome (prophage) to subsequent generations of daughter cells, leading to a population of lysogens—host cells that carry prophage. (6) Induction. Damage to host DNA (indicated by lightning bolt) and other environmental stimuli may cease prophage repression and the virus shifts into lytic replication mode as early viral genes are expressed (4).

the phage repressor protein. Following this event, known as *induction*, the virus shifts into lytic replication and progeny virus particles are released into the environment through cell lysis. While little is known about the factors leading to the establishment of lysogeny in nature, it is generally believed that lysogeny provides a refuge for temperate phages when host abundance is low or conditions are poor for robust viral replication (i.e., cells are starved) (Stewart and Levin, 1984; Marsh and Wellington, 1994).

24.1.3 Impact of the Soil Environment on Virus Replication Mode

The spatial heterogeneity and temporal variability of soils can be extremely high and may select for the predominance of specific viral replication strategies. In particular, lysogenic replication appears to hold a distinct advantage for viruses in soils, as lysogeny supports the long-term maintenance of phages within host populations (Stewart and Levin, 1984; Pantastico-Caldas et al., 1992; Marsh and Wellington, 1994). Recent investigations into the frequency of inducible lysogeny among soil bacteria isolated on laboratory media (Williamson et al., 2008) and within natural assemblages of soil bacteria (Williamson et al., 2007) indicate that inducible lysogens (i.e., bacteria carrying a prophage element) account for approximately 30%, and in some cases over 60%, of the total bacterial population in a given soil. While this is a significant fraction and suggests that many soil phages are temperate, soils also contain appreciable numbers of lytic viruses, many of which have been isolated and characterized (described later). Thus, the importance of lytic viruses to the ecology of soils should not be underestimated.

24.1.4 Detection and Enumeration of Viruses in Soils

Well-characterized viral pathogens are frequently detected in soils using PCR assays, in which specific oligonucleotide primers are designed to amplify the viral DNA (or cDNA generated from viral RNA) from environmental samples. This approach has been used to detect specific plant (Farzadfar et al., 2007; Budge et al., 2008) and animal viruses (Chancellor et al., 2006) in soils. Soil viruses may be detected and enumerated through additional direct or indirect approaches. The most common indirect method for enumerating soil viruses is the plaque assay, which pertains primarily to bacteriophages.

24.1.4.1 Plaque Assay

In the plaque assay, a soil suspension is filtered to remove bacteria and other organisms. The filtrate is then diluted and mixed with an actively growing culture of a specific bacterial host strain and molten agar. The mixture is spread over solid agar media and the molten top agar is allowed to solidify. The plate is then incubated to allow for bacterial growth. If the filtrate contained any infectious phages, this will be apparent as *plaques* or zones of clearing in the bacterial lawn (Figure 24.3a). The number of plaques reflects the number of infectious virus particles present in the soil at the time of sampling. Similar infective assays are used to detect and enumerate animal or plant viruses except that tissue cultures, or in some cases, whole organisms, are used rather than bacteria grown on agar plates. A major shortcoming of infective assays is that detection and enumeration are limited only to specific viruses capable of propagating within the target organism(s), under the specific conditions tested (temperature, humidity, etc.). Thus, infective assays provide no information about total viral abundance in soils.

24.1.4.2 Viral Direct Counts

Total virus abundance in soils may be assessed through transmission electron microscopy (TEM) or epifluorescence microscopy (EFM), each of which enables the direct counting of virus particles. Both of these direct counting procedures require efficient dispersion of soil aggregates and extraction of virus particles from the soil matrix into an aqueous carrier phase. Physical dispersion has been accomplished through bead-beating (Ashelford et al., 2003), mechanical blending (Swanson et al., 2009), or sonication (Williamson et al., 2003, 2005, 2007) of soils suspended in buffer, such as 1% potassium citrate. The suspension is typically filtered to remove bacteria prior to further processing. For EFM enumeration, virus suspensions are passed though 0.02 μm pore filters and the captured virus particles are stained with a fluorochrome that binds to nucleic acids (Williamson et al., 2003).

(a)　　　　　　　　(b)　　　　　　　　(c)
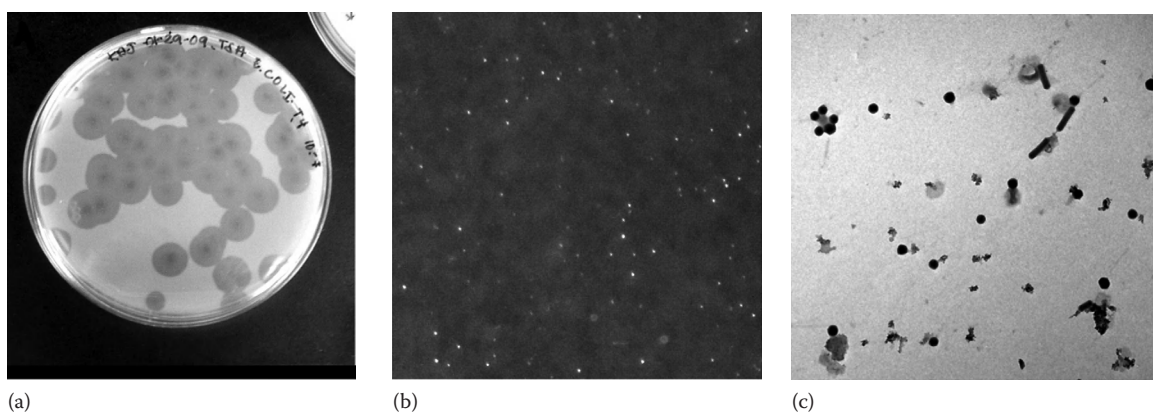
FIGURE 24.3　Methods for detecting and enumerating soil viruses. (a) Plaque assay, the number of infectious virus particles is estimated from the number of infectious centers or plaques apparent on plates of host. (b) Epifluorescence microscopy, digital images of fluorescently labeled virus particles are captured using a CCD camera and typically counted using image analysis software. (c) Transmission electron microscopy, images of virus particles are captured digitally or on film, and counted manually.

Filters are viewed under an appropriate excitation wavelength for the fluorochrome stain and the emitted fluorescence signals from the stained virus genomes are captured digitally using a charge-coupled device (CCD) camera (Figure 24.3b). For TEM enumeration, additional purification such as density gradient centrifugation is necessary prior to loading samples onto EM grids, so as to clearly visualize virus particles. Grids are then stained with an electron-dense heavy metal salt, such as uranyl acetate or phosphotungstic acid, and virus particles are visualized in a transmission electron microscope (Figure 24.3c).

While TEM allows for visualization of viral morphology, electron microscopy entails higher operating costs, increased sample processing time, and poorer capture efficiency leading to lower abundance estimates. EFM is cheaper and allows high sample throughput with much higher capture efficiency but does not allow for visualization of fine structure. Furthermore, EFM may detect false-positive fluorescence signals from nonviral sources; thus, care must be taken in processing images to obtain viral direct count data.

The first TEM-based estimates of viral abundance in an agricultural soil from the United Kingdom placed soil viral abundance on the order $10^8 \, g^{-1}$ (Ashelford et al., 2003). Later EFM-based estimates found that viral abundance varied from $8 \times 10^8 \, g$ dry weight^{-1} (g dw^{-1}) in agricultural soils to $4 \times 10^9 \, g$ dw^{-1} in wetland soils of the temperate Mid-Atlantic region of the United States (Williamson et al., 2005). Attempts at direct counting of viruses in soils from hot deserts such as the Sahara and Namib indicate that extracellular virus particles are below the limits of detection in these ecosystems (Prigent et al., 2005; Prestel et al., 2008). However, soils from the cold desert ecosystem of the McMurdo Dry Valleys, Antarctica, were estimated to contain between 9×10^7 and 8×10^8 viruses g dw^{-1} (Williamson et al., 2007). These patterns of abundance broadly indicate the physical environmental variables that control viral abundance within a given soil ecotype.

24.1.5 Factors Controlling Viral Abundance and Persistence

In temperate soils, viral abundance is significantly correlated to bacterial abundance (Williamson et al., 2007). This supports the generalization that most viruses in these soils are phages and indicates that the presence and abundance of susceptible hosts are a key factor controlling viral abundance. When host organisms are present, particularly in high numbers, new viruses may be continually produced and released into the soil matrix through acute infections. In the absence of susceptible hosts, the abundance of extracellular viruses is controlled by physicochemical properties of the soil environment. Adsorption to soil surfaces, which is presumably the natural state of soil viruses, has been shown to enhance the persistence of infectious virus particles (Lipson and Stotzky, 1986). Thus, the number and types of colloidal surfaces affect total viral abundance in a given soil (Meschke and Sobsey, 2003; Zhuang and Jin, 2003a). Soils with higher clay and organic matter contents generally provide larger surface area for adsorption of viral particles and harbor higher viral abundances.

Soil moisture content is a key factor controlling viral abundance and persistence in soils, as thicker water films have smaller interfacial areas (air–water interface, solid–water interface, and triple-phase boundary), and surface tension created at interfaces tends to inactivate viruses (Gerba, 1984; Jin and Flury, 2002; Zhuang and Jin, 2003b; Torkzaban et al., 2006). The pH of the soil solution (Loveland et al., 1996) is another important abiotic factor controlling viral abundance, with higher abundances generally occurring in neutral pH soils. Lower pH generally favors stronger adsorption of viruses to soil surfaces (Gerba, 1984); thus, it is unclear whether higher viral abundances in neutral soils is due to reduced persistence of intact virions in acid soils or more efficient extraction of virions from neutral soils. Temperature significantly influences viral persistence and abundance in soils (Duboise et al., 1979; Wen et al., 2004). Generally, lower temperatures foster the persistence or preservation of viruses while higher temperatures promote thermal decay of viral particles.

24.1.6 The Virus-to-Bacterium Ratio

The ratio of viral abundance to bacterial abundance (virus-to-bacterium ratio, or VBR) has been used as a metric for determining relationships between viral production/decay and bacterial production/infection in aquatic ecosystems (Wommack and Colwell, 2000). Two significant trends have emerged regarding VBR in all aquatic ecosystems examined to date, despite large disparities in temperature, salinity, and trophic status of these ecosystems: (1) viral abundance is significantly correlated to bacterial abundance, and (2) the range of VBRs is surprisingly narrow, typically varying from 1 to 50 with a mean value of about 10 (Wommack and Colwell, 2000; Weinbauer, 2004). Significant correlations between viral and bacterial abundance suggest a close coupling between viral and bacterial production in aquatic environments, while the narrow range of VBRs may indicate balanced rates of viral production and decay.

In contrast to the relatively narrow range of VBRs in aquatic systems, including marine and freshwater sediments, VBRs of soils encompass a range of several orders of magnitude (Table 24.1). The low VBRs of wetland soils (Williamson et al., 2007) suggest that in perennially wet, temperate soils, viral and bacterial production are tightly linked and viral production and decay are closely balanced. The relatively higher VBRs of agricultural soils are due largely to the fact that bacterial abundances in agricultural soils were lower than those of wetland soils by 100-fold or more (Williamson et al., 2007). Regular tillage and disturbance of agricultural soils is known to negatively impact bacterial abundance (Feng et al., 2003; Helgason et al., 2009) and likely accounts for the decrease in bacterial abundance and concomitant increase in VBR. The higher VBRs of agricultural soils may also indicate a disconnect between viral production and decay, where viruses are produced faster than they are destroyed or removed from the system.

The remarkably high VBRs recorded for Antarctic soils (Williamson et al., 2007) are most likely due to long-term storage

TABLE 24.1 Comparison of Virus-to-Bacterium Ratios across Ecosystems

Ecosystem	Reference	Method	VBR
Aquatic (various)	Wommack and Colwell (2000)	EFM/TEM	0.5–50
Sediment (various)	Helton et al. (2006)	EFM/TEM	0.03–107
Soil (wetland)	Williamson et al. (2007)	EFM	10–60
Soil (agricultural)	Williamson et al. (2007)	EFM	330–470
Soil (Antarctic)	Williamson et al. (2007)	EFM	170–8100
Soil (Sahara)	Prigent et al. (2005)	TEM	$<10^{-3}$

TEM, transmission electron microscopic direct counts; EFM, epifluorescence microscopic direct counts.

of viruses. A multiplicity of growth-limiting conditions, including low temperature, lack of free water, and scarcity of organic substrates for growth, explains the low bacterial abundances in these soils. It is unlikely that the sparse bacterial population is capable of supporting such a burgeoning population of viruses. Since viruses are known to persist longer when adsorbed to surfaces (Lipson and Stotzky, 1986; Vettori et al., 1999) and kept at low temperatures (Duboise et al., 1979; Wen et al., 2004), it is likely that viral abundances obtained for Antarctic soils largely comprise stored viruses.

The very low VBRs for Saharan sands indicate that extracellular viruses are relatively rare in these environments. Indeed, the abundance of extracellular virus particles was below the limits of detection by TEM, whereas bacterial abundance in these sands was about $10^4\,g^{-1}$ (Prigent et al., 2005). Low bacterial abundance in Saharan sands is conflated with a low growth rate, neither of which could be expected to support a large viral population. While the low bacterial abundance and production rates are also characteristic of Antarctic soils, the environmental conditions leading to this situation in the Sahara are completely different and likely explain the conspicuous lack of extracellular viruses. Saharan sands experience large diurnal fluxes of temperature, severe doses of UV radiation, near-perpetual arid conditions, and aeolian abrasion, all of which likely contribute to the destruction of extracellular virus particles shortly after they are released.

While VBR may indicate linkages between host production, viral production, and viral decay in aquatic ecosystems, the value of VBR for soil ecosystems remains contentious, as critical questions surround what VBR is measuring. What fraction of extracted soil viruses are infectious or active at the time of sampling? What fraction of extracted bacteria is active? Do the extracted viruses have anything to do with the extracted bacteria, whatsoever? And how accurate are the estimates of both bacterial and viral abundance used to compute this ratio? In order to answer these questions and advance our understanding of soil viral ecology, additional experiments are needed, which explore (1) survival and/or decay rates for autochthonous soil viruses within various soil ecosystems, (2) site-specific factors leading to the destruction or removal of viruses from soil, and (3) improvements in the extraction of bacteria from soil and in estimations of bacterial extraction efficiency.

24.1.7 Viral Diversity

24.1.7.1 Morphological Approaches

The earliest characterization of soil viral assemblages relied upon TEM to render a morphological assessment of diversity (Williamson et al., 2005). The use of TEM approaches was necessary owing to the slow and problematic development of molecular genetic approaches to specifically assess viral diversity in soils, particularly when compared to the suite of methods then available to study viral assemblages in aquatic environments. Despite the low resolution of TEM-based estimates, the distribution of virus morphotypes observed in six different soils (including agricultural, forest, and wetland soils) indicated that the viral communities of each soil were distinct (Williamson et al., 2005). Interestingly, phages with elongated capsids were observed in five of the six soils. Phage particles with elongated capsids have also been observed in samples of Saharan sands (Prigent et al., 2005). Apart from these few observations, this morphotype has been rarely reported among environmental samples and comprises fewer than 2% of all known phages (Ackermann, 2001). Thus, assemblages of soil viruses present unique and novel features even at this level of resolution. The predominance of tailed viral morphotypes indicated that bacteriophages were the most abundant viral type within all six soil environments. The mean capsid diameter for soil viruses was about 50 nm, with a range of 20–160 nm (Williamson et al., 2005).

24.1.7.2 Molecular Approaches

24.1.7.2.1 Marker Genes

Viruses possess no single, universally conserved phylogenetic marker analogous to the prokaryotic 16S rRNA gene; this has hampered efforts to obtain a comprehensive view of viral diversity and phylogeny. Nevertheless, investigators have used certain marker genes thought to be highly conserved among specific viral families to assess phylogenetic relationships within these families. For example, PCR assays have shown that gene homologues of T7-like DNA polymerase can be commonly found in a broad range of environments, including soils (Breitbart et al., 2006). However, many commonly used phage marker genes (e.g., g23, g20, and T7-like DNA polymerase) may not be routinely found in all soils (Srinivasiah et al., 2008). This limits the applicability of the present array of marker gene PCR assays for examining viral diversity across soils. By and large, most marker gene surveys of viral genetic diversity have been developed for use in aquatic environments. Thus, gene targets that are more specific to autochthonous soil viruses are needed to fully realize the potential of marker gene approaches for assessment of viral diversity in soils.

24.1.7.2.2 Randomly Amplified Polymorphic DNA-PCR

Randomly amplified polymorphic DNA-PCR (RAPD-PCR) relies upon a single primer (typically 10 nucleotides long) to amplify DNA fragments from a heterogeneous mixture of DNA templates. Subsequent separation of RAPD-PCR amplicons by gel electrophoresis generates banding patterns indicative of the underlying

complexity of the original DNA template mixture. In the case of viral assemblages, the template mixture is viral genomic DNA, and the complexity of the RAPD fingerprint serves as a proxy indication of viral genotype richness within a given sample. The approach is termed "random" as the choice of specific primers need not be based on actual sequence information from the target DNA.

RAPD-PCR has been used to assess the richness of virioplankton assemblages in the marine water column (Winget and Wommack, 2008), marine sediments (Helton and Wommack, 2009), and recent experiments indicate the potential utility of this approach in soils (Srinivasiah et al., 2008). However, further methodological refinements are needed before RAPD-PCR can be routinely applied to surveying the composition and diversity of soil viral assemblages.

24.1.7.2.3 Metagenomics

Metagenomics refers to the sequencing of the collective genomic DNA from a microbial assemblage within a specific environmental sample. Metagenomic analysis is a particularly powerful—and increasingly popular—tool for studying the genetic diversity of viral assemblages within environmental samples because sequence data can be obtained for (theoretically) the entire viral assemblage without dependence upon cultivation techniques, marker genes, or prior sequence knowledge. This approach comprises several steps, beginning with the extraction of virus particles from the sample matrix.

24.1.7.2.3.1 Construction of a Viral Metagenome Library The linker-amplified shotgun library, or LASL, has become the *de facto* standard for cloning and sequencing the typically small amounts of viral DNA obtained from environmental samples (Wommack et al., 2009). Once extracted from the soil, viral capsids are denatured and the released DNA is isolated, purified, and sheared into small fragments of one to a few tens of kilobases in size. Shearing is important as it reduces the probability of cloning complete virus genes whose products would be lethal to the cloning system (e.g., holins, lysins). The fragments of viral DNA are end-repaired and ligated to short double-stranded oligonucleotides called linkers. The linker-adapted viral DNA fragments can then be amplified in a PCR reaction, using the linker sequences as priming sites. Amplification *in vitro* not only increases the amount of fragmented viral DNA available for downstream cloning reactions but also removes modified bases, increasing the success of cloning reactions. After size selection of DNA fragments, typically by gel electrophoresis, the linker-amplified viral DNA fragments are cloned into a specific vector, which is then transformed into *Escherichia coli* hosts via electroporation. Cloning vectors used for viral metagenome projects (e.g., pSMART) typically employ transcription terminators flanking the insertion site as an additional protection against the expression of lethal gene products encoded by the viral insert.

Recent technological improvements in the throughput and cost of DNA sequencing have eliminated the necessity of a fragment cloning step (Margulies et al., 2005). Initially, these approaches typically produced shorter sequence read lengths, which perform poorly in subsequent bioinformatic analyses as compared to more established DNA sequencing technology that requires cloning (Wommack et al., 2008). However, manufacturers of second generation sequencing technology are continually striving to increase sequence read length; thus, it is likely that soon, cloning-independent approaches will become the standard for metagenome analyses.

24.1.7.2.3.2 Analysis of Sequence Data in Viral Metagenomes Once a large collection of insert nucleotide sequences within the plasmid clone library has been obtained, individual fragments are assembled into contiguous stretches (*contigs*) with a computer algorithm, according to user-defined match and sequence overlap parameters. Assembly results can then be described by a metric known as the *contig spectrum* (Angly et al., 2005), which indicates the number of assembled contigs containing one sequence (i.e., singletons), two sequences, three, four, and so on. The experimental contig spectrum is used to test predictions about the rank-abundance distribution of viral genotypes within the original environmental sample. In most cases, the power-law rank abundance distribution has produced the smallest error for experimental contig spectra of viral metagenomes (Angly et al., 2005; Fierer et al., 2007). Open reading frames (i.e., putative genes) are typically identified within the contigs using software packages that recognize specific features within the DNA sequence, such as start and stop codons, and ribosome binding sites (Noguchi et al., 2006). This step is usually followed by manual verification and deeper searches to determine the possible taxonomic origin and putative function based on the identity of the closest known homologue.

In the first published metagenomic analysis of soil viral assemblages, the bacterial, fungal, and viral communities of three soils were compared, representing tallgrass prairie, desert, and tropical rainforest ecotypes (Fierer et al., 2007). According to contig spectra analysis, assemblages of soil viruses were found to be extremely rich both on a local and global scale, with up to 10^8 estimated viral genotypes in one soil and only one overlapping sequence in a total of 4577 sequences. These findings suggest that each soil viral assemblage is a distinct microcosm of unique sequences, sharing little in common with viral assemblages from other soils. This contrasts with metagenomic characterizations of marine virus assemblages, which found between 85% and 95% of the most abundant virus genomes were shared among four discrete sampling locations (Sargasso Sea, the Gulf of Mexico, coastal waters of British Columbia, and the Arctic Ocean) (Angly et al., 2006). While marine viral assemblages may be locally diverse but share much in common on a global scale, the available data regarding soil viral assemblages indicates that soil viruses are much more diverse than their marine counterparts. This is consistent with initial results from RAPD-PCR-based assessments of viral genotype richness in soils (Srinivasiah et al., 2008). In all, a number of different techniques are currently available to assess viral diversity in soils. Each technique encompasses innate strengths and weaknesses in terms of cost, ease of analysis, and limits of resolution (Table 24.2).

TABLE 24.2   Comparison of Available Techniques for Assessing Viral Diversity in Soils

Technique	Morphology	Marker Genes
Strengths	1. Abundance data can be obtained simultaneously	1. Low cost, high throughput
	2. Allows for observation of rare or novel morphotypes	2. Allows rapid screening of many samples
	3. Morphology data may be obtained for (theoretically) all viruses in a given sample	3. Can be combined with sequencing to obtain additional information on genetic diversity
	4. Procedures and analysis can be performed by an individual researcher	4. Procedures and analysis can be performed by an individual researcher
		5. Provides information on hosts and their functional diversity in soils
Weaknesses	1. Low throughput, laborious	1. Data may only be obtained for specific viral groups
	2. Low resolution (ICTV family), allows for comparative analyses	2. Relatively low resolution (presence/absence), allows for comparative analyses
	3. Cost may be an issue	3. Development of primers requires prior sequence knowledge
Technique	RAPD-PCR	Metagenomic Libraries
Strengths	1. Low cost, high throughput	1. Extremely high throughput
	2. No prior sequence knowledge required	2. Extremely high resolution (nucleotide level), allows for robust estimates of genotype richness, evenness, as well as comparative analyses
	3. Moderate resolution (banding pattern), can also be combined with sequencing for additional information on genetic diversity	3. Data may be obtained for (theoretically) all dsDNA viruses in a given sample
	4. Procedures and analysis can be performed by an individual researcher	4. Requires no prior sequence knowledge
Weaknesses	1. Data may not be obtained for all viral groups	1. High cost
	2. Allows for comparative analyses	2. Annotation not necessarily high throughput
	3. Development of primers may require extensive trial and error	3. Procedures and analysis typically require a team of researchers
		4. Coverage depends on sampling efficiency and depth of sequencing
		5. Lack of viral taxonomic information to classify unknown sequences

24.1.8 Impacts of Virus Activity in Soils

24.1.8.1 Bacterial Mortality, Clonal Diversity, and Community Succession

As with viruses in marine ecosystems, viruses in soils play important roles in mediating bacterial mortality and controlling the clonal diversity of bacterial assemblages. Until recently, these impacts have only been demonstrated using specific phage-host systems. For example, rhizobia-specific phages (rhizobiophages) can exert selective pressures on the population of soil rhizobia, a broad class of nitrogen-fixing bacteria that engage in symbiotic relationships with leguminous plants. Infection and lysis of sensitive rhizobia strains results in a prevalence of phage-resistant mutant rhizobia that tend to be poor at colonizing plant roots (Kleczkowska, 1971). Apparently, the loss of phage receptor that results in resistance to infection also negatively impacts the microbe–plant interactions that are critical to productive symbioses. In addition, the presence or absence of specific rhizobiophages within a soil ecosystem can dictate which of many competing rhizobia strains are available for successful nodulation (Hashem and Angle, 1990). Beyond the impacts on bacterial mortality and selection, the lytic activity of rhizobiophages influences the efficacy of the legume–rhizobia symbiosis and may contribute to rates of nitrogen fixation and overall soil fertility.

Field experiments focusing on the population dynamics of specific, cultivable bacteriophages in the sugar beet rhizosphere revealed that phage and host dynamics exhibit recurring annual cycles, with temporal dynamics of phage and host populations appearing almost identical over three consecutive years (Ashelford et al., 1999a, 1999b). Thus, soil viruses appear to play important roles in the succession of microbial communities within soils.

Recent microcosm experiments indicate that soil bacterial communities and associated viral assemblages are highly dynamic and responsive to perturbations. The addition of various carbon amendments to soil within microcosm incubations resulted in significant fluctuations in bacterial abundance. Furthermore, a distinct coupling effect was observed, wherein an increase or decrease in bacterial abundance co-occurred with a respective decrease or increase in viral abundance, albeit with a time lag between virus and host fluxes (Srinivasiah et al., 2008). These patterns of staggered fluctuations between soil viruses and soil bacteria closely follow conventional models of predator–prey dynamics. Thus, results from microcosm experiments suggest that viral-mediated mortality is a significant factor controlling bacterial abundance, and by association, biomass turnover, and rates of nutrient cycling in soil ecosystems.

24.1.8.2 Phage Conversion, Host Fitness, and Horizontal Gene Transfer

Temperate phages are known to affect host fitness through a process known as *lysogenic conversion*. Prophage elements may contain genes that are beneficial to host survival and proliferation under specific conditions, thus providing a selective advantage over nonlysogenized hosts (Barksdale and Arden, 1974; Barondess and Beckwith, 1995; Waldor and Mekalanos, 1996).

Prophages also frequently constitute the sole differences in specific gene complement among closely related strains of bacteria and, consequently, the range of relationships between the lysogenic host bacterium and other organisms (Ohnishi et al., 2001; Desiere et al., 2002; Nakagawa et al., 2003; van der Mee-Marquet et al., 2006). Most of the data supporting these observations come from clinical strains of bacteria, and very little data are available regarding these same phage impacts on environmentally relevant strains. However, recent investigations indicate that lysogeny is highly prevalent among autochthonous communities of soil bacteria, with inducible lysogens (i.e., bacteria carrying a prophage element) accounting for 4%–66% of the total population in a given soil (Williamson et al., 2007). Thus, if the impacts of temperate phages observed in clinical settings hold true for environmental bacteria, the high frequency of lysogeny among soil bacteria implies that temperate viruses in soils significantly influence host fitness and evolution through the genetic process of lysogenic conversion.

Viruses may have significant impacts on host population genetics by acting as vectors for the horizontal transfer of genes between dissimilar bacterial strains. Bacteria may horizontally acquire novel gene sets via three mechanisms: (1) transformation, the uptake of naked DNA from the environment by competent cells; (2) conjugation, the transfer of genetic material from one cell to another through a pilus; and (3) transduction, phage-mediated transfer of DNA. Transduction begins as an error during the phage replication process in which host genomic DNA is mistakenly packaged into virus capsids. Two types of transduction are possible, generalized or specialized, depending on the phenotype of the phage mediating the transfer.

Generalized transducing phages, which may be virulent or temperate, carry any gene from the host genome with nearly equal frequency regardless of its location on the chromosome. In contrast, specialized transduction is mediated only by temperate phages that integrate into the host chromosome. Only specific bacterial genes flanking the prophage insertion site are transferred due to improper excision of the phage genome from the host chromosome (Levy and Miller, 1989; Campbell, 2006). While transduction is a rare event leading to a low frequency of gene transfer, the short generation times and fast growth rates of bacteria means that even rare recombinations may quickly alter the genotypic diversity of a bacterial community (Beumer and Robinson, 2005).

In aquatic studies, the presence of particulate matter increased transduction frequencies by providing surface area on which phages and host cells could aggregate and interact (Ripp and Miller, 1995). Soil, especially the clay and organic matter fractions, has an extremely high specific surface area ($80-800 \, m^2 \, g^{-1}$; Hillel, 1998), providing many potential sites for phage and host adsorption and interaction. However, natural transduction has never been observed in soil environments. Furthermore, only a handful of experiments using cultivable phage-host systems have been performed, all of which underscore the potential for transduction in soils (Vettori et al., 1999, 2000; Deb et al., 2003). A more recent study of the importance of phage-mediated gene transfer in soils examined viral assemblages that were obtained from the induction of soil bacteria in atrazine-treated soils. Subsequent PCR screening detected the presence of the atrazine catabolic gene *trzN* within the induced phage community, indicating that this important phenotypic trait could be transferred among bacterial populations through phage vectors (Ghosh et al., 2008). While much uncertainty remains regarding the frequency of transduction in soil environments, the possible impacts on bacterial genetic diversity and evolution are immense.

Acknowledgments

References

Ackermann, H.W. 2001. Frequency of morphological phage descriptions in the year 2000. Arch. Virol. 146:843–857.

Angly, F.E., B. Felts, M. Breitbart, P. Salamon, R.A. Edwards, C. Carlson, A.M. Chan et al. 2006. The marine viromes of four oceanic regions. PLoS Biol. 4:2121–2131.

Angly, F., B. Rodriguez-Brito, D. Bangor, P. McNairnie, M. Breitbart, P. Salamon, B. Felts, J. Nulton, J. Mahaffy, and F. Rohwer. 2005. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. BMC Bioinformatics 6:41–50.

Ashelford, K.E., M.J. Day, M.J. Bailey, A.K. Lilley, and J.C. Fry. 1999a. In situ population dynamics of bacterial viruses in a terrestrial environment. Appl. Environ. Microbiol. 65:169–174.

Ashelford, K.E., M.J. Day, and J.C. Fry. 2003. Elevated abundance of bacteriophage infecting bacteria in soil. Appl. Environ. Microbiol. 69:285–289.

Ashelford, K.E., J.C. Fry, M.J. Bailey, A.R. Jeffries, and M.J. Day. 1999b. Characterization of six bacteriophages of *Serratia liquefaciens* CP6 isolated from the sugar beet phytosphere. Appl. Environ. Microbiol. 65:1959–1965.

Ashelford, K.E., S.J. Norris, J.C. Fry, M.J. Bailey, and M.J. Day. 2000. Seasonal population dynamics and interactions of competing bacteriophages and their host in the rhizosphere. Appl. Environ. Microbiol. 66:4193–4199.

Barksdale, L., and S.B. Arden. 1974. Persisting bacteriophage infections, lysogeny, and phage conversions. Annu. Rev. Microbiol. 28:265–299.

Barondess, J.J., and J. Beckwith. 1995. *Bor* gene of phage-lambda, involved in serum resistance, encodes a widely conserved outer-membrane lipoprotein. J. Bacteriol. 177:1247–1253.

Bertani, G. 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental, systems. J. Bacteriol. 186:595–600.

Beumer, A., and J.B. Robinson. 2005. A broad-host-range, generalized transducing phage (SN-T) acquires 16S rRNA genes from different genera of bacteria. Appl. Environ. Microbiol. 71:8301–8304.

Breitbart, M., J.H. Miyake, and F. Rohwer. 2006. Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol. Lett. 236:245–252.

Budge, G.E., J. Loram, G. Donovan, and N. Boonham. 2008. RNA2 of soil-borne cereal mosaic virus is detectable in plants of winter wheat grown from infected seeds. Eur. J. Plant Pathol. 120:97–102.

Campbell, A. 2006. General aspects of lysogeny, p. 66–73. *In* R. Calendar (ed.) The bacteriophages. Oxford University Press, New York.

Chancellor, D.D., S. Tyagi, M.C. Bazaco, S. Bacvinskas, M.B. Chancellor, V.M. Dato, and F. de Miguel. 2006. Green onions: Potential mechanism for hepatitis A contamination. J. Food Prot. 69:1468–1472.

Christian, P.D., A.R. Richards, and T. Williams. 2006. Differential adsorption of occluded and nonoccluded insect-pathogenic viruses to soil-forming minerals. Appl. Environ. Microbiol. 72:4648–4652.

Deb, C., R. Chakraborty, A.N. Ghosh, N.C. Mandal, T. Mukherjee, and P. Roy. 2003. A generalized transducing thiophage (TPC-1) of a facultative sulfur chemolithotrophic bacterium, *Bosea thiooxidans* CT5, of alpha-proteobacteria, isolated from Indian soil. FEMS Microbiol. Lett. 227:87–92.

Delogu, G., N. Faccini, R. Alberici, A. Gianinetti, and A.M. Stanca. 2003. Soil-borne viruses of barley seriously affect plant growth and grain yield in a monocropping system. Cereal Res. Commun. 31:137–144.

Desiere, F., S. Lucchini, C. Canchaya, M. Ventura, and H. Brussow. 2002. Comparative genomics of phages and prophages in lactic acid bacteria. Antonie Van Leeuwenhoek Int. J. Gen. Mol. Microbiol. 82:73–91.

Duboise, S., B.E. Moore, C.A. Sorber, and B.P. Sagik. 1979. Viruses in soil systems. CRC Crit. Rev. Microbiol. 7:245–285.

Farzadfar, S., R. Pourrahim, A.R. Golnaraghi, and A. Ahoonmanesh. 2007. Surveys of beet necrotic yellow vein virus, beet soil-borne virus, beet virus Q and *Polymyxa betae* in sugar beet fields in Iran. J. Plant Pathol. 89:277–281.

Feng, Y., A.C. Motta, D.W. Reeves, C.H. Burmester, E. van Santen, and J.A. Osborne. 2003. Soil microbial communities under conventional-till and no-till continuous cotton systems. Soil Biol. Biochem. 35:1693–1703.

Fierer, N., M. Breitbart, J. Nulton, P. Salamon, C. Lozupone, R. Jones, M. Robeson et al. 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl. Environ. Microbiol. 73:7059–7066.

Fillhart, R.C., G.D. Bachand, and J.D. Castello. 1998. Detection of infectious tobamoviruses in forest soils. Appl. Environ. Microbiol. 64:1430–1435.

Fuxa, J.R. 2004. Ecology of insect nucleopolyhedroviruses. Agric. Ecosyst. Environ. 103:27–43.

Gerba, C.P. 1984. Applied and theoretical aspects of virus adsorption to surfaces. Adv. Appl. Microbiol. 30:133–168.

Ghosh, D., K. Roy, K.E. Williamson, D.C. White, K.E. Wommack, K.L. Sublette, and M. Radosevich. 2008. Prevalence of lysogeny among soil bacteria and presence of 16S rRNA and *trzN* genes in viral-community DNA. Appl. Environ. Microbiol. 74:495–502.

Hashem, F.M., and J.S. Angle. 1990. Rhizobiophage effects on nodulation, nitrogen-fixation, and yield of field-grown soybeans (Glycine-Max L Merr). Biol. Fertil. Soils 9:330–334.

Helgason, B.L., F.L. Walley, and J.J. Germida. 2009. Fungal and bacterial abundance in long-term no-till and intensive-till soils of the Northern Great Plains. Soil Sci. Soc. Am. J. 73:120–127.

Helton, R.R., and K.E. Wommack. 2009. Seasonal dynamics and metagenomic characterization of estuarine viriobenthos assemblages by randomly amplified polymorphic DNA PCR. Appl. Environ. Microbiol. 75:2259–2265.

Hillel, D. 1998. Environmental soil physics. Harcourt Brace and Company, San Diego, CA.

Jin, Y., and M. Flury. 2002. Fate and transport of viruses in porous media. Adv. Agron. 77:39–51.

Keel, C., Z. Ucurum, P. Michaux, M. Adrian, and D. Haas. 2002. Deleterious impact of a virulent bacteriophage on survival and biocontrol activity of *Pseudomonas fluorescens* strain CHA0 in natural soil. Mol. Plant-Microbe Interact. 15:567–576.

Kleczkowska, J. 1971. Genetic changes in rhizobium bacteria and in their bacteriophages during coexistence. Plant Soil 35:47–56.

Levy, S.B., and R.V. Miller. 1989. Gene transfer in aquatic environments. McGraw-Hill, New York.

Lipson, S.M., and G. Stotzky. 1986. Effect of kaolinite on the specific infectivity of reovirus. FEMS Microbiol. Lett. 37:83–88.

Loveland, J.P., J.N. Ryan, G.L. Amy, and R.W. Harvey. 1996. The reversibility of virus attachment to mineral surfaces. Colloids Surf. A 107:205–221.

Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380.

Marsh, P., and E.M.H. Wellington. 1994. Phage-host interactions in soil. FEMS Microbiol. Ecol. 15:99–107.

Melzer, M.J., and M.J. Bidochka. 1998. Diversity of double-stranded RNA viruses within populations of entomopathogenic fungi and potential implications for fungal growth and virulence. Mycologia 90:586–594.

Meschke, J.S., and M.D. Sobsey. 2003. Comparative reduction of Norwalk virus, poliovirus type 1, F+ RNA coliphage MS2 and *Escherichia coli* in miniature soil columns. Water Sci. Technol. 47:85–90.

Nakagawa, I., K. Kurokawa, A. Yamashita, M. Nakata, Y. Tomiyasu, N. Okahashi, S. Kawabata et al. 2003. Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. Genome Res. 13:1042–1055.

Noguchi, H., J. Park, and T. Takagi. 2006. MetaGene: Prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Res. 34:5623–5630.

Ohnishi, M., K. Kurokawa, and T. Hayashi. 2001. Diversification of *Escherichia coli* genomes: Are bacteriophages the major contributors? Trends Microbiol. 9:481–485.

Pantastico-Caldas, M., K.E. Duncan, C.A. Istock, and J.A. Bell. 1992. Population dynamics of bacteriophage and *Bacillus subtilis* in soil. Ecology 73:1888–1902.

Pourcher, A.M., P.B. Francoise, F. Virginie, G. Agnieszka, S. Vasilica, and M. Gerard. 2007. Survival of faecal indicators and enteroviruses in soil after land-spreading of municipal sewage sludge. Appl. Soil Ecol. 35:473–479.

Prestel, E., S. Salamitou, and M.S. Dubow. 2008. An examination of the bacteriophages and bacteria of the Namib desert. J. Microbiol. 46:364–372.

Prigent, M., M. Leroy, F. Confalonieri, M. Dutertre, and M.S. DuBow. 2005. A diversity of bacteriophage forms and genomes can be isolated from the surface sands of the Sahara Desert. Extremophiles 9:289–296.

Ptashne, M. 1991. A genetic switch. Blackwell Scientific Publications and Cell Press, Cambridge, MA.

Ripp, S., and R.V. Miller. 1995. Effects of suspended particulates on the frequency of transduction among *Pseudomonas aeruginosa* in freshwater environments. Appl. Environ. Microbiol. 61:1214–1219.

Santamaria, J., and G.A. Toranzos. 2003. Enteric pathogens and soil: A short review. Int. Microbiol. 6:5–9.

Srinivasiah, S., J. Bhavsar, K. Thapar, M. Liles, T. Schoenfeld, and K.E. Wommack. 2008. Phages across the biosphere: Contrasts of viruses in soil and aquatic environments. Res. Microbiol. 159:349–357.

Stewart, F.M., and B.R. Levin. 1984. The population biology of bacterial viruses: Why be temperate? Theor. Popul. Biol. 26:93–117.

Swanson, M.M., G. Fraser, T.J. Daniell, L. Torrance, P.J. Gregory, and M. Taliansky. 2009. Viruses in soils: Morphological diversity and abundance in the rhizosphere. Ann. Appl. Biol. 155:51–60.

Torkzaban, S., S.M. Hassanizadeh, J.F. Schijven, and H. van den Berg. 2006. Role of air–water interfaces on retention of viruses under unsaturated conditions. Water Resour. Res. 42:11.

van der Mee-Marquet, N., A.S. Domelier, L. Mereghetti, P. Lanotte, A. Rosenau, W. van Leeuwen, and R. Quentin. 2006. Prophagic DNA fragments in *Streptococcus agalactiae* strains and association with neonatal meningitis. J. Clin. Microbiol. 44:1049–1058.

Vettori, C., E. Gallori, and G. Stotzky. 2000. Clay minerals protect bacteriophage PBS1 of *Bacillus subtilis* against inactivation and loss of transducing ability by UV radiation. Can. J. Microbiol. 46:770–773.

Vettori, C., G. Stotzky, M. Yoder, and E. Gallori. 1999. Interaction between bacteriophage PBS1 and clay minerals and transduction of *Bacillus subtilis* by clay-phage complexes. Environ. Microbiol. 1:347–355.

Waldor, M.K., and J.J. Mekalanos. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 272:1910–1914.

Weinbauer, M.G. 2004. Ecology of prokaryotic viruses. FEMS Microbiol. Rev. 28:127–181.

Wen, K., A.C. Ortmann, and C.A. Suttle. 2004. Accurate estimation of viral abundance by epifluorescence microscopy. Appl. Environ. Microbiol. 70:3862–3867.

Williamson, K.E., M. Radosevich, D.W. Smith, and K.E. Wommack. 2007. Incidence of lysogeny within temperate and extreme soil environments. Environ. Microbiol. 9:2563–2574.

Williamson, K.E., M. Radosevich, and K.E. Wommack. 2005. Abundance and diversity of viruses in six Delaware soils. Appl. Environ. Microbiol. 71:3119–3125.

Williamson, K.E., J.B. Schnitker, M. Radosevich, D.W. Smith, and K.E. Wommack. 2008. Cultivation-based assessment of lysogeny among soil bacteria. Microb. Ecol. 56:437–447.

Williamson, K.E., K.E. Wommack, and M. Radosevich. 2003. Sampling natural viral communities from soil for culture-independent analyses. Appl. Environ. Microbiol. 69:6628–6633.

Winget, D.M., and K.E. Wommack. 2008. Randomly amplified polymorphic DNA PCR as a tool for assessment of marine viral richness. Appl. Environ. Microbiol. 74:2612–2618.

Wommack, K.E., S.R. Bench, J. Bhavsar, D. Mead, and T. Hanson. 2009. Isolation independent methods of characterizing phage communities 2: Characterizing a metagenome, p. 279–289. *In* M.R.J. Clokie and A.M. Kropinski (eds.) Bacteriophages: Methods and protocols, Vol. 2: Molecular and applied aspects. Humana Press, New York.

Wommack, K.E., J. Bhavsar, and J. Ravel. 2008. Metagenomics: Read length matters. Appl. Environ. Microbiol. 74:1453–1463.

Wommack, K.E., and R.R. Colwell. 2000. Virioplankton: Viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64:69–114.

Yin, X., L.R. Zeph, and G. Stotzky. 1997. A simple method for enumerating bacteriophages in soil. Can. J. Microbiol. 43:461–466.

Zhuang, J., and Y. Jin. 2003a. Virus retention and transport as influenced by different forms of soil organic matter. J. Environ. Qual. 32:816–823.

Zhuang, J., and Y. Jin. 2003b. Virus retention and transport through Al-oxide coated sand columns: Effects of ionic strength and composition. J. Contam. Hydrol. 60:193–209.

24.2 Structure and Function of Prokaryotes in Soil

Michael Schloter

24.2.1 Introduction

Prokaryotes can be regarded, based on pure numbers, as the most important group of organisms in soils. It has been estimated that 1 g of soil harbors up to 10 billion microorganisms, belonging to up to 100,000 different species (Amann et al., 1995). Craig Venter compared the amount of genetic information stored in 1 g of soil to that of 4000 human genomes. But not only the numbers and the diversity of microbes are impressive, the amount of stored carbon and nitrogen in the microbial biomass exceeds the mass, which is stored in the plant biomass (Whitman et al., 1998). Prokaryotes are generally regarded as the catalysts for the major turnover processes in soil. Some metabolic steps, such as the fixation of atmospheric nitrogen and the biosynthesis of methane, are exclusively performed by bacteria and archaea. Furthermore prokaryotes are capable of efficiently degrading many pollutants. Finally, microbial communities highly influence plant growth positively as so-called plant growth–promoting rhizobacteria or negatively as phytopathogens. Most prokaryotes can adapt to extreme conditions, such as high salinity, drought or water stress, which is necessary to survive the changing conditions in soils. Therefore, the structure and activity of prokaryotic communities serves as important parameters for soil quality and one of the major goals of agricultural management is the steering of soil microbial communities ever since, to improve sustainable crop yield and quality.

Compared to eukaryotes, prokaryotes are characterized by a relatively simple structural level. They are unlike most eukaryotes in many cases unicellular and have no nucleus. Often they are much smaller than eukaryotes. Most prokaryotes are between 1 and 10 μm. Generation time of prokaryotes is often much faster compared to eukaryotes. Furthermore, prokaryotes never harbor organelles like mitochondria or chloroplasts. However, prokaryotes do possess some internal structures, such as cytoskeletons. The genome of prokaryotes is characterized by a DNA–protein complex in the cytosol called the nucleoid, which lacks a nuclear envelope. Often prokaryotes contain small ring-like DNA fragments, called plasmids, which are easily transferable between organisms. These plasmids encode, for example, for resistance genes, which make prokaryotes tolerant against antibiotics or heavy metals. Prokaryotes reproduce through asexual reproduction. While prokaryotes are still commonly imagined to be strictly unicellular, most are capable of forming so-called biofilms. Both eukaryotes and prokaryotes contain large RNA/protein structures called ribosomes, which are involved in protein expression. Among the prokaryotes, two phylogenetically distinct domains could be distinguished: bacteria and archaea. It is generally accepted that the

TABLE 24.3 Some Important Differences between Prokaryotes and Eukaryotes

	Prokaryotes (Eubacteria and Archaea), pro = "before" and karyon = "nucleus"	Eukaryotes
Size of individual cells	1–10 μm	10 μm to 1 mm
Cell wall	Peptidoglycan (eubacteria)	Cellulose and chitin (plants and fungi)
Cellular organization	Mostly single cells	Mostly differentiation in several cell types
Organelles	No	Yes (mitochondria, ER, chloroplasts, etc.)
Location of genome	Naked ring-like DNA	Nucleus
Genome structure	One chromosome and plasmids	Several chromosomes (exon–intron structure)
Metabolism	Aerobic and anaerobic	Aerobic and anaerobic
RNA and proteins	Synthesis in same compartment	RNA synthesis in nucleus; protein synthesis in cytoplasm
Size of ribosomes	70S (30S + 50S)	80S (40S + 60S)
Reproduction	Binary fission	Meiosis (sexual reproduction) mitosis (cell division)
Evolution		More advanced having arisen from prokaryote-like predecessors

first living organisms were prokaryotes, approximately 3.5 billion years ago. Even today, prokaryotes are perhaps the most successful and abundant form of life. Details about prokaryotes can be found in reviews by Woese (1994) and Ogunseitan (2005). Some differences between Eukaryotes and Prokaryotes are given in Table 24.3.

However, for a long period, no tools were available to characterize microbes from soil, and even today, we still have huge difficulties in describing microbial diversity in this habitat.

24.2.2 The Methodological Challenge

Almost 30 years ago Vigdis Torsvik (Torsvik, 1980) published an article in *Soil Biology and Biochemistry* describing for the first time the isolation of bacterial DNA from soil. This paper is regarded as the initialization of molecular microbial ecology in soil.

Up to that date, most papers that had been dealing with microbial diversity and the corresponding physiological properties were based on isolation and traditional cultivation-based approaches, as it had been assumed that most microbes from soil can be easily isolated. And indeed this approach was useful in a way to get several thousand isolates in hand and to describe them. However, studies from the last decades have proven that this number of organisms is just the tip of the iceberg and true diversity in soil might be than 1000 above that, what is known

from cultivation-based approaches (Amann et al., 1995). Direct, simple staining of microorganisms in different natural compartments have indicated that the actual number of cells that can be found in different environmental samples exceed diversity based on plate counting by 2–4 orders of magnitude (Amann et al., 1995). These data gave a first indication that in general, less than 1% of the microflora from a soil sample can be assessed with simple cultivation methods. Unfortunately, some authors have greatly simplified these findings and concluded that generally only 1% of all species can be isolated. This statement could be disproven simply by improving cultivation techniques, especially by adapting the media and the cultivation conditions to the appropriate environmental conditions. For example, Bruns et al. (2002) could show that the addition of cyclic AMP and homoserine lactones to traditional media, improves the cultivation efficiency of heterotrophic bacteria from the Black Sea. There are several hypotheses that explain reasons why microbes from the environment are difficult to be cultivated.

1. The so-called noncultivable cells are genetically closely related to culturable representatives, but their physiological status in the environment makes their isolation difficult or impossible. This interpretation seems plausible, because it is well known from pure culture studies, that under certain conditions, for example, starvation, microbes can turn into a status that is described as "alive but not culturable" (Binnerup et al., 1993). If this hypothesis is the only valid one, this would mean that up to 99% of the yet noncultured microbes are closely related to the 1% of microbes that have been isolated so far and just differ in their physiological status from the well-characterized isolates. However, molecular data indicate that diversity in soil is far increased to what we know so far from the cultured isolates.

2. Therefore, there is general agreement among scientists that there must be an additional reason why microbes from soil cannot be cultured easily. The second hypothesis, which might explain the difficulties in cultivation of the majority of microorganisms from environmental samples, assumes that many of these microorganisms belong to completely new lineages, which are phylogenetically not related to isolated microbes, which makes their isolation on standard media impossible. If this hypothesis is true, there must be still tens of thousands of new species in the nature, which until now were not isolated. A good indication on the proportion of isolated to yet undiscovered microbes is based on a comparison of sequences from marker genes (e.g., 16S rRNA genes) from isolates and PCR products of the corresponding gene from directly extracted DNA from soil (Pace, 1997). It is therefore not surprising that even today the majority of research in the field of "molecular microbial ecology," is based on the question how to identify new species from environmental samples and how to link their dynamics and activity to ecosystem functioning.

3. A third hypothesis derives from the fact that the nutrient supply in soils usually is very limited and extremely complex. Typical soil bacteria belong mostly to the so-called k-strategists, with relatively long generation times and high substrate affinity for complex substrates. Common culture conditions, however, are more likely to "r-strategists" with fast generation times and nutrients with a low complexity adjusted. Even taking generation times of only a few hours into account for typical soil bacteria would result in incubation times of several years to acquire a visible colony on an agar plate (Bakken, 1997; Vieira and Nahas, 2005).

The first step toward a broader assessment of soil microbes from soil was done in the eighties of the last century when PCR was introduced into soil microbiology and nucleic acids (mainly DNA) from soil samples was directly extracted and amplified to build a catalog of information, with the aim to clarify the question, "What is where?" In the pioneering work by Torsvik et al. (1990), the diversity of microbial communities in nature was studied with the help of DNA–DNA reassociation kinetics. The results indicated that DNA, directly extracted from the soil by fractional centrifugation, is much more complex than what had been expected at that time and that 1 g of soil harbors up to 10,000 different genomes (Torsvik et al., 1990). This study provided the basis for all investigations in microbial ecology.

The genes of the rRNA have been used as a marker for genetic diversity in the environment in the last two decades. The rRNA is composed of several conserved areas. Besides strongly conserved areas, which are similar in almost all organisms, the rRNA genes also contain variable and hypervariable areas, which allow a distinction up to the species level. The different conserved areas of the rRNA can be used also as "molecular clock" meaning that differences between two organisms in the rRNA gene sequences directly correlate with their phylogeny (Gupta, 2000, 2004; Rappe and Giovannoni, 2003). The highly conserved flanking areas on the other hand are ideal target regions for primers and can be used for an amplification of the corresponding variable regions by PCR. Until today, in public accessible databases approximately 65,000 rRNA sequences have been deposited (www.ncbi.nlm.nih.gov). This enormous amount of data allows a very precise phylogenetic assignment of unknown sequences from environmental DNA. Therefore, rRNA genes are often called the "gold standard" in microbial ecology. The characterization of the obtained PCR fragments can be done by different fingerprinting methods. This includes denaturing gradient gel electrophoresis (DGGE), the analysis of single-strand conformation polymorphism (SSCP), or restriction length polymorphism (RFLP) (Liu et al., 1997; Muyzer et al., 1998; Dohrmann and Tebbe, 2004). This technique will be described in this book by Kornelia Smalla. However, since the introduction of capillary sequencing technology in microbial ecology amplicons are often directly cloned and sequenced. This approach offers the advantage that the sequences can be directly analyzed with the help of appropriate software programs (e.g., ARB, Ludwig et al., 2004)

and used as the basis for the development of fluorescence in situ hybridization (FISH), which can be used for the in situ detection of the corresponding phylotypes in soil (full rRNA cycle approach, Amann et al., 1995).

Interestingly, even today, almost 30 years after the first extraction of nucleic acids from soil has been described, no standard protocol is available. There are only some very general criteria for nucleic acids extracted from soils available, if they should be amplified by PCR (Thakuria et al., 2008). This include the ratio OD 260/230 nm respectively 260/280 nm, which give information about the quality of nucleic acids extracted and the amount of humic substances coextracted, which might inhibit PCR reaction. Interestingly, in the paper stated above, it could be shown that not only the quantity and quality of DNA extracted from soil is depending on the protocol used but also the obtained diversity pattern, as different methods show different extraction efficiencies of nucleic acids for selected microbial groups from soil. This can result in changes in diversity patterns from the same soil, depending on the methods used for extraction of nucleic acids (Bakken and Frostegard, 2006).

However, also molecular, PCR-based methods only allow the detection of a part of the overall microbial diversity in soil. This fact is mainly related to general features of the PCR reaction, as mainly the amplicons generated in the first cycles of PCR determine the final composition of PCR products at the end after 30 or 35 cycles. This so-called preferential amplification of certain target sequences is a well-described phenomena and becomes more important if so-called universal primer sets are used to describe the whole bacterial or archaeal community (Becker et al., 2000). If more targeted primer sets are used to describe the diversity of certain phylogenetic groups, this bias becomes less importance. Furthermore, the generation of microbial fingerprints based on the rRNA gene does only allow a description of communities based on the species level. Ecotypes and strains cannot be distinguished by this approach. Therefore, individual ecotypes or strains must be characterized by other methods (Schloter et al., 2000).

Therefore, it is often suggested to use a combination of methods to describe microbial communities in soil that are not only based on the genotype of microbial communities but also in their phenotype (e.g., analysis of phospholipid acid pattern from the outer membrane or cell wall of microbes, PLFA; Zelles, 1997). These approaches are complementary to the molecular analysis, though mainly detection on the level up to the domains and large groups is possible and given environmental conditions can influence the results. However, PLFA analysis compared to PCR-based methods is less biased, since the extracts are analyzed by GC/MS directly. For detailed studies on the level of ecotypes up to now, only cultivation-dependent methods have been described (Schloter et al., 2000), including all problems mentioned above.

It was therefore suggested by Tebbe and Schloter (2006) that various methodological approaches should be combined to describe microbial communities in soil. The corresponding "polyphasic approach" is shown in Figure 24.4. Such a concept may help to overcome possible methodological limitations that

Consequences for the ecosystem

Sensitivity of the method

Microbial biomass
(z.B. Substrate induced respiration) (1)

Domains and families
(PLFA pattern) (2)

Species
(PCR-based methods) (3)

Intraspecies diversity (Microdiversity) (4)
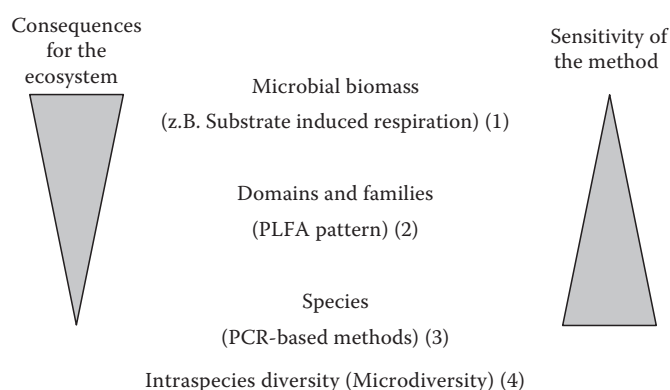
FIGURE 24.4 Hierarchic–polyphasic approach to characterize microbial diversity in soils (1) Anderson et al. (1978); (2) Zelles (1997); (3) Amann et al. (1995); and (4) Schloter et al. (2000).

can occur when only one technique is applied on the one hand and to describe diversity of microbes at different taxonomic and phylogenetic levels.

In future, however, we might see a second revolution in microbial ecology, as more and more "high throughput sequencing" approaches become popular and are also used in microbial ecology. These new methods allow a genetic assessment of whole "metagenomes" from soil without generation of PCR products in very short time periods and to very cheap prices. However, there are still huge problems on how to handle and evaluate the generated sequence data (Urich et al., 2008).

24.2.3 Parameters Controlling the Composition of Prokaryotic Communities in Soils

Plants are highly influencing the composition of bacteria and archaea in the rhizosphere. These interactions between plants and prokaryotes are manifold and complex. They range from general nutrient mobilization and implementation to production of specific signaling molecules that allow a specific interaction. These interactions form the basis for both a natural nutrient management of plants as well as for healthy vegetation. However, it should not been overseen that also bulk soil microflora influences composition of bacteria and fungi in the rhizosphere (De Ridder-Duine et al., 2005).

The interactions between plants and prokaryotes are mainly driven by the carbon flow in the rhizosphere. The compartment of soil, that is directly influenced by the root, has been defined as rhizosphere. The rhizosphere is characterized by a high microbial biomass and activity and is considered as the most active soil habitat. The plants drive, by the excretion of exudates and nutrients, the composition of bacteria and archaea in the rhizosphere. The main components of carbon within the rhizosphere are organic acids, amino acids, and sugars in the rhizosphere of young roots and high-polymeric substances in the rhizosphere of older roots (Killham and Yeomans, 2001). In addition, certain signaling molecules such as furanones formed by the plant and bacterial *N*-acyl homoserine lactones induce

a strong cross talk between biota in the rhizosphere (Teplitski et al., 2000). Recent papers have shown additionally the importance of plant derived phenols, for the regulation of microbial activities in the rhizosphere (Bais et al., 2006). The mean turnover rates of organic compounds are approximately 2.5 years in rhizosphere soils. In contrast, in habitats with a relatively inactive microflora, such as the B horizon of an arable soil, turnover rates from 1000 to 2000 years for organic compounds have been calculated (Whitman et al., 1998).

Conversely, plants benefit largely from most of the chemical transformation products catalyzed by bacteria and archaea. The interaction between plant and prokaryotes can be even closer and symbiosis-like structures can be formed. The best known symbiosis is that involving rhizobia and legumes (Werner, 1992). In some cases, also other highly specific interactions between different microorganisms and some plant species have been described (Kowalchuk et al., 2002; Wardle, 2005). A typical example is the production of antibiotics by biocontrol-active microorganisms in the rhizosphere, resulting in a decreased risk of infection for the plant by phytopathogenic bacteria or fungi (Raaijmakers et al., 1997).

In addition to the rhizosphere, plants highly influence microbial community structure and function also in the litter layer, as the quality and quantity of the leaves that reach the ground drives the composition of the degrader community and consequently the kinetics of nutrient mobilization (Aneja et al., 2006).

Besides the interactions with plants, prokaryotes form interactions with most other biota in the soil like fungi, protozoa, nematodes, earthworms, and insects (Czarnetzki and Tebbe, 2003). Especially for organisms that belong to higher trophic levels, bacteria and archaea play an important role in the gastrointestinal tract for the degradation of complex organic material.

Of course, also other factors such as soil type, rainfall conditions, temperature profiles, etc. play an important role for the formation and dynamics of microbial communities in soil.

In agricultural used soils, in addition to factors mentioned above, the type of agricultural management plays an important role for microbes. Mainly, the quality and quantity of fertilizers and pesticides used and the types of tillage and the cropping sequence influence the composition of prokaryotic as well as eukaryotic populations in soil. The individual variables are summarized in Figure 24.5.

24.2.4 Structure of Prokaryotic Communities in Soil

Both cultivation-dependent and cultivation-independent methods support the hypothesis that soils are one of the ecosystems with the highest microbial diversity on earth. Even soils, where a reduced diversity is expected, like in contaminated areas, diversity indices calculated were far above the expected (Perez de Mora et al., 2006). Investigations based on clone libraries of 16S rRNA gene amplicons have also shown that it is practically impossible to get identical sequence from the same soil even if more than 1000 clones are investigated (Podar et al., 2007).

Most of the previously identified sequences belong to the group of Proteobacteria (Zavarzin et al., 1991; Gupta, 2000), Cytophagales (Dungan et al., 2003), Actinobacteria (Holmes et al., 2000; Ritz et al., 2004), or to the Gram-positive bacteria with low GC content (Felske et al., 1998). In the last two decades, quite often sequences from the Holophaga-Acidobacterium have been described from different soil samples (Ludwig et al., 1997; Barns et al., 1999; Felske et al., 1999) detected. Also, members of the group of Verrucomicrobia seem to be important members of soil microbial communities (Barns et al., 1999). Interestingly, there is increasing evidence that also sequences from organisms, which have been considered so far as less important from soil, such as planctomycetes (Wang et al., 2002) and spirochetes (Varm et al., 1994) can be easily detected in soil. Even the presence of high numbers of archaea in soils has been well described
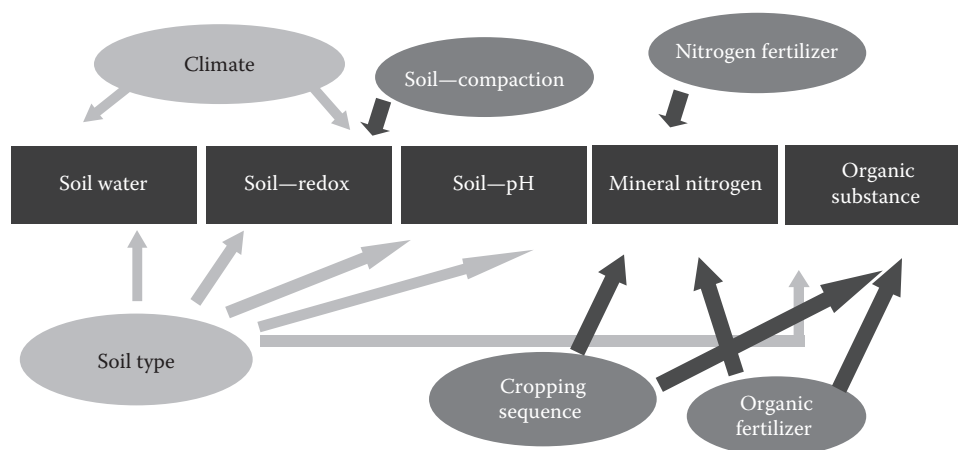
FIGURE 24.5   Factors influencing microbial community structure and function in arable soils (light color: site; dark color: management).

based on typical biomarkers as well as on the basis of molecular methods (Leininger et al., 2006).

In comparison to the data published for the domains Bacteria and Archaea, results on the basis of 18S rRNA or its sequences for the description of fungal communities are rare; thus, a generalization, as described above for fungal communities is hardly possible at the moment based on the existing data.

24.2.5 From Structure to Function of Prokaryotic Communities in Soils

The structure of microbial communities does not necessarily help to clarify their function. Exceptions are simple ecosystems with relatively few species, such as microbial biofilms in hot springs. Therefore, studies on structure–function relationships were initially limited to those ecosystems. However, if processes and nutrient fluxes in complex biogeosystems should be better understood, information about microbial enzymes that are related to a particular turnover process and their expression in soils are of essential importance, as most of the ecologically relevant functions are distributed among different taxonomic groups of microbes in soil. For example, bacteria from the phyla Actinobacteria, Proteobacteria, Firmicutes, and Bacteroidetes are able to reduce cellulose (Ulrich and Wirth, 1999). The same is true for bacteria that are able to reduce nitrate or to fix atmospheric nitrogen (Zehr et al., 2003). The extent to which these functional redundancies contribute to the stability of soils is still unclear. However, it is clear that the development of such genetic diversity pattern in soils is strongly influenced by the microhabitat structure of soils (Gattinger et al., 2002; Webster et al., 2002; Franklin and Mills, 2003; Ritz et al., 2004).

To assess the potential of a soil for a specific turnover process, the amount and diversity of genes catalyzing the corresponding reaction serves as a good indicator. However, the analysis by PCR-based methods is often difficult, as appropriate primer systems are missing and often functionally identical genes in different organisms have low homology levels. One reason for this high divergence might be related to the fact that organisms, which share the same function, have developed independently during evolution, and, therefore, the coding sequences in particular at the nucleic acid level are strongly divergent. For example, exoproteases have a history of over 3.5 billion years (Mrkonjic et al., 2008). Therefore, not surprisingly in bacteria three classes of proteases have developed, which share the same function but are completely different in their structure and genetic composition (neutral-, metallo- and serine-proteases). If the entire bacterial community of a soil covering proteolytic activities should be measured for each of these enzyme classes, specific primers have to be developed. Other functional genes are significantly higher conserved, which indicates not coevolution but horizontal gene transfer. For example, the genes responsible for resistance against certain antibiotics are highly conserved in most bacteria mainly on the amino acid level

(Heuer and Smalla, 2007). Similarly, most of the genes coding for enzymes for the degradation of xenobiotics show high similarity. Many of these genes are localized on mobile genetic elements (Springael and Top, 2004). Overall, a transfer of complete gene sequences in soil between microorganisms is not a rare event (Clerc and Simonet, 1998). Especially in nutrient-rich microhabitats (such as the rhizosphere) and under the appropriate selection pressure, a transfer of genetic information can easily been demonstrated (Hoffmann et al., 1998; Smit et al., 1998; Thimm et al., 2001).

The alignments for the development of primer systems for the detection of functional genes can be done using existing sequences from databases in silico. However, the number of sequences available for functional genes is often very low and cannot be compared to the number of sequences present in databases for ribosomal genes, especially for the 16S rRNA coding gene. With a modification of the PCR, besides a qualitative assessment of diversity, also the quantification of genes of interest is possible (real time PCR; Sharma et al., 2007b).

However, if the actual—not the potential—physiological performance of soils should be related to specific microbial activities, it is not enough to determine the amount of the corresponding genes on DNA basis. Further investigations including the transcription or expression level are highly needed. Approaches for the extraction and analysis of mRNA (Sharma et al., 2006, 2007a) and proteins (Roberts and Jones, 2008) have been developed in recent years. Mainly the short half-life of mRNA is a major methodological and conceptual problem, even if for most genes in soil half times >5 h can be assumed. Therefore, the analysis of mRNA might be useful but is more a snapshot of an actual status at the time of sampling and varies highly in time and space. At the protein level, it is important that a targeted approach for the detection of individual proteins is used, since a total extraction of all proteins from soil is still extremely difficult (Hartmann et al., 2005; Nannipieri, 2006). Overall, however, a collection of all three levels of the genome–transcriptome–proteome is needed to link processes and fluxes in soils to microbial activity pattern and performance (Figure 24.6).
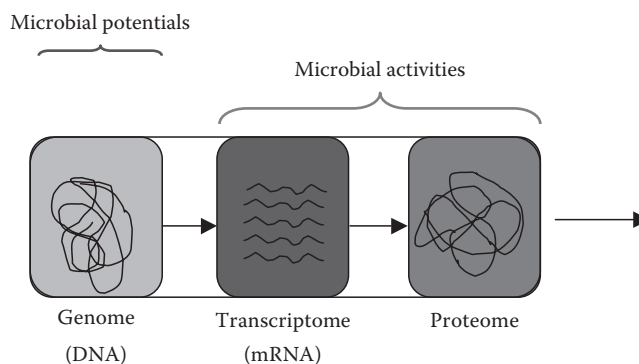
FIGURE 24.6 Assessment of microbial diversity at different levels.

References

Amann, R.I., W. Ludwig, and K.H. Schleifer. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59:143–169.

Anderson, J.P.E., and K.H. Domsch. 1978. A physiological method for the quantitative measurement of microbial biomass in soils. Soil Biol. Biochem. 10:215–221.

Aneja, M., S. Sharma, M. Schloter, and J.C. Munch. 2006. Microbial degradation of beech litter—Influence of soil type and litter quality on the structure and function of microbial populations involved in the turnover process. Microb. Ecol. 52:127–135.

Bais, H.P., T.L. Weir, L.G. Perry, S. Gilroy, and J.M. Vivanco. 2006. The role of root exudates in rhizosphere interactions with plants and other organisms. Ann. Rev. Plant Biol. 57:233–266.

Bakken, L.R. 1997. Culturable and unculturable bacteria in soil, p. 47–61. *In* J.D. van Elsas, E.M.H. Wellington, and J.T. Trevors (eds.) Modern soil microbiology. Marcel Decker Press, New York.

Bakken, L.R., and Å. Frostegård. 2006. Nucleic acid extractions from soil. *In* K. Smalla, and P. Nannipieri (eds.): Series on soil biology: Nucleic acids and proteins in soil. 350p. Springer Press, pp. 49–67.

Barns, S.M., S.L. Takala, and C.R. Kuske. 1999. Wide distribution and diversity of members of the bacterial kingdom Acidobacterium in the environment. Appl. Environ. Microbiol. 65:1731–1737.

Becker, S., P. Böger, R. Oehlmann, and A. Ernst. 2000. PCR bias in ecological analysis: A case study for quantitative taq nuclease assays in analyses of microbial communities. Appl. Environ. Microbiol. 66:4945–4953.

Binnerup, S.J., D.F. Jensen, H. Thordal-Christiansen, and J. Soerensen. 1993. Detection of viable, but non-culturable *Pseudomonas fluorescens* DF57 in soil using a microcolony epifluorescence technique. FEMS Microbiol. Ecol. 12:97–105.

Bruns, A., H. Cypionka, and J. Overmann. 2002. Cyclic AMP and acyl homoserine lactones increase the cultivation efficiency of heterotrophic bacteria from the central Baltic Sea. Appl. Environ. Microbiol. 66:3978–3987.

Clerc, S., and P. Simonet. 1998. A review of available systems to investigate transfer of DNA to indigenous soil bacteria. Antonie. Van. Leeuwenhoek. 73:15–23.

Czarnetzki, A.B., and C.C. Tebbe. 2003. Diversity of bacteria associated with Collembola—A cultivation-independent survey based on PCR-amplified 16S rRNA genes. FEMS Microbiol. Ecol. 49:217–227.

De Ridder-Duine, A.S., G. Kowalchuk, P.J.A. Klein Gunnewiek, W. Smant, J.A. van Veen, and W. de Boer. 2005. Rhizosphere bacterial community composition in natural stands of *Carex arenaria* (sand sedge) is determined by bulk soil community composition. Soil Biol. Biochem. 37:349–357.

Dohrmann, A.B., and C.C. Tebbe. 2004. Microbial community analysis by PCR-single-strand conformation polymorphism (PCR-SSCP), p. 334–360. *In* A.D.L. Akkermans, J.D. van Elsas, and F.J. de Bruijn (eds.) Molecular microbial ecology manual. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Dungan, R.S., A. Mark, I. Scott, and R. Yates. 2003. Effect of propargyl bromide and 1,3-dichloropropene on microbial communities in an organically amended soil. FEMS Microbiol. Ecol. 43:75–87.

Felske, A., H. Rheims, A. Wolterink, E. Stackebrandt, and A.D.L Akkermans. 1999. Ribosome analysis reveals prominent activity of an uncultured member of the class Actinobacteria in grassland soils. Microbiology 143:2983–2989.

Felske, A., A. Wolterink, R. VanLis, and A.D.L. Akkermans. 1998. Phylogeny of the main bacterial 16S rRNA sequences in Drentse A grassland soils (the Netherlands). Appl. Environ. Microbiol. 64:871–879.

Franklin, R.B., and A.L. Mills. 2003. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. FEMS Microbiol. Ecol. 44:335–346.

Gattinger, A., R. Ruser, M. Schloter, and J.C. Munch. 2002. Microbial community structure varies in different soil zones of a potato field. J. Plant Nutr. Soil Sci. 165:421–428.

Gupta, R.S. 2000. The phylogeny of proteobacteria: Relationships to other eubacterial phyla and eukaryotes. FEMS Microbiol. Rev. 24:367–402.

Gupta, R.S. 2004. The phylogeny and signature sequences characteristics of Fibrobacteres, Chlorobi, and Bacteroidetes. Crit. Rev. Microbiol. 30:123–143.

Hartmann, A., R. Pukall, M. Rothballer, S. Gantner, S. Metz, M. Schloter, and B. Mogge. 2005. Microbial community analysis in the rhizosphere by in situ and ex situ application of molecular probing, biomarker and cultivation techniques, p. 449–469. *In* D. Werner (ed.) The rhizosphere. Springer Publications, Heidelberg, Germany.

Heuer, H., and K. Smalla. 2007. Horizontal gene transfer between bacteria. Environ. Biosaf. Res. 6:3–13.

Hoffmann, A., T. Thimm, M. Droge, E.R.B. Moore, J.C. Munch, and C.C. Tebbe. 1998. Intergeneric transfer of conjugative and mobilizable plasmids harbored by *Escherichia coli* in the gut of the soil microarthropod *Folsomia candida* (Collembola). Appl. Environ. Microbiol. 64:2652–2659.

Holmes, A.J., J. Bowyer, H.P. Holley, M. Montgomery, and M.R. Gillings. 2000. Diverse, yet to be-cultured members of the Rubrobacter subdivision of the Actinobacteria are widespread in Australian arid soils. FEMS Microbiol. Ecol. 33:111–120.

Killham, K., and C. Yeomans. 2001. Rhizosphere carbon flow measurement and implications: From isotopes to reporter genes. Plant Soil 232:91–96.

Kowalchuk, G.A., F.A. De Souza, and J.A. van Veen. 2002. Community analysis of arbuscular mycorrhizal fungi associated with *Ammophila arenaria* in Dutch coastal sand dunes. Mol. Ecol. 11:571–581.

Leininger, S., T. Urich, M. Schloter, L. Schwark, J. Qi, G.W. Nicol, J.I. Prosser, S.C. Schuster, and C. Schleper. 2006. Archaea predominate among ammonium oxidizing prokaryotes in soil. Nature 442:806–809.

Liu, W.T., T.L. Marsh, H. Cheng, and L.J. Forney. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. 63:4516–4522.

Ludwig, W., S.H. Bauer, M. Baue, I. Held, G. Kirchhof, R. Schulze, I. Huber, S. Spring, A. Hartmann, and K.H. Schleifer. 1997. Detection and in situ identification of representatives of a widely distributed new bacterial phylum. FEMS Microbiol. Lett. 153:181–190.

Ludwig, W., O. Strunk, R. Westram et al. 2004. ARB: A software environment for sequence data. Nucleic Acids Res. 32:1363–1371.

Mrkonjic, M., M. Engel, A. Gattinger, A. Embacher, U. Bausenwein, M. Sommer, J.C. Munch, and M. Schloter. 2008. Spatial and temporal variability of soil proteolytic bacteria and their activities in an arable field. Soil Biol. Biochem. 40:1646–1653.

Muyzer, G., T. Brinkhoff, U. Nübel, C. Santegoeds, H. Schäfer, and C. Wawer. 1998. Denaturing gradient gel electrophoresis (DGGE) in microbial ecology, p. 1–27. *In* A.D.L. Akkermans, J.D. van Elsas, and F.J. de Bruijn (eds.) Molecular microbial ecology manual. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Nannipieri, P. 2006. Role of stabilized enzymes in microbial ecology and enzyme extraction from soil with potential applications in soil proteomics, p. 75–94. *In* P. Nannipieri and K. Smalla (eds.) Soil biology. Vol. 8: Nucleic acids and proteins. Springer Verlag, Heidelberg, Germany.

Ogunseitan, O. 2005. Microbial diversity. Blackwell Publishing, Malden, MA.

Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. Science 276:734–740.

Perez de Mora, A., P. Jaeckel, and M. Schloter. 2006. Effects of plant growth and different amendments on microbial community structure and function in heavy metal contaminated soils. Soil Biol. Biochem. 38:327–341.

Podar, M., C.B. Abulencia, M. Walcher, D. Hutchison, K. Zemgler, J.A. Garcia, T. Holland, D. Cotton, L. Hauser, and M. Keller. 2007. Targeted access to the genomes of low-abundance organisms in complex microbial communities. Appl. Environ. Microbiol. 73:3205–3214.

Raaijmakers, J.M., D.M. Weller, and L.S. Thomashow. 1997. Frequency of antibiotic-producing *Pseudomonas* spp. in natural environments. Appl. Environ. Microbiol. 63:881–887.

Rappe, M., and S.J. Giovannoni. 2003. The uncultured microbial majority. Ann. Rev. Microbiol. 57:369–394.

Rediers, H., P.B. Rainey, J. Vanderleyden, and R. De Mot. 2005. Unrevealing the secret lives of bacteria. Microbiol. Mol. Biol. Rev. 69:217–261.

Ritz, K., W. McNicol, N. Nunan, S. Grayston, P. Millard, D. Atkinson, A. Gollotte et al. 2004. Spatial structure in soil chemical and microbiological properties in an upland grassland. FEMS Microbiol. Ecol. 49:191–205.

Roberts, P., and D.L. Jones. 2008. Critical evaluation of methods for determining total protein in soil solutions. Soil Biol. Biochem. 40:1485–1495.

Schloter, M., M. Lebuhn, T. Heulin, and A. Hartmann. 2000. Ecology and evolution of bacterial microdiversity. FEMS Microbiol. Rev. 24:647–660.

Sharma, S., M. Aneja, J.C. Munch, and M. Schloter. 2007a. mRNA approaches to study microbial functionality in soil, p. 337–352. *In* D.J. van Elsas, J.T. Trevors, and E.M.H. Wellington (eds.) Modern soil microbiology. Marcel Dekker, New York.

Sharma, S., H. Karl, and M. Schloter. 2006. Genomic analyses of microbial processes in biogeochemical cycles, p. 14–41. *In* J.E. Cooper and J.R. Rao (eds.) Molecular techniques for soil and rhizosphere microorganisms. CABI Publisher, New Delhi, India.

Sharma, S., V. Radl, B. Hai, K. Kloos, M. Mrkonjic, and M. Schloter. 2007b. Quantification of functional genes in soil by PCR. J. Microbiol. Methods 68:445–452.

Smit, E., A. Wolters, and J.D. van Elsas. 1998. Self-transmissible mercury resistance plasmids with gene-mobilizing capacity in soil bacterial populations: Influence of wheat roots and mercury addition. Appl. Environ. Microbiol. 64:1210–1219.

Springael, D., and E. Top. 2004. Horizontal gene transfer and microbial adaptation to xenobiotics: Newtypes of mobile genetic elements and lessons from ecological studies. Trends Microbiol. 12:53–58.

Tebbe, C., and M. Schloter. 2006. Discerning the diversity of prokaryotes and their impact on agriculture, p. 81–100. *In* G. Benkiser (ed.) Biodiversity in agroecosystems. Springer Verlag, Heidelberg, Germany.

Teplitski, M., W.D. Bauer, and J.B. Robinson. 2000. Plants secrete substances that mimic bacterial *N*-acyl homoserine lactone signal activities and affect population density-dependent behaviours in associated bacteria. Mol. Plant-Microbe Interact. 13:637–648.

Thakuria, D., O. Schmidt, M.M. Siurtain, D. Egan, and F.M. Doohan. 2008. Importance of DNA quality in comparative soil microbial community structure analysis. Soil Biol. Biochem. 40:1390–1403.

Thimm, T., A. Hoffmann, I. Fritz, and C.C. Tebbe. 2001. Contribution of the earthworm *Lumbricus rubellus* (Annelida, Oligochaeta) to the establishment of plasmids in soil bacterial communities. Microb. Ecol. 41:341–351.

Torsvik, V.L. 1980. Isolation of bacterial DNA from soil. Soil Biol. Biochem. 12:15–21.

Torsvik, V., J. Goksoyr, and F.L. Daae. 1990. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. 56:782–787.

Torsvik, V., and O. Øvreås. 2002. Microbial diversity and function in soil: From genes to ecosystems. Curr. Opin. Microbiol. 5:240–245.

Ulrich, A., and S. Wirth. 1999. Phylogenetic diversity and population densities of culturable cellulolytic soil bacteria across an agricultural encatchment. Microb. Ecol. 37:238–247.

Urich, T., A. Lanzen, J. Qi, D. Huson, C. Schleper, and S. Schuster. 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLoS One 3:e2527.

Varm, A., B. Krishna Kolli, J. Paul, S. Saxena, and H. König. 1994. Lignocellulose degradation by microorganisms from termite hills and termite guts: A survey on the present state of art. FEMS Microbiol. Rev. 15:9–28.

Vieira, F., and E. Nahas. 2005. Comparison of microbial numbers in soils by using various culture media and temperatures. Microbiol. Res. 160:197–202.

Wang, J., C. Jenkins, R.I. Webb, and J.A. Fuerst. 2002. Isolation of Gemmata-like and Isosphaera-like planctomycete bacteria from soil and freshwater. Appl. Environ. Microbiol. 68:417–422.

Wardle, D.A. 2005. How plant communities influence decomposer communities, p. 119–138. *In* R.A. Bardgett, M.B. Usher, and D.W. Hopkins (eds.) Biological diversity and function in soils. Cambridge University Press, Cambridge, U.K.

Webster, G., T.M. Embley, and J.I. Prosser. 2002. Grassland management regimens reduce small-scale heterogeneity and species diversity of beta-proteobacterial ammonia oxidizer populations. Appl. Environ. Microbiol. 68:20–30.

Werner, D. 1992. Symbiosis of plants and microbes. Chapman and Hall, London, U.K.

Whitman, W.B., D.C. Coleman, and W. Wiebe. 1998. Prokaryotes: The unseen majority. Proc. Natl. Acad. Sci. USA 95:6578–6583.

Woese, C.R. 1994. There must be a prokaryote somewhere: Microbiology's search for itself. Microbiol. Rev. 58:1–9.

Zavarzin, G.A., E. Stackebrandt, and R.G. Murray. 1991. A correlation of phylogenetic diversity in the Proteobacteria with the influences of ecological forces. Can. J. Microbiol. 37:1–6.

Zehr, J.P., B.D. Jenkins, S.M. Short, and G.F. Steward. 2003. Nitrogenase gene diversity and microbial community structure: A cross-system comparison. Environ. Microbiol. 5:539–554.

Zelles, L. 1997. Phospholipid fatty acid profiles in selected members of soil microbial communities. Chemosphere 35:275–294.

24.3 Soil Fungi

R. Greg Thorn

24.3.1 Introduction

Soil is all of the unconsolidated mineral and organic materials on the ground surface that is differentiated from the parent materials below and the unincorporated and undigested litter above by the biological activities of fungi, bacteria, invertebrates, and other soil organisms. This definition includes the L, F, H, and O organic horizons and the mineral horizons from A to C (Fanning and Balluff, 1989; Bridges, 1997) and could be expanded to include the aerial "soil" that forms on the surfaces of trees that are covered with mosses, lichens, and vascular plant epiphytes in humid regions (Nadkarni et al., 2004; Enloe et al., 2006). The nature of the inorganic parent material and the activities of plants and soil organisms together make a soil what it is. The dominant soil

organisms, both in terms of processes and biomass, are the fungi (de Boer et al., 2005; Joergensen and Wichern, 2008). It is difficult to overemphasize the importance of fungi in soil. In many soils, the biomass of fungi exceeds that of all other soil organisms combined (excluding plant roots) by a factor of 10–1. In soil, as elsewhere, the primary and best known role of fungi is the decomposition and mineralization of complex, recalcitrant compounds of plant and animal origins, such as cellulose, hemicellulose, lignin, and chitin (Rayner and Boddy, 1988; Schwarze et al., 2004; Boddy et al., 2007). Other major roles of soil fungi include their involvements in beneficial and detrimental symbioses with plant roots. These range from the mutually beneficial mycorrhizae that enable land plants to exist in the face of nutrient and water limitations and other stresses to the plant pathogens that annually cause billions of dollars in losses to world crops (Agrios, 2004; Smith and Read, 2008; Section 24.4). Less well known, but possibly of great importance in the functioning of soil ecosystems, are the many connections made by fungi between virtually all soil organisms, at all trophic levels (Thorn, 1997).

The purpose of this chapter is to provide a brief introduction to the modern classification of those organisms that are fungi or resemble true fungi, what these organisms do, and modern methods for the analysis of their presence and activity. A selection of more detailed references is included for particular aspects.

24.3.2 The Diversity of Fungi and Fungus-Like Organisms and Their Roles in Soil

Although the strict definition of mycology is the study of fungi, many of the organisms that mycologists have studied are outside of the Kingdom Fungi (Figure 24.7). These organisms are united by a suite of fungus-like characteristics: they are non-photosynthetic chemoheterotrophs that reproduce by spores. Many have extracellular digestion and absorptive nutrition and most have a filamentous growth form. This broad definition has included the filamentous bacteria (actinomycetes), slime molds that are more closely related to the amoebae, "water molds" that prove to be achlorophyllous heterokont algae, and true members of the Kingdom Fungi. True fungi are the closest relatives of the Animal Kingdom (not of plants), and fungi and animals are derived from the slime molds. Fungi can be divided among five phyla: the Chytridiomycota, Zygomycota, Glomeromycota, Ascomycota, and Basidiomycota, although the Chytridiomycota and Zygomycota as currently defined are not monophyletic groups (Schüßler et al., 2001; Hibbett et al., 2007).

24.3.2.1 Actinomycetes

These prokaryotic organisms, although sometimes called the "ray fungi," are bacteria (=Eubacteria), not fungi. They form a part of the high GC subdivision of the Gram-positive lineage, along with nonfilamentous, important soil bacteria such as *Arthrobacter* and human and animal pathogens such as *Mycobacterium* (Woese, 1987; Embley and Stackebrandt, 1994; Anderson and Wellington, 2001; Zhi et al., 2009). However, as the trivial name actinomycetes suggests (the ending-mycetes
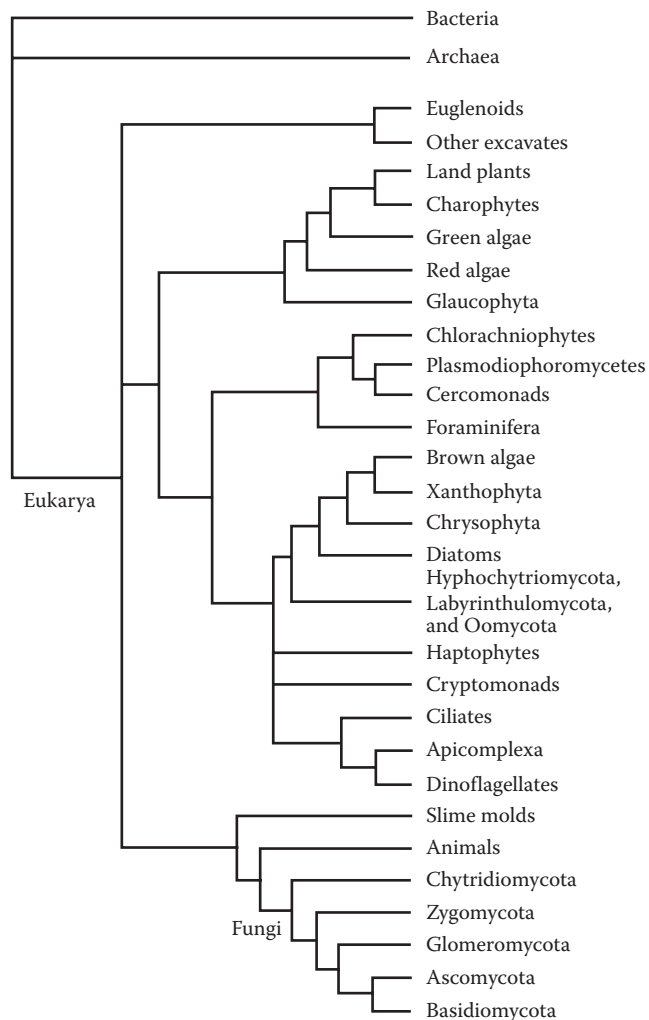
FIGURE 24.7 A phylogenetic tree indicating the place of various groups of fungi and fungus-like organisms (indicated in bold font), based on Archibald and Keeling (2004), Lutzoni et al. (2004), and Keeling et al. (2005). Many groups of organisms are left out (e.g., actinomycetes, not shown, are members of domain Bacteria), branch lengths are not proportional to sequence divergence, and the tree is not rooted. The relationships of the three domains Archaea, Bacteria, and Eukarya (Wheelis et al., 1992) are not resolved.

refers to fungi), they possess a number of fungus-like characteristics, including the hyphal (tubular) growth form, the production of extracellular enzymes (including cellulases but not lignin peroxidases; Ball and Trigo, 1997; Mason et al., 2001), and absorptive nutrition. These characteristics combine to make them important decomposers of organic particles in soil (Kirby, 2006; Elfstrand et al., 2008; Jayasinghe and Parkinson, 2008). Certain thermophilic actinomycetes are predominant in high-temperature composts (40°C–60°C), and are important in the formation of composted substrate for commercial cultivation of the button mushroom, *Agaricus brunnescens* (Fermor and Grant, 1985; Savoie, 1998). Like the Gram-negative rhizobia and selected cyanobacteria, actinomycetes in the genus *Frankia* form nitrogen-fixing symbioses with the roots of various woody plants (Jeong et al., 1999; Wall, 2000; Vessey et al., 2004; Bothe et al., 2007). Nitrogen fixation by fungi or other eukaryotes have not been convincingly demonstrated (Postgate, 1998). Saprotrophic actinomycetes are abundant in soil (Williams et al., 1984) and produce the compound geosmin, which is responsible for the characteristic odor of freshly tilled soil (Gerber and Lechevalier, 1965). Actinomycetes, particularly *Streptomyces* spp., are commonly recovered during the isolation of soil fungi but are readily recognized by their tufted growth form and extremely fine (<0.5 μm) hyphae. Media for the selective isolation of actinomycetes frequently make use of the antifungal antibiotic cycloheximide (actidione), and antibacterial antibiotics to which actinomycetes are tolerant, and may contain colloidal chitin as the sole carbon and nitrogen source (Goodfellow and Cross, 1974). There is no realistic estimate available of the number of species of actinomycetes inhabiting soil. However, over 670 described species of the important genera *Actinomyces*, *Frankia*, *Geodermatophilus*, *Streptomyces*, *Streptosporangium*, *Streptoverticillium*, and *Thermoactinomyces* are accepted in the *List of Prokaryotic names with Standing in Nomenclature* (Euzéby, 2009; http://www.bacterio.cict.fr/), and this is likely to be a tremendous underestimate of the real diversity of these organisms in soil (cf. Stach et al., 2003).

24.3.2.2 Protozoan "Fungi"

The fungus-like protists include the slime molds and their relatives in the phylum Mycetozoa (myxomycetes or myxogastria, dictyostelids, and protostelids), plus plasmodiophorids of the phylum Cercozoa, and acrasids of the phylum Percolozoa (Alexopoulos et al., 1996; Archibald and Keeling, 2004; Kirk et al., 2008). The phylum name Mycetozoa (fungus animals) refers to their mix of fungus-like and animal-like characters: reproduction by spores and ingestion of food by phagocytosis. The acrasids, dictyostelids, protostelids, and myxomycetes occur in soil as amoebae, and the latter may be abundant in the root rhizosphere (Amewowor and Madelin, 1991). These amoeba-like cells may aggregate to form motile "slime stages" that are multicellular pseudoplasmodia (in dictyostelids) or multinucleate plasmodia (in myxomycetes) that continue to feed on bacteria, other microbes, and small organic particles; those of certain myxomycetes can be conspicuous on lawns and garden beds of wood-chip mulch (Martin and Alexopoulos, 1969; Raper, 1984; Stephenson and Stempen, 1994). The plasmodiophoromycetes (Phytomyxea) include important root pathogens of cultivated crucifers (cabbage and cole crops) and potatoes, and may also serve as vectors of plant–pathogenic viruses (Braselton, 2001). Owing to their abundance in soils and their lack of cell walls during their trophic phase, the protistan "fungi" may contribute a substantial proportion of the eukaryotic DNA recovered during lysis procedures for the isolation of bacterial DNA from soil. This group includes 1165 described species, many of which live in habitats other than soil, such as rotting wood, bark, and dung (Kirk et al., 2008).

24.3.2.3 Stramenopile "Fungi"

Fungus-like members of the Kingdom Stramenopila (or Chromista) include the Oomycota and Hyphochytriomycota, which have often gone by the name of "water molds" (Alexopoulos et al., 1996; Dick, 2001a, 2001b; Fuller, 2001). Like the actinomycetes, the oomycetes (or peronosporomycetes) have a number of fungus-like characteristics, including the hyphal growth form, the production of extracellular enzymes, and absorptive nutrition (Dick, 2001b). Unlike most true fungi, members of both the oomycota and hyphochytriomycota have motile stages known as zoospores. These are powered by two different types of flagella, the whiplash and the tinsel flagellum, in common with the motile stages of chromistan algae (the diatoms and the golden and brown algae). The hyphochytriomycota are a small group known primarily as parasites of oomycota, but the oomycota are a fairly large group that include the important soilborne plant pathogens *Pythium* and *Phytophthora* (Buczacki, 1983; Martin, 1992; Erwin and Ribeiro, 1996; Arcate et al., 2006). There are approximately 1000 described species of stramenopile "fungi" of which perhaps half occur in freshwater and marine habitats (Kirk et al., 2008).

24.3.2.4 Kingdom Fungi, Phylum Chytridiomycota

Soil chytrids are a relatively small group, including plant pathogens such as *Synchytrium* and *Olpidium*, nematode parasites such as *Catenaria*, parasites of algae such as *Chytridium*, and saprobes such as *Allomyces* and *Chytriomyces* (Barr, 2001; James et al., 2006b). The soil chytrids possess motile zoospores powered by a single, posterior flagellum, and are thought to be close to the common ancestor of fungi and animals. Phylogenetic analyses that have included more than one representative of each of the chytrids and zygomycetes have indicated that the chytrids may form a monophyletic group, splitting the zygomycetes (Van de Peer and De Wachter, 1997) or that both groups may be polyphyletic and interspersed one among the other (Nagahama et al., 1995; James et al., 2006b). Together, the chytrids and zygomycetes differ considerably from the Ascomycota and Basidiomycota in that they are typically coenocytic, with multiple genetically different nuclei per cell, whereas the Ascomycota and Basidiomycota typically have uninucleate primary mycelia and binucleate (dikaryotic) secondary mycelia. Many of the approximately 700 described species of chytrids (Kirk et al., 2008) occur in freshwater or marine environments, and I know of no estimate of the number of species occurring in soil.

24.3.2.5 Kingdom Fungi, Phylum Zygomycota

Soil-inhabiting members of the Zygomycota belong in two groups of very different morphology and ecology: the predominantly saprobic zygomycetes (Mucorales and relatives) and the Zoopagales, Kickxellales, and Entomophthorales, which are parasitic on insects and other invertebrates, fungi, and humans (O'Donnell, 1979; Humber, 1989; Alexopoulos et al., 1996; Benny et al., 2001; White et al., 2006; Kirk et al., 2008). A third group formerly treated as the Glomerales (misspelled Glomales) has since been shown to represent an independent phylum, Glomeromycota (Schüßler et al., 2001; see later). As mentioned above, recent phylogenetic analyses that have included multiple examples of both chytrids and Zygomycota suggest that neither forms a monophyletic group. The saprobic and mycoparasitic zygomycetes (Mucorales and relatives) are more basal and cluster together with some or all of the chytrids, while the Glomeromycota appear to be the sister group to the "higher fungi" (subphylum Dikarya, Hibbett et al., 2007), the Ascomycota and Basidiomycota (Nagahama et al., 1995; James et al., 2006a; Figure 24.7). However, all of these groups share general morphological features such as broad, tubular, thin-walled hyphae, the coenocytic condition, and positive (red) staining of their hyphal walls with diazonium blue B (Summerbell, 1985; Prillinger et al., 2002).

In the earlier literature, the saprobic Mucorales were often referred to as "sugar fungi." The accompanying view held that these organisms are rapid colonists making use of the soluble sugars before secondary colonists become established during the succession of different fungi on rich and freshly deposited substrates such as dung and fallen fruits (Garrett, 1963). In reality, the rapid appearance of Mucorales on such substrates reflects their rapid rates of growth and sporulation; in many cases, the "secondary colonists" are already there and digesting the more recalcitrant substrates but are slower to sporulate and come to our notice (Pugh, 1974; Cooke and Rayner, 1984). Still, the term "sugar fungi" accurately portrays the typical growth substrate of these fungi. They are not noted for their extracellular degradative enzymes and may frequently depend on the breakdown products provided by lignocellulose-degrading ascomycetes or basidiomycetes (Cooke and Rayner, 1984). From here, it may have been a short step to direct, biotrophic parasitism of the decomposer fungi, a nutritional mode found among some Mucorales and related groups (O'Donnell, 1979; Benny et al., 2001). Of the 1065 described species of Zygomycota (including Entomophthorales, Kickxellales, Mucorales, and Zoopagales) accepted in the *Dictionary of the Fungi* (Kirk et al., 2008), the majority might be found in soil or close to soil in various organic substrates.

24.3.2.6 Kingdom Fungi, Phylum Glomeromycota

The phylum Glomeromycota contains a single order, Glomerales (formerly misspelled Glomales and treated within the Zygomycota) (Morton, 1988; Morton and Benny, 1990; Schüßler et al., 2001; Redecker and Raab, 2006; http://www.lrz-muenchen.de/~schuessler/amphylo/). It appears that this lineage goes back at least to the early Cambrian, when they may have been symbiotic with cyanobacteria and, later, with the ancestors of modern land plants (Pirozynski and Malloch, 1975). These are arguably among the most important of all of the fungi: Approximately 80% of the world's land plants depend upon mycorrhizal symbioses with these fungi for survival in soils that are nutrient-limiting or prone to drought or other stresses (Malloch et al., 1980; van der Heijden et al., 1998; Selosse et al., 2004; de Boer et al., 2005; Smith and Read, 2008; Section 24.4). As in chytrids

and zygomycetes, their hyphae are coenocytic, and recent analyses have shown that many of the nuclei in a single spore may be genetically distinct (Jansa et al., 2002; Sanders, 2002). *Geosiphon*, a sac-like soil organism that forms an endosymbiosis with a photosynthetic cyanobacterium (the cyanobacterium is inside the *Geosiphon*) is a basal member of the Glomeromycota (Gehrig et al., 1996; Schüßler, 2002). This endosymbiosis is most likely very old: the origin of the Glomeromycota is thought to be some 350–550 million years ago based on fossil evidence (Pirozynski, 1981; Pirozynski and Dalpé, 1989) or DNA sequence divergence (Berbee and Taylor, 1993; Simon et al., 1993; Taylor and Berbee, 2006), and the cyanobacteria date back to 2.9–3.4 billion years before present (Schopf, 1993). Approximately 215 species of Glomeromycota in 14 genera are known (http://www.lrz-muenchen.de/~schuessler/amphylo/), and their ability to form symbioses with a far greater diversity of land plants remains an interesting puzzle. It is possible that the complex nuclear makeup of individuals in this group provides the genetic variation required to achieve symbioses with such a broad range of host plants and in a broad range of habitats. However, some studies have shown that there is more host or habitat specificity than has generally been recognized, perhaps a function of identifications based on their very limited morphology (Lovelock et al., 2003; Vandenkoornhuyse et al., 2003; Sykorová et al., 2007; Appoloni et al., 2008). If true, appropriate species of these fungi could be vital in survival or restoration of rare or endangered vascular plant species, and the community of arbuscular mycorrhizal fungi determines plant community diversity and productivity (van der Heijden et al., 1998).

24.3.2.7 Kingdom Fungi, Phylum Ascomycota (Including Most Yeasts and Most Asexual Forms Previously Included in the "Deuteromycetes")

The Ascomycota include species of diverse nutritional modes, from saprotrophs through plant and animal pathogens, mycoparasites, and mycorrhizal associates. The Ascomycota include fungi that reproduce sexually through the production of meiotic spores in a cell called an ascus (plural asci), as well as their phylogenetic relatives that are only known to reproduce asexually or rarely sexually. The inclusion of asexual relatives may seem obvious to the nonmycological reader, but the tradition in mycology has been to recognize separate taxa for the sexual and asexual forms (e.g., Webster, 1981; Hawksworth et al., 1983). Filamentous asexual forms have been called imperfect fungi, fungi imperfecti, deuteromycetes, hyphomycetes, or molds. There have been separate names and classifications all the way from genus and species (e.g., asexual *Aspergillus nidulans* versus sexual *Emericella nidulans*) up to phylum ("Deuteromycota" versus Ascomycota, Margulis and Schwartz, 1988). The vast majority of deuteromycete species have not been connected with any sexual state (indeed, it is thought that many may not have a sexual state) (Seifert and Gams, 2001; Domsch et al., 2007). However, there are a number of morphological, physiological, and molecular criteria by which these fungi can be recognized as belonging to

the Ascomycota (see later), so there seems little justification for retaining separate, artificial classifications above the level of the genus. *Aspergillus*, *Cladosporium*, *Fusarium*, and *Penicillium* will remain familiar genera of soil fungi, even if some species are connected with sexual states that have different names. The Ascomycota also include the majority of species of the yeasts, predominantly cellular (nonfilamentous) organisms that usually reproduce asexually by budding or fission (Kurtzman and Fell, 1998; Barnett et al., 2000). In yeasts, too, species known only to reproduce asexually are barred from genera typified by species that reproduce sexually. The genus *Candida* consists of a polyphyletic assemblage of asexual species of diverse affinities. However, the cellular growth form of yeasts is an adaptation to life in moist or liquid environments, and some members of the Basidiomycota have independently developed the yeast life form (see later). Thus, yeasts and molds are artificial categories of fungi, useful to describe their appearance and ecology but not their classification.

A particular disadvantage of this system has been that a single genus name used for asexual forms (e.g., *Sclerotium* or *Rhizoctonia*) may include species of both Ascomycota and Basidiomycota: two phyla in one genus. Likewise, maintenance of a separate class Deuteromycetes has hidden the fact that we do not know the correct phylum for many of the species included. For instance, the economically important fungus causing cotton root rot, *Phymatotrichopsis omnivora*, is only known to reproduce asexually. Despite ultrastructural and physiological evidence to the contrary, it was widely reported as an asexual member of the Basidiomycota until DNA sequence analyses placed it in the family Rhizinaceae (Pezizales) of the Ascomycota (Marek et al., 2009). With new knowledge from ultrastructural, physiological, and molecular studies, this situation is gradually improving; the genus *Rhizoctonia* has recently been subdivided along phylogenetic lines (Roberts, 1999). Recent molecular phylogenetic analyses have included the asexual forms among their sexually reproducing relatives and have greatly improved our level of knowledge of the broad relationships within the Ascomycota (Blackwell et al., 2007; see http://www.fieldmuseum.org/myconet/outline.asp).

The Ascomycota and their asexual relatives are certainly the largest group of fungi, with over 64,000 described species (Kirk et al., 2008). About 20,000 described species form lichens: stable symbiotic associations between fungi and green algae or cyanobacteria, in which the fungal partner (mycobiont) forms a characteristic structure called a thallus that encloses and protects the alga or cyanobacterium (photobiont) (Hawksworth, 1988; Ahmadjian, 1993; Grube and Hawksworth, 2007). As with mycorrhizae, the lichen habit or symbiosis has multiple separate evolutionary origins amongst Ascomycota and Basidiomycota. These separate origins, long recognized on the basis of morphology, are also "suggested" by molecular phylogenetic analysis (Gargas et al., 1995). Lichens are important in soil formation by the colonization and physicochemical degradation of bare rocks (Hale, 1983; Nash, 2008) and form part of the cryptogamic crusts stabilizing and providing nutrients to desert soils (West, 1990;

Johansen, 1993; Belnap et al., 2001), but they are not normally considered soil fungi. Many other species of Ascomycota are parasites or endophytes of aerial plant parts, are human or animal pathogens, or inhabit dung or other specialized substrates. Some of these have a soilborne stage or a significant reservoir in soil. Perhaps 15,000 or more described species of Ascomycota might be found in soil or in litter and other organic substrates on soil.

Persons isolating fungi from substrates such as soil, rotting wood or litter, or diseased plants or animals frequently recover cultures of morphologically undistinguished fungi that lack the sexual reproductive structures necessary to identify them as members of the Ascomycota or Basidiomycota. These fungi may sporulate asexually but belong to a group that has never been connected with any sexual state, or may be nonsporulating in pure culture. The latter, "sterile" forms are often referred to in the literature as sterile white mycelium, sterile dark mycelium, or *mycelium radicis atrovirens* (Latin for greenish black mycelium from roots). These names are unhelpful, in terms of physiological, taxonomic, and ecological precision, and should no longer be necessary or accepted in publication, given the many tools with which we can now quite simply make a more accurate identification.

Hyphae of Ascomycota have simple septal pores; that is, their septa thin to a circular pore that provides a cytoplasmic connection between adjacent cells. Almost all members of the Basidiomycota that grow in culture as filamentous colonies (not yeasts) have what are called dolipore septa, a donut-like thickening around the septal pore that may be surrounded by a perforate or imperforate, membranous septal pore cap (Wells, 1978, 1994; Moore, 1985). Some Basidiomycota also form clamp connections (looping connections across septa) in their dikaryotic mycelia, and these are diagnostic if present. For those fungi without clamp connections, the dolipore septum or lack of it can usually be seen with a good light microscope in light-colored hyphae more than 1 μm in diameter if they are stained with trypan blue (Sneh et al., 1991) or with Congo red in an aqueous solution of ammonia or sodium dodecyl sulfate (Nawawi et al., 1977; Clemençon, 1998). Laser confocal microscopy should also provide a good method for detection of dolipore or simple septal pores in living or fixed materials, particularly if cell walls are stained with a fluorescent dye such as calcofluor white. Dolipore septa and simple septal pores are readily identified with transmission electron microscopy (TEM) in appropriate sections. If material is fixed, sectioned, and examined with TEM, Ascomycota can usually be distinguished from Basidiomycota even if septal pores are not seen in the section or sections examined. Most Ascomycota have bilayered hyphal walls, with a narrow, electron-dense outer layer and a broad, pale inner layer. In contrast, most Basidiomycota have multilamellate hyphal walls, with multiple narrow light and dark layers as seen in TEM (Bracker, 1967; Alexopoulos et al., 1996). The hyphal walls of Ascomycota and Basidiomycota (especially yeasts, and less dependably filamentous forms) also differ in their reaction to staining with diazonium blue B. The hyphal walls of Basidiomycota stain red or

purplish (a pretreatment with KOH is necessary for filamentous forms), whereas the walls of most Ascomycota are nonstaining (Hagler and Ahearn, 1981; Summerbell, 1985). As mentioned above, hyphae of Glomeromycota and Zygomycota also stain with diazonium blue B, but these can easily be distinguished from members of the Basidiomycota by their large diameter and coenocytic nature. Physiologically, cultures can often be determined as Ascomycota or Basidiomycota by their pattern of tolerance or sensitivity to sodium chloride, benomyl, and cycloheximide (Hutchison, 1990). Finally, great progress has been made in the extraction and sequencing of fungal DNA, enabling the rapid and certain identification—at least at the phylum level—of an unknown strain by comparison with available sequences of known strains (e.g., Bruns et al., 1991; Marek et al., 2009).

24.3.2.8 Kingdom Fungi, Phylum Basidiomycota (Including Some Yeasts and Some Asexual Forms Previously Included in the "Deuteromycetes")

The Basidiomycota include two of the most important groups of soil fungi: species forming ectomycorrhizae of forest trees (Smith and Read, 2008; Section 24.4), and species that are the dominant decay organisms of plant polymers such as lignin, hemicellulose, and cellulose (Rayner and Boddy, 1988; Boddy et al., 2007). A few basidiomycetes, such as *Rhizoctonia solani*, are very important plant pathogens of field and forage crops ranging from wheat to rice to alfalfa. Others are root pathogens of tropical tree crops such as rubber and oil palms, or are root pathogens or cause rot and butt rots of tropical and temperate forest trees and turf (Ploetz et al., 1994; Roberts, 1999; Smiley et al., 2005). The Ascomycota and their asexual relatives are usually thought to be the major group of soil fungi, both in number of species and in biomass. This may be true in agricultural soils and in native tropical soils, but in many temperate and boreal forest soils, the basidiomycetes probably outnumber and outweigh the ascomycetes in at least the organic-rich soil horizons (Frankland, 1982; Entry et al., 1992). Many form extensive mycelia with considerable biomass. The most famous of these is *Armillaria*, the "humongous fungus" of which one individual mycelium was calculated to occupy 15 ha and have a weight of 10 Mg (Smith et al., 1992); another individual *Armillaria* thought to occupy 40 ha has been reported (Shaw and Roth, 1976). Clearly, the Basidiomycota include the fungi with the best ligninolytic abilities (Orth et al., 1993; Worrall et al., 1997; Baldrian, 2008) and are therefore likely to be the most important in decomposition of plant litter in most soils. Approximately 10,000 of the over 31,000 described species of Basidiomycota may be found in soil or litter (Kirk et al., 2008).

As with the situation in the Ascomycota, a number of asexual or predominantly asexual fungi belong in the Basidiomycota. Their natural relationships have been hidden (perhaps especially to nonmycologists) by their inclusion in the artificial class "Deuteromycetes." Formerly orphaned fungi include "molds" such as the *Sporotrichum* (=*Chrysosporium*) state of

Phanerochaete chrysosporium and many yeasts frequently encountered in soil, such as *Rhodotorula*. Recent molecular phylogenetic analyses of Basidiomycota (summarized in, e.g., Hibbett and Thorn, 2001; Hibbett et al., 2007) have greatly improved our knowledge of the relationships of these fungi and provide a framework into which the asexual forms and unidentified isolates from soil or mycorrhizae may be placed and identified (Bruns et al., 1998; Kõljalg et al., 2005; http://unite.ut.ee/).

Despite their numbers and biomass in soil, the Basidiomycota are rarely recovered in surveys of soil fungi based on isolation into culture, because few produce significant numbers of propagules in soil in the way that the asexual Ascomycota do. On isolation plates, the few hyphal propagules of Basidiomycota are tremendously outnumbered by mold spores and outcompeted by the more rapidly growing mold fungi. Because of this, the Basidiomycota remain the "missing link" in soil mycology (Chesters, 1949), 60 years after Chesters first made that claim. In order to overcome our inability to routinely isolate decomposer basidiomycetes from soil, Thorn et al. (1996) developed a method of plating washed organic particles from soil on a medium made selective for basidiomycetes with benomyl (Edgington et al., 1971; Worrall, 1991; Johnson, 1994) and with guaiacol as an indicator of laccase(s) and peroxidase(s) (these enzymes convert colorless guaiacol to a deep red quinone). The addition of dicloran (2,6-dichloro-4-nitroaniline; no longer widely available) to the medium should provide control of Mucorales (which are tolerant to benomyl), for soils or litters in which these fungi are a problem (Worrall, 1991); hymexazol (3-hydroxy-5-methylisoxazole) would be an alternative (Tsao and Guy, 1977). The use of these or other semiselective inhibitors in soil (for instance, to assay the activities of particular taxonomic groups of fungi) is beset with complications that may lessen their usefulness, such as nontarget effects and quenching of the inhibitory effects of the selected compound by physical or chemical components of soil (Nannipieri et al., 2003).

24.3.3 Methods for the Characterization of Communities and Activities of Soil Fungi

Among the methods available for the characterization of communities and activities of soil fungi are direct counts (Bloem et al., 1995) and immunological detection (Frankland, 1984; Clausen, 1997), differential inhibition with antibacterial and antifungal biocides and measurement of CO_2 evolution or other indicators of physiological activity (Anderson and Domsch, 1975; Kjøller and Struwe, 1982; Bailey et al., 2003), chloroform fumigation and incubation or extraction methods for total microbial biomass (Harris et al., 1997), incorporation of ^{13}C-labeled substrates into biomarkers such as ergosterol (Boschker et al., 1998; Baath 2001; Neufeld et al., 2007), activity stains applied to fungi in soil smears (Tsuji et al., 1995; Li et al., 2004), assays on enzymes extracted from soil and litter (Wood, 1979; Savoie, 1998; Osono, 2007), isolation and quantitation of fungal biomarkers such as ergosterol or chitin-derived glucosamine (Swift, 1973;

Tunlid and White, 1992; Zhao et al., 2005), isolation of fungi into culture, and isolation and characterization of fungal DNA or RNA from soil. A brief overview of methods for isolating fungi from soil and isolation and characterization of soil fungal DNA and RNA are all that can be included here.

24.3.3.1 Methods for Isolating Fungi from Soil

Methods for the isolation into culture of various fungi from soil have recently been compiled and reviewed (see Mueller et al., 2004). Keys for the identification of common genera and species of soil fungi of temperate areas are available (e.g., Domsch et al., 2007), but identification of many isolates from tropical soils or from temperate or tropical litter must be sought in the large mycological literature, to which Kirk et al. (2008) provide an entry. In choosing a method for isolating fungi from soil, one should first define the objective; is it to obtain a complete list for one sample, or a suite of predominant species by which to compare two or more samples, or to recover a specific species or ecological group of interest? If it is to obtain a "complete" list, be prepared for a challenge: Christensen (1989), using just one method and medium, found that 8–15 previously unseen species of microfungi were added with every 100 incremental isolates past 1100. No one method, medium, or incubation temperature will recover all fungi in any sample (Carreiro and Koske, 1992). A number of sources discuss methods and media for the selective recovery of particular groups of fungi (see Mueller et al., 2004). The greatest difficulty with any of these methods is the inability to relate the species and numbers of isolates obtained to the identities and importance of fungi in the native soil.

24.3.3.2 Isolation and Characterization of Nucleic Acids from Soil

Methods for the isolation of DNA and RNA from soil have seen great improvements in recent years, with off-the-shelf kits yielding nucleic acids of good quality (e.g., suitable for PCR) much more rapidly and reliably than laborious techniques that were common in the 1990s (e.g., Holben et al., 1988). Most early studies of soil DNA were directed toward characterization of the soil bacterial community, but in recent years, the field has really opened wide: studies of fungal community membership (e.g., O'Brien et al., 2005; Lynch and Thorn, 2006) have been joined by studies detecting particular genes (Section 28.1) and gene expression (Section 28.2). Individual fungal taxa of interest (e.g., plant or human pathogens) can be detected by probing with taxon-specific probes (Bruns and Gardes, 1993; McArthur et al., 2001), amplification using taxon-specific primers (PCR-based detection assays; Gardes and Bruns, 1993; Lievens et al., 2006; van der Linde et al., 2008), or by using the extracted nucleic acids to probe taxonomic arrays of oligonucleotides bound on glass slides (microarrays) or nylon membranes (macroarrays) (e.g., Tambong et al., 2006). The relative biomass of selected fungi in soil may be estimated using real-time PCR (qPCR; e.g., Landeweert et al., 2003). Community analyses have been done by PCR amplification of rRNA or other genes with conserved fungal primers (or broadly specific primers, e.g., Redecker, 2000; Lynch and Thorn, 2006;

Appoloni et al., 2008) and subsequent cloning and sequencing of amplified products (e.g., Vandenkoornhuyse et al., 2002; Martin and Rygiewicz, 2005; O'Brien et al., 2005). Community analyses based on banding patterns (e.g., RFLP or DGGE) of amplified community DNA, without accompanying sequence data that allow identification of the community members, are not recommended because of the inability to compare between sites or studies. "Next generation" sequencing using 454 or Titanium pyrosequencing platforms (e.g., Roesch et al., 2007; Tringea and Hugenholtz, 2008) produces millions of short (30 to perhaps 200 bp) sequences from a sample, permitting much broader and semiquantitative analyses of fungal communities from soil, litter, or other complex substrates. These approaches allow us to increase the sample size of fungal communities investigated and may better answer the question of what proportion of the fungal community can be isolated into culture. The future of soil mycology just got more interesting, again.

24.3.4 Conclusions

Two facts became apparent as I revised this review, first written about 10 years ago. First, there have been tremendous changes in the classification (and, as a result, naming) of fungi and fungus-like organisms at all levels from their Kingdom placement to their species names. An outline of the higher classification of Fungi has been published (Hibbett et al., 2007), and these changes have been included in the latest edition of the *Dictionary of the Fungi* (Kirk et al., 2008). The latter also has separate sections for fungus-like organisms belonging in the Heterokonta or "Protozoa" (Kirk et al., 2008). It will be some time before these changes make it into textbooks of mycology, let alone those for plant pathology, botany, or general biology. However, as in all fields of biology, a classification that more closely reflects the evolutionary history of the organisms will be more predictive and useful than our historical classifications that were misled by morphological convergences.

Secondly, new molecular techniques are allowing a rapid acquisition of knowledge about the composition and functioning of complex fungal communities that was simply unattainable even recently. Applied critically, these new methods are rejuvenating mycology and bringing the fundamental importance of fungi to the attention of ecologists and soil biologists. It is time for fungi to be given their due.

References

Agrios, G.N. 2004. Plant pathology. 5th edn. Academic Press, Amsterdam, the Netherlands.

Ahmadjian, V. 1993. The lichen symbiosis. John Wiley, New York.

Alexopoulos, C.J., C.W. Mims, and M. Blackwell. 1996. Introductory mycology. 4th edn. John Wiley, New York.

Amewowor, D.H.A.K., and M.F. Madelin. 1991. Numbers of myxomycetes and associated microorganisms in the root zones of cabbage (*Brassica oleracea*) and broad bean (*Vicia faba*) in field plots. FEMS Microbiol. Lett. 86:69–82.

Anderson, J.P.E., and K.H. Domsch. 1975. Measurement of bacterial and fungal contributions to respiration of selected agricultural and forest soils. Can. J. Microbiol. 21:314–322.

Anderson, A.S., and E.M.H. Wellington. 2001. The taxonomy of *Streptomyces* and related genera. Int. J. Syst. Evol. Microbiol. 51:797–814.

Appoloni, S., Y. Lekberg, M.T. Tercek, C.A. Zabinski, and D. Redecker. 2008. Molecular community analysis of arbuscular mycorrhizal fungi in roots of geothermal soils in Yellowstone National Park (USA). Microb. Ecol. 56:649–659.

Arcate, J.M., M.A. Karp, and E.B. Nelson. 2006. Diversity of peronosporomycete (oomycete) communities associated with the rhizosphere of different plant species. Microb. Ecol. 51:36–50.

Archibald, J.M., and P.J. Keeling. 2004. Actin and ubiquitin protein sequences support a cercozoan/foraminiferan ancestry for the plasmodiophorid plant pathogens. J. Eukaryot. Microbiol. 51:113–118.

Baath, E. 2001. Estimation of fungal growth rates in soil using C-14-acetate incorporation into ergosterol. Soil Biol. Biochem. 33:2011–2018.

Bailey, V.L., J.L. Smith, and H. Bolton. 2003. Novel antibiotics as inhibitors for the selective respiratory inhibition method of measuring fungal:bacterial ratios in soil. Biol. Fertil. Soils 38:154–160.

Baldrian, P. 2008. Enzymes of saprotrophic basidiomycetes, p. 19–42. *In* L. Boddy, J.C. Frankland, and P. van West (eds.) Ecology of saprotrophic basidiomycetes. Academic Press, London, U.K.

Ball, A.S., and C. Trigo. 1997. The role of actinomycetes in plant litter decomposition. Recent Res. Dev. Soil Biol. Biochem. 1:9–20.

Barnett, J.A., R.W. Payne, and D. Yarrow. 2000. Yeasts: Characteristics and identification. 3rd edn. Cambridge University Press, Cambridge, U.K.

Barr, D.J.S. 2001. Chytridiomycota, p. 93–112. *In* D.J. McLaughlin, E.G. McLaughlin, and P.A. Lemke (eds.) The Mycota. Vol. VIIa. Systematics and evolution. Springer-Verlag, New York.

Belnap, J., J. Hilty Kaltenecker, R. Rosentreter, J. Williams, S. Leonard, and D. Eldridge. 2001. Biological soil crusts: Ecology and management. Technical Reference 1730-2. U.S. Department of the Interior, Denver, CO.

Benny, G.L., R.A. Humber, and J.B. Morton. 2001. Zygomycota: Zygomycetes, p. 113–146. *In* D.J. McLaughlin, E.G. McLaughlin, and P.A. Lemke (eds.) The Mycota. Vol. VIIa, Systematics and evolution. Springer-Verlag, Berlin, Germany.

Berbee, M.L., and J.W. Taylor. 1993. Dating the evolutionary radiations of the true fungi. Can. J. Bot. 71:1114–1127.

Blackwell, M., D.S. Hibbett, J.W. Taylor, and J.W. Spatafora. 2007. Research coordination networks: A phylogeny for kingdom fungi (Deep Hypha). Mycologia 98:829–837.

Bloem, J., M. Veninga, and J. Shepherd. 1995. Fully automatic determination of soil bacterium numbers, cell volumes, and frequencies of dividing cells by confocal laser scanning microscopy and image analysis. Appl. Environ. Microbiol. 61:926–936.

Boddy, L., J.C. Frankland, and P. van West. 2007. Ecology of saprotrophic basidiomycetes. Elsevier Academic, Amsterdam, the Netherlands.

Boschker, H.T.S., S.C. Nold, P. Wellsbury, D. Bos, W. de Graaf, R. Pel, R.J. Parkes, and T.E. Cappenberg. 1998. Direct linking of microbial populations to specific biogeochemical processes by 13C-labelling of biomarkers. Nature (London) 392:801–805.

Bothe, H., S.J. Ferguson, and W.E. Newton (eds.) 2007. Biology of the nitrogen cycle. Elsevier, Amsterdam, the Netherlands.

Bracker, C.E. 1967. Ultrastructure of fungi. Annu. Rev. Phytopathol. 5:343–374.

Braselton, J.P. 2001. Plasmodiophoromycota, p. 81–91. *In* D.J. McLaughlin, E.G. McLaughlin, and P.A. Lemke (eds.) The Mycota. Vol. VIIa, Systematics and evolution. Springer-Verlag, Berlin, Germany.

Bridges, E.M. 1997. World soils. 3rd edn. Cambridge University Press, London, U.K.

Bruns, T.D., and M. Gardes. 1993. Molecular tools for the identification of ectomycorrhizal fungi—Taxon-specific oligonucleotide probes for suilloid fungi. Mol. Ecol. 2:233–242.

Bruns, T.D., T.M. Szaro, M. Gardes, K.W. Cullings, J.J. Pan, D.L. Taylor, T.R. Horton, A. Kretzer, M. Garbelotto, and Y. Li. 1998. A sequence database for the identification of ectomycorrhizal basidiomycetes by phylogenetic analysis. Mol. Ecol. 7:257–272.

Bruns, T.D., T.J. White, and J.W. Taylor. 1991. Fungal molecular systematics. Annu. Rev. Ecol. Syst. 22:525–564.

Buczacki, S.T. (ed.) 1983. Zoosporic plant pathogens. Academic Press, London, U.K.

Carreiro, M.M., and R.E. Koske. 1992. Room temperature isolations can bias against selection of low temperature microfungi in temperate forest soils. Mycologia 84:886–900.

Chesters, C.G.C. 1949. Concerning fungi inhabiting soil. Trans. Br. Mycol. Soc. 32:197–216.

Christensen, M. 1989. A view of fungal ecology. Mycologia 81:1–19.

Clausen, C.A. 1997. Immunological detection of wood decay fungi—An overview of techniques developed from 1986 to the present. Intl. Biodet. Biodegrad. 39:133–143.

Clemençon, H. 1998. Observing the dolipore with the light microscope. Inoculum 49:3.

Cooke, R.C., and A.D.M. Rayner. 1984. Ecology of saprotrophic fungi. Longman, London, U.K.

de Boer, W., L.B. Folman, R.C. Summerbell, and L. Boddy. 2005. Living in a fungal world: Impact of fungi on soil bacterial niche development. FEMS Microbiol. Rev. 29:795–811.

Dick, M.W. 2001a. Straminipilous fungi: Systematics of the peronosporomycetes, including accounts of the marine straminipilous protists, the plasmodiophorids, and similar organisms. Kluwer Academic, Dordrecht, the Netherlands.

Dick, M.W. 2001b. Peronosporomycetes, p. 39–72. *In* D.J. McLaughlin, E.G. McLaughlin, and P.A. Lemke (eds.) The Mycota. Vol. VIIa, Systematics and evolution. Springer-Verlag, Berlin, Germany.

Domsch, K.H., W. Gams, and T.-H. Anderson. 2007. Compendium of soil fungi. 2nd edn. IHW Verlag, Eching, Germany.

Edgington, L.V., K.L. Khew, and G.L. Barron. 1971. Fungitoxic spectrum of benzimidazole compounds. Phytopathology 61:42–44.

Elfstrand, S., J. Lagerlof, K. Hedlund, and A. Martensson. 2008. Carbon routes from decomposing plant residues and living roots into soil food webs assessed with C-13 labelling. Soil Biol. Biochem. 40:2530–2539.

Embley, T.M., and E. Stackebrandt. 1994. The molecular phylogeny and systematics of the actinomycetes. Annu. Rev. Microbiol. 48:257–289.

Enloe, H.A., R.C. Graham, and S.C. Sillett. 2006. Arboreal histosols in old-growth redwood forest canopies, northern California. Soil Sci. Soc. Am. J. 70:408–418.

Entry, J.A., C.L. Rose, and K. Cromack. 1992. Microbial biomass and nutrient concentrations in hyphal mats of the ecotmycorrhizal fungus *Hysterangium setchellii* in a conifer forest soil. Soil Biol. Biochem. 24:447–453.

Erwin, D.C., and O.K. Ribeiro. 1996. Phytophthora diseases worldwide. APS Press, St. Paul, MN.

Euzéby, J.P. 2009. List of prokaryotic names with standing in nomenclature. Available online at: http://www.bacterio.cict.fr/ (last accessed August 7, 2009).

Fanning, D.S., and M.C. Balluff. 1989. Soil: Morphology, genesis, and classification. John Wiley, New York.

Fermor, T.R., and W.D. Grant. 1985. Degradation of fungal and actinomycete mycelia by *Agaricus bisporus*. J. Gen. Microbiol. 131:1729–1734.

Frankland, J.C. 1982. Biomass and nutrient cycling by decomposer basidiomycetes, p. 241–261. *In* J.C. Frankland, J.N. Hedger, and M.J. Swift (eds.) Decomposer basidiomycetes. Cambridge University Press, Cambridge, U.K.

Frankland, J.C. 1984. Autecology and the mycelium of a woodland litter decomposer, p. 241–260. *In* D.H. Jennings and A.D.M. Rayner (eds.) The ecology and physiology of the fungal mycelium. Cambridge University Press, Cambridge, U.K.

Fuller, M.S. 2001. Hyphochytriomycota, p. 73–80. *In* D.J. McLaughlin, E.G. McLaughlin, and P.A. Lemke (eds.) The Mycota. Vol. VIIa, Systematics and evolution. Springer-Verlag, Berlin, Germany.

Gardes, M., and T.D. Bruns. 1993. ITS primers with enhanced specificity for basidiomycetes—Application to the identification of mycorrhizae and rusts. Mol. Ecol. 2:113–118.

Gargas, A., P.T. DePriest, M. Grube, and A. Tehler. 1995. Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. Science (Washington, DC) 268:1492–1495.

Garrett, S.D. 1963. Soil fungi and soil fertility. Pergamon Press, Oxford, U.K.

Gehrig, H., A. Schüβler, and M. Kluge. 1996. Geosiphon pyriforme, a fungus forming endocytobiosis with Nostoc (Cyanobacteria), is an ancestral member of the Glomales: Evidence by SSU rRNA analysis. J. Mol. Evol. 43:71–81.

Gerber, N.N., and H.A. Lechevalier. 1965. Geosmin, an earth-smelling substance isolated from actinomycetes. Appl. Microbiol. 13:935–938.

Goodfellow, M., and T. Cross. 1974. Actinomycetes, p. 269–302. *In* C.H. Dickinson and G.J.F. Pugh (eds.) Biology of plant litter decomposition. Vol. 2. Academic Press, London, U.K.

Grube, M., and D.L. Hawksworth. 2007. Trouble with lichen: The re-evaluation and reinterpretation of thallus form and fruit body types in the molecular era. Mycol. Res. 111:1116–1132.

Hagler, A.N., and D.G. Ahearn. 1981. Rapid diazonium blue B test to detect basidiomycetous yeasts. Intl. J. Syst. Bacteriol. 31:204–208.

Hale, M. 1983. Biology of lichens. 3rd edn. E. Arnold, London, U.K.

Harris, D., R.P. Voroney, and E.A. Paul. 1997. Measurement of microbial biomass N:C by chloroform fumigation–incubation. Can. J. Soil Sci. 77:507–514.

Hawksworth, D.L. 1988. The variety of fungal-algal symbioses, their evolutionary significance, and the nature of lichens. Bot. J. Linn. Soc. (London) 96:3–20.

Hawksworth, D.L., B.C. Sutton, and G.C. Ainsworth. 1983. Ainsworth & Bisby's dictionary of the fungi. 7th edn. Commonwealth Mycological Institute, Kew, U.K.

Hibbett, D.S., M. Binder, J.F. Bischoff, M. Blackwell, P.F. Cannon, O. Eriksson, S. Huhndorf et al. 2007. A higher-level phylogenetic classification of the fungi. Mycol. Res. 111:509–547.

Hibbett, D.S., and R.G. Thorn. 2001. Homobasidiomycetes, p. 121–168. *In* D.J. McLaughlin, E.G. McLaughlin, and P. Lemke (eds.) The Mycota. Vol. VIIb. Systematics and evolution. Springer-Verlag, Berlin, Germany.

Holben, W.E., J.J. Jansson, B.K. Chelm, and J.M. Tiedje. 1988. DNA probe method for the detection of specific microorganisms in the soil bacterial community. Appl. Environ. Microbiol. 54:703–711.

Humber, R.A. 1989. Synopsis of a revised classification for the Entomophthorales (Zygomycotina). Mycotaxon 34:441–460.

Hutchison, L.J. 1990. Studies on the systematics of ectomycorrhizal fungi in axenic culture. IV. The effect of some selected fungitoxic compounds upon linear growth. Can. J. Bot. 68:2172–2178.

James, T.Y., F. Kauff, C.L. Schoch, P.B. Matheny, V. Hofstetter, C.J. Cox, G. Celio et al. 2006a. Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature (London) 443:818–822.

James, T.Y., P.M. Letcher, J.E. Longcore, S.E. Mozley-Standridge, D. Porter, M.J. Powell, G.W. Griffith, and R. Vilgalys. 2006b. A molecular phylogeny of the flagellated fungi (Chytridiomycota) and a proposal for a new phylum (Blastocladiomycota). Mycologia 98:860–871.

Jansa, J., A. Mozafar, S. Banke, B.A. McDonald, and E. Frossard. 2002. Intra- and intersporal diversity of ITS rDNA sequences in *Glomus intraradices* assessed by cloning and sequencing, and by SSCP analysis. Mycol. Res. 106:670–681.

Jayasinghe, B.A.T.D., and D. Parkinson. 2008. Actinomycetes as antagonists of litter decomposer fungi. Appl. Soil Ecol. 38:109–118.

Jeong, S.C., N.J. Ritchie, and D.D. Myrold. 1999. Molecular phylogenies of plants and *Frankia* support multiple origins of actinorhizal symbioses. Mol. Phylogenet. Evol. 13:493–503.

Joergensen, R.G., and F. Wichern. 2008. Quantitative assessment of the fungal contribution to microbial tissue in soil. Soil Biol. Biochem. 40:2977–2991.

Johansen, J.R. 1993. Cryptogamic crusts of semiarid and arid lands of North America. J. Phycol. 28:139–147.

Johnson, G.C. 1994. A comparison of the efficacy of six media for isolating wood-decaying hymenomycetes. Can. J. Microbiol. 41:104–107.

Keeling, P.J., G. Burger, D.G. Durnford, B.F. Lang, R.W. Lee, R.E. Pearlman, A.J. Roger, and M.W. Gray. 2005. The tree of eukaryotes. Trends Ecol. Evol. 20:670–676.

Kirby, R. 2006. Actinomycetes and lignin degradation. Adv. Appl. Microbiol. 58:25–168.

Kirk, P.M., P.F. Cannon, D.W. Minter, and J.A. Stalpers (eds.). 2008. Dictionary of the fungi. 10th edn. CABI, Wallingford, U.K.

Kjøller, A., and S. Struwe. 1982. Microfungi in ecosystems: Fungal occurrence and activity in litter and soil. Oikos 39:389–422.

Kõljalg U., K.H. Larsson, K. Abarenkov, R.H. Nilsson, I.J. Alexander, U. Eberhardt, S. Erland et al. 2005. UNITE: A database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytol. 166:1063–1068.

Kurtzman, C.P., and J.W. Fell. 1998. The yeasts: A taxonomic study. 4th edn. Elsevier, Amsterdam, the Netherlands.

Landeweert, R., C. Veenman, T.W. Kuyper, H. Fritze, K. Wernars, and E. Smit. 2003. Quantification of ectomycorrhizal mycelium in soil by real-time PCR compared to conventional quantification techniques. FEMS Microbiol. Ecol. 45:283–292.

Li, Y., W.A. Dick, and O.H. Tuovinen. 2004. Fluorescence microscopy for visualization of soil microorganisms—A review. Biol. Fertil. Soils 39:301–311.

Lievens, B., M. Brouwer, A.C.R.C. Vanachter, B.P.A. Cammue, and B.P.H.J. Thomma. 2006. Real-time PCR for detection and quantification of fungal and oomycete tomato pathogens in plant and soil samples. Plant Sci. (Shannon) 171:155–165.

Lovelock, C.E., K. Andersen, and J.B. Morton. 2003. Arbuscular mycorrhizal communities in tropical forests are affected by host tree species and environment. Oecologia 135:268–279.

Lutzoni, F., F. Kauff, C.J. Cox, D. McLaughlin, G. Celio, B. Dentinger, M. Padamsee et al. 2004. Assembling the fungal tree of life: Progress, classification and evolution of subcellular traits. Am. J. Bot. 91:1446–1480.

Lynch, M.D.J., and R.G. Thorn. 2006. Diversity of basidiomycetes in Michigan agricultural soils. Appl. Environ. Microbiol. 72:7050–7056.

Malloch, D., K.A. Pirozynski, and P.H. Raven. 1980. Ecological and evolutionary significance of mycorrhizal symbioses in vascular plants (a review). Proc. Natl. Acad. Sci. USA 77:2113–2118.

Marek, S.M., K. Hansen, M. Romanish, and R.G. Thorn. 2009. Molecular systematics of the cotton root rot pathogen, *Phymatotrichopsis omnivora*. Persoonia 22:63–74.

Margulis, L., and K.V. Schwartz. 1988. Five kingdoms: An illustrated guide to the phyla of life on earth. 2nd edn. W.H. Freeman, New York.

Martin, F.N. 1992. *Pythium*, p. 39–49. *In* L.L. Singleton, J.D. Mihail, and C.M. Rush (eds.) Methods for research on soilborne phytopathogenic fungi. APS Press, St. Paul, MN.

Martin, G.W., and C.J. Alexopoulos. 1969. The myxomycetes. University of Iowa Press, Iowa.

Martin, K.J., and P.T. Rygiewicz. 2005. Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. BMC Microbiol. 5:28.

Mason, M.G., A.S. Ball, B.J. Reeder, G. Silkstone, P. Nicholls, and M.T. Wilson. 2001. Extracellular heme peroxidases in actinomycetes: A case of mistaken identity. Appl. Environ. Microbiol. 67:4512–4519.

McArthur, F.A., M.O. Baerlocher, N.A.B. MacLean, M.D. Hiltz, and F. Baerlocher. 2001. Asking probing questions: Can fluorescent in situ hybridization identify and localise aquatic hyphomycetes on leaf litter? Int. Rev. Hydrobiol. 86:429–438.

Moore, R.T. 1985. The challenge of the dolipore/parenthosome septum, p. 175–212. *In* D. Moore, L. Casselton, D.A. Wood, and J.C. Frankland (eds.) Developmental biology of higher fungi. Cambridge University Press, Cambridge, U.K.

Morton, J.B. 1988. Taxonomy of VA mycorrhizal fungi: Classification, nomenclature, and identification. Mycotaxon 32:267–324.

Morton, J.B., and G.L. Benny. 1990. Revised classification of arbuscular mycorrhizal fungi (zygomycetes): A new order, Glomales, two new suborders, Glomineae and Gigasporineae, and two new families, Acaulosporaceae and Gigasporaceae, with an emendation of Glomaceae. Mycotaxon 37:471–491.

Mueller, G.M., G.F. Bills, and M.S. Foster (eds.). 2004. Biodiversity of fungi: Inventory and monitoring methods. Elsevier, Amsterdam, the Netherlands.

Nadkarni, N.M., D. Schaefer, T.J. Matelson, and R. Solano. 2004. Biomass and nutrient pools of canopy and terrestrial components in a primary and a secondary montane cloud forest, Costa Rica. For. Ecol. Manage. 198:223–236.

Nagahama, T., H. Sato, M. Shimazu, and J. Sugiyama. 1995. Phylogenetic divergence of the entomophthoralean fungi: Evidence from nuclear 18S ribosomal RNA gene sequences. Mycologia 87:203–209.

Nannipieri, P., J. Ascher, M.T. Ceccherini, L. Landi, G. Pietramellara, and G. Renella. 2003. Microbial diversity and soil functions. Eur. J. Soil Sci. 54:655–670.

Nash, T.H., III. 2008. Lichen biology. 2nd edn. Cambridge University Press, Cambridge, U.K.

Nawawi, A., J. Webster, and R.A. Davey. 1977. *Dendrosporomyces prolifer* gen. et sp. nov., a basidiomycete with branched conidia. Trans. Br. Mycol. Soc. 68:59–63.

Neufeld, J.D., M. Wagner, and J.C Murrell. 2007. Who eats what, where and when? Isotope-labelling experiments are coming of age. ISME J. 1:103–110.

O'Brien, H.E., J.L. Parrent, J.A. Jackson, J.-M. Moncalvo, and R. Vilgalys. 2005. Fungal community analysis by large-scale sequencing of environmental samples. Appl. Environ. Microbiol. 71:5544–5550.

O'Donnell, K.L. 1979. Zygomycetes in culture. Department of Botany. University of Georgia, Athens, GA.

Orth, A.B., D.J. Royse, and M. Tien. 1993. Ubiquity of lignin-degrading peroxidases among various wood-degrading fungi. Appl. Environ. Microbiol. 59:4017–4023.

Osono, T. 2007. Ecology of ligninolytic fungi associated with leaf litter decomposition. Ecol. Res. 22:955–974.

Pirozynski, K.A. 1981. Interactions between fungi and plants through the ages. Can. J. Bot. 59:1824–1827.

Pirozynski, K.A., and Y. Dalpé. 1989. Geological history of the Glomaceae with particular reference to mycorrhizal symbiosis. Symbiosis 7:1–36.

Pirozynski, K.A., and D.W. Malloch. 1975. The origin of land plants: A matter of mycotrophism. Biosystems 6:153–164.

Ploetz, R.C., G.A. Zentmyer, W.T. Nishijima, K.G. Rohrbach, and H.D. Ohr (eds.). 1994. Compendium of tropical fruit diseases. APS Press, St. Paul, MN.

Postgate, J. 1998. Nitrogen fixation. Cambridge University Press, Cambridge, U.K.

Prillinger, H., K. Lopandic, W. Schweigkofler, R. Deak, H.J. Aarts, R. Bauer, K. Sterflinger, G.F. Kraus, and A. Maraz. 2002. Phylogeny and systematics of the fungi with special reference to the Ascomycota and Basidiomycota. Chem. Immunol. 81:207–295.

Pugh, G.J F. 1974. Terrestrial fungi, p. 303–336. *In* C.H. Dickinson and G.J.F. Pugh (eds.) Biology of plant litter decomposition. Vol. 2. Academic Press, London, U.K.

Raper, K.B. 1984. The dictyostelids. Princeton University Press, Princeton, NJ.

Rayner, A.D.M., and L. Boddy. 1988. Fungal decomposition of wood: Its biology and ecology. Wiley, Chichester, U.K.

Redecker, D. 2000. Specific PCR primers to identify arbuscular mycorrhizal fungi within colonized roots. Mycorrhiza 10:73–80.

Redecker, D., and P. Raab. 2006. Phylogeny of the Glomeromycota (arbuscular mycorrhizal fungi): Recent developments and new gene markers. Mycologia 98:885–895.

Roberts, P. 1999. Rhizoctonia-forming fungi: A taxonomic guide. Royal Botanic Gardens, Kew, U.K.

Roesch, L.F.W., R.R. Fulthorpe, A. Riva, G. Casella, A.K.M. Hadwin, A.D. Kent, S.H. Daroub, F.A.O. Camargo, W.G. Farmerie, and E.W. Triplett. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. 1:283–290.

Sanders, I.R. 2002. Ecology and evolution of multigenomic arbuscular mycorrhizal fungi. Am. Nat. 160:S128–S141.

Savoie, J.M. 1998. Changes in enzyme activities during early growth of the edible mushroom, *Agaricus bisporus*, in compost. Mycol. Res. 102:1113–1118.

Schopf, J.W. 1993. Microfossils of the early archean apex chert: New evidence of the antiquity of life. Science (Washington, DC) 260:640–646.

Schüßler, A. 2002. Molecular phylogeny, taxonomy, and evolution of *Geosiphon pyriformis* and arbuscular mycorrhizal fungi. Plant Soil 244:75–83.

Schüßler, A., D. Schwarzott, and C. Walker. 2001. A new fungal phylum, the Glomeromycota: Phylogeny and evolution. Mycol. Res. 105:1413–1421.

Schwarze, F.W.M.R., J. Engels, and C. Mattheck. 2004. Fungal strategies of wood decay in trees, W. Linnard (transl.). Springer-Verlag, Berlin, Germany.

Seifert, K.A., and W. Gams. 2001. Taxonomy of anamorphic fungi, p. 307–347. *In* D.J. McLaughlin, E.G. McLaughlin, and P.A. Lemke (eds.) The Mycota. Vol. VIIa, Systematics and evolution. Springer-Verlag, Berlin, Germany.

Selosse, M.A., E. Baudoin, and P. Vandenkoornhuyse. 2004. Symbiotic microorganisms, a key for ecological success and protection of plants. C.R. Biol. 327:639–648.

Shaw, C.G., III, and L.F. Roth. 1976. Persistence and distribution of a clone of *Armillaria mellea* in ponderosa pine forest. Phytopathology 66:1210–1213.

Simon, L., J. Bousquet, R.C. Lévesque, and M. Lalonde. 1993. Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. Nature (London) 363:67–69.

Smiley, R.W., P.H. Dernoeden, and B.B. Clarke. 2005. Compendium of turfgrass diseases. 3rd edn. APS Press, St. Paul, MN.

Smith, M.L., J.N. Bruhn, and J.B. Anderson. 1992. The fungus *Armillaria bulbosa* is among the largest and oldest living organisms. Nature (London) 356:428–431.

Smith, S.E., and D.J. Read. 2008. Mycorrhizal symbiosis. 3rd edn. Academic Press, Amsterdam, the Netherlands.

Sneh, B., L. Burpee, and A. Ogoshi. 1991. Identification of *Rhizoctonia* species. APS Press, St. Paul, MN.

Stach, J.E.M., L.A. Maldonado, A.C. Ward, M. Goodfellow, and A.T. Bull. 2003. New primers for the class Actinobacteria: Application to marine and terrestrial environments. Environ. Microbiol. 5:828–841.

Stephenson, S.L., and H. Stempen. 1994. Myxomycetes: A handbook of slime molds. Timber Press, Portland, OR.

Summerbell, R.C. 1985. The staining of filamentous fungi with diazonium blue B. Mycologia 77:587–593.

Swift, M.J. 1973. The estimation of mycelial biomass by determination of hexosamine content of wood decayed by fungi. Soil Biol. Biochem. 5:321–332.

Sykorová, Z., A. Wiemken, and D. Redecker. 2007. Cooccurring *Gentiana verna* and *Gentiana acaulis* and their neighboring plants in two Swiss upper montane meadows harbor distinct arbuscular mycorrhizal fungal communities. Appl. Environ. Microbiol. 73:5426–5434.

Tambong, J.T., A.W.A.M. de Cock, N.A. Tinker, and C.A. Lévesque. 2006. Oligonucleotide array for identification and detection of *Pythium* species. Appl. Environ. Microbiol. 72:2691–2706.

Taylor, J.W., and M.L. Berbee. 2006. Dating divergences in the fungal tree of life: Review and new analyses. Mycologia 98:838–849.

Thorn, G. 1997. The fungi in soil, p. 63–127. *In* J.D. van Elsas, J.T. Trevors, and E.M.H. Wellington (eds.) Modern soil microbiology. Marcel Dekker, New York.

Thorn, R.G., C.A. Reddy, D. Harris, and E.A. Paul. 1996. Isolation of saprophytic basidiomycetes from soil. Appl. Environ. Microbiol. 62:4288–4292.

Tringea, S.G., and P. Hugenholtz. 2008. A renaissance for the pioneering 16S rRNA gene. Curr. Opin. Microbiol. 11:442–446.

Tsao, P.H., and S.O. Guy. 1977. Inhibition of Mortierella and *Pythium* in a Phytophthora-isolation medium containing hymexazol. Phytopathology 67:796–801.

Tsuji, T., Y. Kawasaki, S. Takeshima, T. Sekiya, and S. Tanaka. 1995. A new fluorescence staining assay for visualizing living microorganisms in soil. Appl. Environ. Microbiol. 61:3415–3421.

Tunlid, A., and D.C. White. 1992. Biochemical analysis of biomass, community structure, nutritional status, and metabolic activity of microbial communities in soil, p. 229–262. *In* G. Stotzky and G.M. Bollag (eds.) Soil biochemistry. Vol. 7. Marcel Dekker, New York.

Van de Peer, Y., and R. De Wachter. 1997. Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. J. Mol. Evol. 45:619–630.

van der Heijden, M.G.A., J.N. Klironomos, M. Ursic, P. Moutoglis, R. Streitwolf-Engel, T. Boller, A. Wiemken, and I.R. Sanders. 1998. Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. Nature (London) 396:69–72.

van der Linde, S., I. Alexander, and I.C. Anderson. 2008. A PCR-based method for detecting the mycelia of stipitate hydnoid fungi in soil. J. Microbiol. Methods 75:40–46.

Vandenkoornhuyse, P., S.L. Baldauf, C. Leyval, J. Straczek, and J.P.W. Young. 2002. Extensive fungal diversity in plant roots. Science (Washington, DC) 295:2051.

Vandenkoornhuyse, P., K.P. Ridgway, I.J. Watson, A.H. Fitter, and J.P.W. Young. 2003. Co-existing grass species have distinctive arbuscular mycorrhizal communities. Mol. Ecol. 12:3085–3095.

Vessey, J.K., K. Pawlowski, and B. Bergman. 2004. Root-based N_2-fixing symbioses: Legumes, actinorhizal plants, *Parasponia* sp. and cycads. Plant Soil 266:205–230.

Wall, L.G. 2000. The actinorhizal symbiosis. J. Plant Growth Regul. 19:167–182.

Wang, B., and Y.L. Qiu. 2006. Phylogenetic distribution and evolution of mycorrhizas in land plants. Mycorrhiza 16:299–363.

Webster, J. 1981. Introduction to fungi. 2nd edn. Cambridge University Press, Cambridge, U.K.

Wells, K. 1978. The fine structure of septal pore apparatus in the lamellae of *Pholiota terrestris*. Can. J. Bot. 56:2915–2924.

Wells, K. 1994. Jelly fungi, then and now! Mycologia 86:18–48.

West, N.E. 1990. Structure and function of microphytic soil crusts in wildland ecosystems of arid to semi-arid regions. Adv. Ecol. Res. 20:179–223.

Wheelis, M.L., O. Kandler, and C.R. Woese. 1992. On the nature of global classification. Proc. Natl. Acad. Sci. USA 89:2930–2934.

White, M.M., T.Y. James, K. O'Donnell, M.J. Cafaro, Y. Tanabe, and J. Sugiyama. 2006 (publ. 2007). Phylogeny of the Zygomycota based on nuclear ribosomal sequence data. Mycologia 98:872–884.

Williams, S.T., S. Lanning, and E.M.H. Wellington. 1984. Ecology of actinomycetes, p. 481–528. *In* M. Goodfellow, M. Mordarski, and S.T. Williams (eds.) Biology of actinomycetes. Academic Press, London, U.K.

Woese, C.R. 1987. Bacterial evolution. Microbiol. Rev. 51:221–271.

Wood, D.A. 1979. A method for estimating biomass of *Agaricus bisporus* in a solid substrate, compost wheat straw. Biotechnol. Lett. 1:255–260.

Worrall, J.J. 1991. Media for the selective isolation of hymenomycetes. Mycologia 83:296–302.

Worrall, J.J., S.E. Anagnost, and R.A. Zabel. 1997. Comparison of wood decay among diverse lignicolous fungi. Mycologia 89:199–219.

Zhao, X.R., Q. Lin, and P.C. Brookes. 2005. Does soil ergosterol concentration provide a reliable estimate of soil fungal biomass? Soil Biol. Biochem. 37:311–317.

Zhi, X.Y., W.J. Li, and E. Stackebrandt. 2009. An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. Int. J. Syst. Evol. Microbiol. 59:589–608.

24.4 Mycorrhizae

Raffaella Balestrini

Valeria Bianciotto

Paola Bonfante

24.4.1 Introduction

Mycorrhizal fungi are specialized root symbionts that engage in intimate association with a great diversity of plants (Smith and Read, 2008; Bonfante and Genre, 2010). The term mycorrhiza implies an association of fungi with roots (from the Greek "mycos," meaning fungus and "rhiza," meaning root), relationships, called mycorrhizal associations. These are also found between fungi and the underground gametophytes of many bryophytes and pteridophytes, as well as the sporophytes of most pteridophytes and the roots of seed plants (Bonfante and Genre, 2008; Smith and Read, 2008). The most understood function of such symbioses concerns the improvement of plant mineral nutrient acquisition, in exchange for carbon compounds, which results in positive host growth responses. Mycorrhizal fungi can also perform many other significant roles, including protection of the plant from biotic and abiotic stress, for instance, by altering host environmental tolerances to water deficit or pollutants or reducing susceptibility to soilborne pathogens (Aroca et al., 2007; Smith and Read, 2008).

Structures similar to arbuscular mycorrhiza have been observed in plant fossils from the Early Devonian (Selosse and Le Tacon, 1998), whereas fossil ectomycorrhiza have been found in the middle Eocene (Le Page et al., 1997). The variety of mycorrhizal associations established between plants and fungi has been divided into seven categories (arbuscular mycorrhiza, ectomycorrhiza, ectendomycorrhiza, arbutoid mycorrhiza, monotropoid mycorrhiza, ericoid mycorrhiza, orchid mycorrhiza—Table 24.4), primarily on the basis of the structural characteristics of the symbiotic interfaces and the taxonomic identity of the symbionts (Smith and Read, 2008).

Mycorrhizal associations are found in most annual and perennial plants. About two-thirds of these plants, mostly herbaceous species belonging to crops (e.g., maize and rice), are symbiotic with arbuscular mycorrhizal (AM) fungi. Ericoid mycorrhizae are ecologically important as efficient organic matter (OM) degraders but are mainly restricted to heathlands (Read and Perez-Moreno, 2003). While ectomycorrhizae (ECM) are made up of a relatively small number of plants (about 8000 species, see Section 24.4.2.2), their ecological importance is amplified by their extensive occupancy of biomes (Martin, 2007). These associations dominate in boreal, temperate, Mediterranean and some subtropical forest ecosystems (Read and Perez-Moreno, 2003). Pinaceae, members of which form the major component of the vast boreal forests in the northern hemisphere, and Fagaceae, dominants or codominants of the northern and some southern hemisphere temperate forests, as well as tropical forests in Southeast Asia, are predominantly ectomycorrhizal species (Smith and Read, 2008). Some genera of shrubs are also routinely found to be ECM (i.e., Cistaceae). These plant species, through the symbioses with ECM fungi, have been able to acquire metabolic capabilities that in turn have allowed otherwise unavailable ecological niches to be utilized.

This chapter focuses on the most relevant aspects of mycorrhizal symbioses, focusing on the nutrient exchanges between the partners. It also provides insights into new approaches that can be used to study these fungi directly in the field (metagenomics or environmental genomics).

24.4.2 Mycorrhizal Fungi: A Heterogeneous Group of Soil Fungi

24.4.2.1 Arbuscular Mycorrhizal Fungi

The most ancient mycorrhizal fungi, on the basis of evidence from paleobotanical studies and the analysis of rDNA-based phylogenies, are those that form arbuscular mycorrhizae. Their earliest fossil record dates back to the Ordovician, 460 million

TABLE 24.4   Characteristics of the Several Mycorrhizal Types

Kinds of Mycorrhizae	Arbuscular Mycorrhizae	Ectomycorrhizae	Ericoid Mycorrhizae	Orchid Mycorrhizae	Ecto-Endomycorrhizae	Arbutoid Mycorrhizae	Monotropoid Mycorrhizae
Mycobiont	Glomeromycota	Basidiomycota Ascomycota (Zygomycetes)	Ascomycota	Basidiomycota	Basidiomycota Ascomycota	Basidiomycota	Basidiomycota
Phytobiont	Bryophyta Pteridophyta Gymnospermae Angiospermae	Gymnospermae Angiospermae	Ericales Byophyta	Orchidales	Gymnospermae Angiospermae	Ericales	Monotropoideae
Intracellular structures	+	−	+	+	+	+	+
Fungal mantle	−	+	−	−	±	±	+
Hartig net	−	+	−	−	+	+	+

years ago (Redecker et al., 2000), but molecular clock-based inferences estimate their origin as much earlier (Tehler et al., 2000; Berbee and Taylor, 2001; Schüβler et al., 2001), possibly as far back as between 1200 and 1400 Ma ago (Heckman et al., 2001). Thus, it has been hypothesized that symbiosis played a key role in the colonization of lands by plants (Taylor et al., 1995; Phipps and Taylor, 1996). AM fungi have been accommodated inside the phylum Glomeromycota, which is a sister clade to Asco- and Basidiomycota (Schüβler et al., 2001), on the basis of the 18S ribosomal gene sequences. More information is required to further support the definition of this phylum, and to clarify evolutionary patterns within the group, although information from different genomic loci/genes is continuously being collected (e.g., Helgason et al., 2003; Corradi et al., 2004; van der Heijden et al., 2004; Da Silva et al., 2006; Lee and Young, 2009).

Glomeromycota were considered a species-poor group, with about 214 species recognized by conventional spore-based morphology, based on a traditional, morphological species concept. Recently, new taxa have been identified and have been named *Ambispora*, *Intraspora*, *Pacispora*, *Kuklospora*, *Otospora*, and *Diversispora* (www.amf-phylogeny.com). Since these few morphospecies colonize about two-thirds of modern plants, arbuscular mycorrhizae appear to be not only the most ubiquitous and abundant terrestrial symbiosis but also nonspecific, with most AM fungi readily colonizing almost any susceptible plant species. Although some plant species enter into specialized relationships with only a few AM fungi, others appear to associate with as many as 20 different AM fungi, thus hosting very diverse AM fungal communities that include currently unknown taxa. Possible relationships between plant responsiveness to mycorrhiza and their selectivity toward specific AM fungi still have to be elucidated (van der Heijden, 2002). Their uniqueness is mirrored not only by this taxonomical status but also by their biological traits. AM fungi are obligate biotrophs; although AM fungal spores can germinate in the absence of host plants (presymbiotic phase), they depend on the establishment of symbiosis with a living photoautotrophic partner for the development of the intracellular colonization. This step is mandatory to complete the fungal life cycle and produce the next generation of spores. AM fungi are usually aseptate and coenocytic, with hundreds of nuclei sharing the same cytoplasm. At the same time, individual spores contain hundreds of nuclei and the question of how the different polymorphic DNA sequence variants are distributed between genomes or nuclei is still being studied (Parniske, 2008). Although there is no confirmed report of a sexual stage in the life cycle of AM fungi, Croll et al. (2008) have recently suggested that AMF are promiscuous, in the sense that an individual will form hyphal connections with several other members of the population and this creates the possibility of genetic exchanges. In particular, they have suggested that the exchange of nuclei among AMF isolates provides the opportunity for the fusion of genetically different nuclei and recombination (Croll et al., 2008; Croll and Sanders, 2009).

The sequencing of representative AM fungi, such as *Glomus intraradices* (Martin et al., 2008b), is currently in progress (http://www.jgi.doe.gov/DOEmicrobes2004.html), but due to the obligate biotrophic status of the fungus, it has encountered unexpected difficulties, rendering its achievement far from immediate (Martin et al., 2008b). Waiting for the production of a completely annotated and assembled *G. intraradices* genome, the increasing number of EST sequences obtained inside the *Glomus* Genome Consortium (GGC; INRA *Glomus* database: http://mycor.nancy.inra.fr/genomeResources.php) might provide new clues to strengthen our understanding of the biology of these fungi (Martin et al., 2008b). The assembly and annotation of the mitochondrial genome has been concluded, and it is already available in Gene Bank (Lee and Young, 2009; Young, 2009).

24.4.2.2 ECM Fungi

Ectomycorrhizal fungi are best known for their fruiting structures (e.g., toadstools), which often grow next to tree trunks in woodlands. However, morphological recognition has shown that ectomycorrhizal roots often contain fungi that cannot be linked to epigeous fruiting bodies (Anderson and Cairney, 2007). Most of the ECM plant species are trees or shrubs belonging to Betulaceae, Cistaceae, Dipterocarpaceae, Fagaceae,

Pinaceae, Myrtaceae, Salicaceae, and several tribes of Fabaceae (Martin, 2007). Although they represent a relatively small number of plants, the ecological importance of these symbioses is amplified by their extensive occupancy of biomes (Read and Perez-Moreno, 2003). Ectomycorrhizal fungi include at least 6000 species and, unlike AMs, ECM symbiosis involves a diverse range of fungi: a large number of basidiomycetes, some ascomycetes, and a handful of zygomycetes belonging to the genus *Endogone* (Smith and Read, 2008). Fossil evidence on macrofungi is very limited (Taylor and Taylor, 1997). It is likely that a period of rapid diversification of the basidiomycetes occurred in the Cretaceous, when plants with ECM associations became important. Ectomycorrhizal Basidiomycetous taxa, like Cortinariaceae, Boletales, Amanitaceae, and Russulaceae, probably originated in this period. The rapid diversification of these fungi continues to this day, driven by increasing host and habitat specificity (Brundrett, 2002). Unlike AM fungi, ECM fungi show conventional sexual mechanisms, with possibly a few exceptions such as the asexual widespread ascomycete *Cenococcum geophilum* (Girlanda et al., 2007). Ectomycorrhizal basidiomycetes and ascomycetes are not a phylogenetically separate group, but an assemblage of very different fungal species that have independently developed a symbiotic lifestyle over the past 130–180 million years (Hibbett et al., 2000; Brundrett, 2002). The change between saprotrophic and mycorrhizal lifestyles probably happened convergently, and perhaps many times, during the evolution of these fungal lineages (Hibbett et al., 2000). This may have facilitated the evolution of ectomycorrhizal lineages with a broad range of physiological and ecological functions that partly reflect the activities of their disparate saprotrophic ancestors. Although a few Basidiomycota clades are exclusively ectomycorrhizal (e.g., bolets), most clades comprise both ectomycorrhizal and saprotrophic species (e.g., Tricholomataceae), suggesting that the symbiotic ability involves a limited number of genes or key enzyme activities (Martin, 2007). The genome projects of several fungal species with distinct and varied lifestyles (i.e., saprobic, symbiotic, and parasitic strategies) will offer a great opportunity to reveal how evolution has played a crucial part in shaping fungal genomes to suit their own specialized lifestyles (Pain and Hertz-Fowler, 2008). In recent years, sequencing projects of ectomycorrhizal fungi have been launched, since they grow in pure culture and offer excellent opportunities for DNA investigation. Among the basidiomycetes, the genome of the ECM *Laccaria bicolor* has just been sequenced, providing new insights into the mutually beneficial relationship between the host and fungus (Martin et al., 2008a). The analysis of cell wall–degrading enzymes present in the genome has offered information on the adaptation capacity (Martin et al., 2008a; Martin and Nehls, 2009). As far as ascomycetes are concerned, *Tuber melanosporum* genome sequencing has been recently completed (http://mycor.nancy.inra.fr/; *Tuber* Genome Consortium, Martin et al., 2010), and it represents another opportunity to carry out a deeper investigation of the processes by which symbiotic organisms interact with plant roots in the ecosystem.

The use of DNA-based genotyping methods has recently modified our understanding of the specificity of mycorrhizal fungi toward their host plants. Most ECM fungi and plants are capable of association with multiple partners, although specialists occur in ECM symbiosis in both fungi and plants (e.g., Trappe, 1962; Molina and Trappe, 1982; Molina et al., 1992; Horton and Bruns, 1998; Horton et al., 1999; Cullings et al., 2000). For example, observations have shown that the host range of ericoid mycorrhizal fungi may also be extended to ectomycorrhizal plants in nature (Bergero et al., 2000; Vrålstad et al., 2000; Tedersoo et al., 2006) and, vice versa, typical ECMs like truffles or Russula have been detected in orchids (Girlanda et al., 2007; Selosse and Roy, 2009).

24.4.2.3 Other Mycorrhizal Fungi

Mycorrhizal fungi that associate with members of the Ericaceae and the related families include fungi that are all ascomycetes and generally do not form mycorrhizae with other vascular plants (Smith and Read, 2008). Studies on DNA sequences of fungi from these plants in Australia, Europe, and North America have revealed two or more distantly related groups of fungi involved in ericoid mycorrhizae (McLean et al., 1999; Monreal et al., 1999; Sharples et al., 2000). *Hymenoscyphus*-like fungi associate with Ericales and bryophytes throughout the world but other taxa are more restricted to specific geographic regions (Chambers et al., 1999; Read et al., 2000). It is not certain whether ericoid mycorrhizal fungi exist primarily as soil saprotrophs or as mycorrhizal associates of plants. If they are less dependent on plants than AM or ECM fungi, their capacity to form mycorrhizal associations would not be a driving factor in their evolution (Brundrett, 2002). Many ericaceous species colonize habitats where most of the nutrients in the soil are in organic form. It is widely accepted that the main benefit conferred to the ericaceous host plant by mycorrhizal infection is the enzymatic degradation of organic polymers in the soil and the transfer of some of the resulting products to the root (Perotto et al., 2002). Thanks to their mycorrhizal status, host plants can access otherwise unavailable organic N and P (Näsholm and Persson, 2001). In addition to nutritional benefits, there are other advantages that are likely to be of ecological significance. The ability of the fungal partner to sequester and, in some cases, to metabolize metal ions that are toxic to the plant appears to be important (Martino et al., 2003). This combination of nutritional and nonnutritional attributes of these associations contribute significantly to the ability of these plants to grow in contaminated soils as well as in natural habitats (Smith and Read, 2008).

Members of the Ericales species, with monotropoid or arbutoid mycorrhizae (ECM-like associations), generally have much higher host–fungus specificity than other ECM associations. The simultaneous formation of ectomycorrhiza in tree hosts and ectoendomycorrhiza in plants of the Arbutoideae and Monotropoideae subfamilies (family Ericaceae) has long been known (Girlanda et al., 2007). Orchids have mycorrhizal associations with soil fungi that are believed to be essential for seed germination and to assist the growth of adult plants (reviewed in

Waterman and Bidartondo, 2008). Most orchids have fairly specific fungal associates that vary according to the host species and habitat (Otero et al., 2002; Ma et al., 2003; Suarez et al., 2006). Most of these fungi are assigned to the anamorphic form genus *Rhizoctonia* (Roberts, 1999). Rhizoctonia-forming fungi constitute a polyphyletic group of fungi from three basidiomycetes families (Sebacinaceae, Ceratobasidiaceae, and Tulasnellaceae). It seems most likely that orchid fungi are a disparate group with many separate origins and the recruitment of new fungal lineages by orchids is an ongoing process. These fungi rarely fruit in an axenic culture, thus making their identification difficult when vegetative characteristics are used alone. The advent of molecular systematics has revolutionized the study of mycorrhizae as it allows direct identification without axenic isolation (Waterman and Bidartondo, 2008). The benefits provided by orchids to their mycorrhizal fungi, if any, are not clear, as these fungi seem to grow just as well without their hosts as they do with them. Several nongreen orchids (mycoheterotrophic) are known to rely solely on C from their mycorrhizal fungi (Selosse and Roy, 2009). In fact, green plants (mixotrophic) that can obtain additional carbon from their symbiotic fungi have also been discovered. Among them, orchids such as *Epipactis*, *Cephalanthera*, *Plantanthera*, and, more recently, some small green perennial shrubs of the Ericaceae family (referred to as pyroloids) have been shown to receive variable amounts of C from their mycorrhizal fungi (reviewed in Selosse and Roy, 2009). The mycorrhizal fungi that are associated with pyroloids and orchids belong to diverse fungal taxa (ascomycetes and basidiomycetes) that usually form ectomycorrhizal on tree roots. In some studies, it has been reported that these same fungi can be simultaneously detected as associated to the trees (Selosse et al., 2004; Julou et al., 2005). These data suggest that mixotrophic plants that live in forests derive photosynthates by sharing symbiotic fungi (Selosse and Roy, 2009).

24.4.3 Diversity in Mycorrhizal Features: Mycorrhizal Fungi Develop Structures Outside and Inside the Roots

According to their ability to colonize the root cells, mycorrhizal fungi are divided into two main categories: endo- and ectomycorrhizae. Since a detailed structural description of mycorrhizae is not the main aim of this chapter, we here only briefly review the main structural features that characterize mycorrhizal morphologies. More detailed information can be found in Peterson et al. (2004). Endomycorrhizal hyphae adopt a variety of colonization patterns during their penetration of host root cells. AM fungi depend on their hosts on a great extent and cannot survive for long without them. Once the fungus has overcome the epidermal layer, it develops an appressoria and grows inter- and intracellularly all along the root in order to spread fungal structures. During the symbiotic phase, some Glomeromycota (i.e., *Glomus* species) develop structures, called vesicles, that fill the cellular spaces, and probably act as storage pools. Only when the fungus has reached the cortical layers, does a peculiar

branching process start, which leads to highly branched arbuscules (Bonfante et al., 2009; Bonfante and Genre, 2010). These are the key structures that are produced during the symbioses, and are considered the site of nutrient exchange (Figure 24.8a). During endosymbiosis, regardless of the involved partners, a new apoplastic space, based on membrane proliferation, is built around all the intracellular hyphae (Bonfante, 2001). In AM, this new compartment is known as the interfacial compartment, and it consists of the invaginated membrane, the cell wall-like material, and the fungal wall, and plasma membrane (Balestrini and Bonfante, 2005). Colonization of the Ericales takes place via simple hyphal structures whose organization is not modified to any great extent during the presymbiotic and the intraradical symbiotic phases. Hyphae of ericoid fungi can form loops and bundles outside the host plants, appressoria and coils inside the epidermal root cells (Figure 24.8b). Basidiomycete fungi produce specialized coils (pelotons) in the orchid cells of the protocorm, which is the first developmental stage in orchids, as well as in roots and rhizomes (Figure 24.8c). The interface in both Ericales and orchid endomycorrhizae is morphologically similar to that of AM fungi but different in composition. This indicates that the invagination of the host membrane around the intracellular hyphae is a common biotrophic interaction, while the molecules that are detected at the interface depend to a great extent on the nutritional capabilities of the fungus and mirror its relationship with the host plant.

During the symbiotic phase, ectomycorrhizal fungi form a fungal sheath (the mantle), consisting of aggregated hyphae that surround the root surface. This mycelium is linked to extramatrical hyphae that explore the substrate and are responsible for the mineral nutrition and water uptake of the symbiotic tissues. From the inner zone of the mantle, some hyphae penetrate between the root cells to form the Hartig net (Figure 24.8d), an intercellular hyphal network inside the root tissues where metabolites are exchanged. The hyphae always remain apoplastic and can colonize epidermal and cortical cell layers (Barker et al., 1998). In ECM, plant and fungal cell walls are always in direct contact, forming the interface between the two partners; it has also been observed that the fungus can cause subtle changes in the host cell walls (Balestrini et al., 1996; Martin et al., 2007).

24.4.4 Nutrient Exchanges

The success of mycorrhizal fungi in time and space is mainly linked to the nutritional benefits they confer to their plant hosts: They take up phosphate (Pi) and other macronutrients as well as microelements and water from the soil and deliver them to the plant. The fungus, in turn, receives photosynthetic carbohydrates. The extraradical mycorrhizal mycelium (ERMM) captures water and nutrients from the soil and is an important sink for host carbon. Inside the roots, carbohydrates and mineral nutrients are then exchanged across the interface between the plant and the fungus (Smith and Smith, 1990). In recent years, several studies have been performed with the aim of investigating the transfer of nutrients between symbionts and, from the
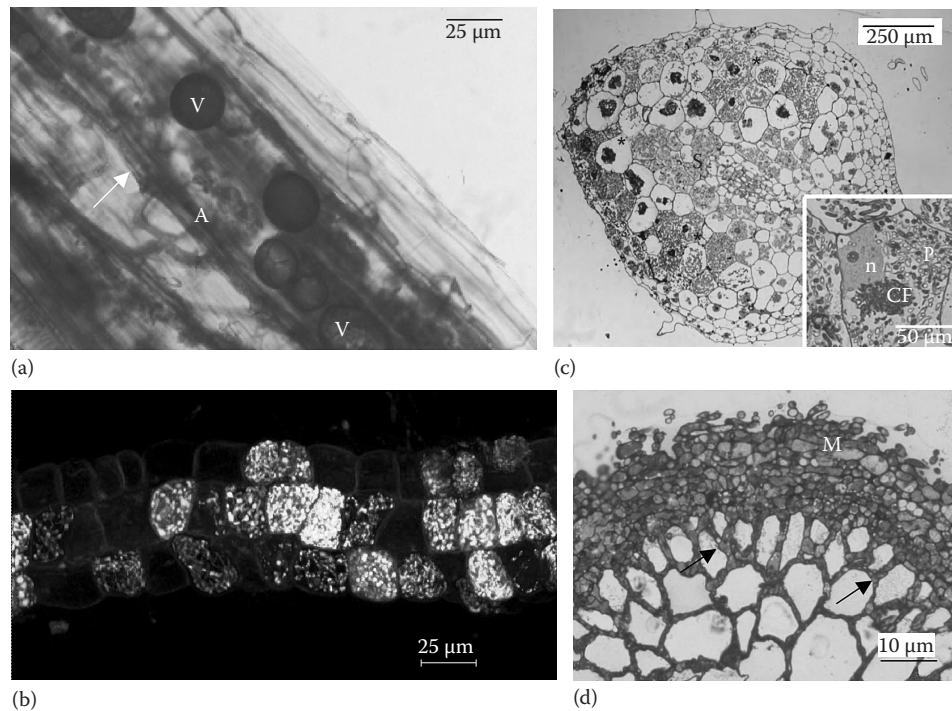
FIGURE 24.8 Mycorrhizal features: (a), (b), (c) different endomycorrhizal associations and (d) ectomycorrhiza. (a) Arbuscular mycorrhiza (AM). Clover mycorrhizal root staining with cotton blue. Typical fungal structures are evident in the cortical layers: vesicles (V); arbuscule (A). The arrows indicate the intercellular hyphae. (b) Ericoid mycorrhiza. Confocal microscopic images of *Vaccinium myrtillus* roots colonized by *Oidiodendron maius* expressing enhanced green fluorescent protein (EGFP). Fluorescent fungal coils are visible in the root cells observed 1 year after fungal inoculation. Some mycorrhizal cells are clearly fluorescent whereas others fluoresce poorly, possibly due to a loss in viability of the fungal hyphae after prolonged infection. (c) Orchid mycorrhiza. Protocorm of *Serapias vomeracea* mycorrhized in vitro with *Tulasnella* spp. The asterisks indicate the region colonized by the fungus with the characteristic pelotons. S, region rich in starch. In the inset: higher magnification of colonized cells. P, peloton; CF, collapsed fungus; n, plant nucleus. (d) Transverse section of an ectomycorrhiza of *Tilia* sp. and *Tuber borchii*, showing the mantle (M) ensheathing the root and the Hartig net mycelium progressing among the host cells (arrows). (Images a, b, c, and d courtesy of Andrea Berruti, Elena Martino, Michele Rodda, and Mara Novero.)

literature, it seems clear that most of the information available is for AMs and ECM. For this reason, in this section, we will focus on these mycorrhizal categories.

24.4.4.1 Phosphate

Improved phosphate uptake is the main benefit for plants in AM symbiosis (Bucher, 2007; Javot et al., 2007). The extensive hyphal network of AM fungi influences the physicochemical properties of the soil and directly or indirectly contributes to the release of phosphate from inorganic complexes of low solubility (Parniske, 2008). The fungal phosphate transporters that are expressed in the extraradical mycelium are involved in the uptake of phosphate from the substrate. The characterization of a high-affinity Pi transporter (PT) in *Glomus versiforme*, which is expressed in extraradical hyphae (Harrison and van Buuren, 1995), was a milestone in the definition of AM fungi as biofertilizers. Detailed gene expression data have been obtained for the orthologous gene identified in *G. intraradices* (Maldonado-Mendoza et al., 2001) and another PT ortholog (*GmosPT*) has been described in *Glomus mosseae*, a particularly important AM fungus for natural and agricultural ecosystems (Benedetto et al., 2005). Current data suggest that Pi, taken up by the extraradical

mycelium from soil solutions, is translocated through the AM fungal hyphae as polyphosphate (poly-Pi). After hydrolysis in the arbuscule, the phosphorous is then released as Pi into the periarbuscular space (Javot et al., 2007). The subsequent import of Pi across the periarbuscular membrane into the plant cell is mediated by plant transporters. The study of plant PTs and of their functioning thus represents a key point to fully understand this process. Mycorrhiza-specific PTs, which are responsible for plant Pi uptake in arbuscule-containing cells, have recently been characterized in many plants: potato, tomato, rice, maize, barley, *Medicago truncatula*, and *Lotus japonicus* (Rausch et al., 2001; Harrison et al., 2002; Paszkowski et al., 2002; Glassop et al., 2005; Nagy et al., 2005; Guether et al., 2009a). Recent data, based on expression analysis of laser-dissected cells, have demonstrated that the five so far identified *LePT* genes in tomato (Nagy et al., 2005) are differentially expressed in cortical cell populations during the interaction with *G. mosseae* (Balestrini et al., 2007). All tomato PT genes are consistently expressed inside the arbusculated cells, suggesting that plants guarantee maximum Pi uptake through the activation of the whole gene family. On the fungal side, the presence of *GmosPT* transcripts in the arbusculated cells offers sound confirmation of the results

obtained by Benedetto et al. (2005) on the whole mycorrhizal root. The discovery that five plant genes and one fungal PT gene are consistently expressed in the arbusculated cells provides a new scenario for plant–fungus nutrient exchanges. The efflux of phosphate probably occurs in competition with its uptake and the fungus might exert control over the amount of phosphate delivered to the plant.

As far as ectomycorrhizal symbiosis is concerned, the host plants rely to a great extent on the fungus for Pi uptake. This might be due to the ability of ECM fungi to take up complex or immobilized forms of these nutrients that the root alone cannot absorb effectively. Phosphate in a soil solution is taken up by ECM hyphae and then translocated to the host roots, its absorption and efflux likely being regulated by intracellular Pi and inorganic polyphosphates (poly-Pi) pools. Although several low- and high-affinity PTs have been identified in the genome sequence of *L. bicolor* (Martin et al., 2008a) and other symbiotic fungi, the molecular processes controlling Pi uptake in ECM fungi still remain unknown. Among the 13 low- and high-affinity PTs of *Populus trichocarpa*, one is expressed specifically in endomycorrhiza, whereas another family member is preferentially expressed in ectomycorrhizal root tips (Martin, 2007). It might be possible that, in an analogous manner to the periarbuscular membrane localization of the PT in endomycorrhiza, another PT is localized in the plasma membrane of the host cortical cells in contact with the fungal Hartig net in ectomycorrhiza.

24.4.4.2 Nitrogen

In addition to their function in P nutrition, AM and, especially, ECM fungi play a pivotal role in the uptake of nitrogen (N) by plants. The results obtained over the last decade have provided support to the hypothesis (Read, 1991) that, at a global scale, mycorrhizal fungi may be making significant contributions to ecosystem nutrient cycling (Read and Perez-Moreno, 2003). Ectomycorrhizal fungi have biochemical and physiological attributes that make them highly efficient in scavenging organic N sources (Read et al., 2004). They may use soluble amino acids, and some of these fungi have highly developed proteolytic capabilities that enable them to directly access macromolecular N from the soil (Abuzinadah and Read, 1989). Functional complementation of a yeast strain deficient in amino acid transporters has led to the identification of amino acid transporters from *Amanita muscaria* (Nehls et al., 1999) and *Hebeloma cylindrosporum* (Chalot et al., 2002; Wipf et al., 2002), which seem to be involved in both the uptake of amino acids from the soil solution and the retention of amino acids under N-deprivation conditions. Evidence is now accruing that AM fungi may also have access to organic N sources (Hodge et al., 2001; Leigh et al., 2008), thanks to the expression of specific amino acid permeases (Cappellazzo et al., 2008). AM and ECM mycelia are also effective scavengers of either ammonium or nitrate, inorganic forms of N, and ECM mycelia are particularly effective in the uptake of ammonium. The molecular bases of nitrate and ammonium uptake have been investigated, and transporters as well as assimilating enzymes have been characterized in ECM

fungi such as *Pisolithus*, *Laccaria*, and *Tuber* (Jargeat et al., 2003; Javelle et al., 2003a, 2003b; Montanini et al., 2003, 2006). Ammonium absorbed by ERMM, or derived from nitrate reduction, may be rapidly assimilated into amino acids, mostly glutamine, and subsequently incorporated into mycelial proteins or translocated to the host; glutamine is also considered as the main translocation form in ECM mycelia (Martin and Botton, 1993). In *Tuber*, the glutamine synthase gene is highly expressed both during fruit body ripening (Lacourt et al., 2002) and N starvation (Montanini et al., 2003). Nitrate and high-affinity ammonium transporters are differentially expressed, at least 1.5-fold, in response to N deprivation (Montanini et al., 2006). Recent studies of ECM and AM mycorrhiza have highlighted the direct transfer potential of ammonia from fungal to plant cells at the symbiotic interface (Chalot et al., 2006). Together with the expression of putative ammonium exporter genes in the ectomycorrhizal fungus *A. muscaria*, expression analysis of a high-affinity ammonium importer from *Populus tremula-tremuloides* has revealed that its expression is root-specific, is affected by N nutrition, and strongly increases in a N-independent manner upon ectomycorrhiza formation, thus suggesting that ammonium could act as a direct N source, which is delivered by the fungus in ECM symbiosis (Selle et al., 2005). In AM fungi, elegant isotope-labeling experiments in combination with gene expression data (Govindarajulu et al., 2005) have demonstrated that inorganic N (nitrate and ammonium) is taken up by the extraradical mycelium, incorporated into amino acids, and translocated to the intraradical mycelium, mainly as arginine. Nitrogen is then transferred from the fungus to the plant as ammonium, without any loss of C skeleton, thanks to the catabolic arm of the urea cycle, which releases ammonium from arginine. This proposed mechanism requires that the enzymes for N assimilation are expressed differently in the extraradical and the intraradical mycelium. Transcripts encoding enzymes that assimilate inorganic N have in fact been detected in the extraradical mycelium, while those involved in arginine breakdown are located in the intraradical fungal structures (Govindarajulu et al., 2005; Lopez-Pedrosa et al., 2006). Furthermore, in agreement with the idea that NH_3/NH_4^+ is thought to be the preferential form of N released by the fungus and uptaken by the plant (Chalot et al., 2006), a mycorrhizal-specific *Lotus AMT* gene has recently been identified (Guether et al., 2009b).

24.4.4.3 Carbon

As obligate biotrophs, AM fungi receive all of their carbohydrates from the host plant. Many ECM fungi instead likely acquire carbon (C) both via host photosynthesis and assimilation following the degradation of soil C polymers, thus possibly reducing the amount of C needed from the host plant (Girlanda et al., 2007; Martin et al., 2007). In vivo NMR, coupled with ^{13}C labeling, and transcriptome profiling studies, are beginning to unravel some secrets of C transport processes at the molecular level in mycorrhizal symbioses (Pfeffer et al., 1999; Bago et al., 2000, 2002; Lammers et al., 2001; Nehls et al., 2001; Ferrol et al., 2002; Jakobsen et al., 2002; Simard et al., 2002). In both AM and

ECM mycorrhiza, sucrose, the primary transport carbohydrate in plants, can only be used as a C source by fungi if hydrolyzed by plant cell wall-bound invertases. Mycorrhizal fungi partly convert it into fungal storage metabolites (such as lipids, trehalose, and glycogen) and partly transfer it to the extraradical mycelium (Lammers et al., 2001; Nehls, 2008; Parniske, 2008). Fungal hexose transporters have been identified in ectomycorrhizal fungi and the fungal hexose uptake capacity is remarkably increased in the Hartig net hyphae of the model fungi *A. muscaria* and *L. bicolor*. It has also been suggested that the plant might be involved in the regulation of the carbohydrate drain toward the fungus. A poplar monosaccharide transporter gene has been shown to be upregulated in ECM (Grunze et al., 2004), suggesting that root cells can compete with fungal hyphae for the hexoses present in the apoplast of the Hartig net region (Nehls, 2008). Instead, no fungal hexose transporter-coding gene has so far been characterized in AM fungi. A member of a novel clade of hexose transporters has been identified in the symbiotic fungus *Geosiphon pyriformis*, which associates to cyanobacteria. Due to its phylogenetical position inside Glomeromycota and the capacity to establish endosymbiosis acting as a host, it has been suggested that a similar transporter might be expressed in arbuscules and localized in the fungal interface membrane in AM fungi (Schüßler et al., 2006). However, the relationship between *Geosiphon* (the fungal host) and the bacteria (the symbiont) is topologically reverse compared to usual AMs. However, mycoheterotrophic plants associate not only to orchid fungi (Selosse and Roy, 2009) but also with AM fungi, suggesting that the plants are likely to receive C from the fungus (Winther and Friedman, 2008). This would require AM fungi to have C transporters that work in the efflux direction. Recently, Cameron et al. (2008) have demonstrated a bidirectional C flux for an orchid–mycorrhizal symbiosis, although with a net plant-to-fungus C flux. In addition, due to the rapidity of the bidirectional flux, a dynamic transfer at an interfacial apoplast has been suggested instead of the reliance on the so far hypothesized digestion of fungal pelotons (Cameron et al., 2008).

In conclusion, although details of molecular mechanisms for mineral nutrients and C exchange between AM partners have emerged rapidly from genetic studies on nutrient transport and metabolic pathways, the full dissection of host–fungal interactions is still problematic. Consistent with their extensive genetic variability, mycorrhizal associations have different functional abilities and, therefore, offer distinct advantages to the host plant. Some mycorrhizal fungi are particularly effective in scavenging organic N, and are associated with plants for which the acquisition of N is crucial (Peter et al., 2001); others, instead, are more effective at P uptake and transport.

24.4.5 Perspectives

Mycorrhizal fungi also play an important role in nutrient cycles and in plant protection against pathogens and environmental stresses. It is therefore important to know and understand the role played by this important category of microorganisms. This can be obtained through the study of their biodiversity in land ecosystems and their role in ecological interactions, as well described in van der Heijden and Sanders (2002) and Finlay (2008). In spite of the enormous steps that have been made, thanks to molecular biology, the *databases* that gather the sequences of the fungal species, including mycorrhizal ones, are still not sufficiently developed to allow a complete analysis of soil communities. In order to overcome the limitations connected to the impossibility, in many cases, of growing fungi in cultures and also in order to obtain an overall vision of the fungal community, ecologists have begun to use an approach called environmental genomics or metagenomics (Chivian et al., 2008; Hugenholtz and Tyson, 2008). The term metagenomics refers to the idea that the total DNA extracted from an environmental sample represents the whole microorganism community and can be considered as a single genome (metagenome or megagenome). This approach promotes the study of communities of organisms found in nature without the cultivation and isolation of individual species. In this way, the majority of organisms existing in nature can be characterized. A further advancement in these studies is based on pyrosequencing, a new generation of high-throughput platforms. An approach based on 454 GS-FLX pyrosequencing has recently been used to analyze mycorrhizal communities in different soils in the Mediterranean and temperate forest environments (Buée et al., 2009; Öpik et al., 2009; Lumini et al., 2010). In future, investigations will take advantage of the new 454-FLX-titanium chemistry that allows longer read sequences to be obtained than with 454 GS-FLX. The Genome Sequencer (GS) FLX Titanium is an ultra high-throughput tool that can be used to generate 1 million sequencing readings with an average reading of about 500 nucleotides in a single operation run. This will allow detection of unknown mycorrhizal communities in molecular ecological researches and allow the dream of all soil microbiologists to be true, that is, of detecting all the fungi in a soil and of understanding their function.

Acknowledgments

References

Abuzinadah, R.A., and D.J. Read. 1989. The role of proteins in the nitrogen nutrition of ectomycorrhizal plants. V. Nitrogen transfer in birch (*Betula pendula*) grown in association with mycorrhizal and nonmycorrhizal fungi. New Phytol. 112:61–68.

Anderson, I.C., and J.W.G. Cairney. 2007. Ectomycorrhizal fungi: Exploring the mycelial frontier. FEMS Microbiol. Rev. 31:388–406.

Aroca, R., R. Porcel, and J.M. Ruiz-Lozano. 2007. How does arbuscular mycorrhizal symbiosis regulate root hydraulic properties and plasma membrane aquaporins in *Phaseolus vulgaris* under drought, cold or salinity stresses? New Phytol. 173:808–816.

Bago, B., P.E. Pfeffer, and Y. Shachar-Hill. 2000. Carbon metabolism and transport in arbuscular mycorrhizas. Plant Physiol. 124:949–957.

Bago, B., W. Zipfel, R.M. Williams, J. Jun, R. Arreola, P.J. Lammers, P.E. Pfeffer, and Y. Sachar-Hill. 2002. Translocation and utilization of fungal storage lipid in the arbuscular mycorrhizal symbiosis. Plant Physiol. 128:108–124.

Balestrini, R., and P. Bonfante. 2005. The interface compartment in arbuscular mycorrhizae: A special type of plant cell wall? Plant Biosyst. 139:8–15.

Balestrini, R., J. Gomez-Ariza, L. Lanfranco, and P. Bonfante. 2007. Laser microdissection reveals that transcripts for five plant and one fungal phosphate transporter genes are contemporaneously present in arbusculated cells. Mol. Plant-Microbe Interact. 20:1055–1062.

Balestrini, R., M.G. Hahn, and P. Bonfante. 1996. Location of cell-wall components in ectomycorrhizae of *Corylus avellana* and *Tuber magnatum*. Protoplasma 191:55–69.

Barker, S.J., D. Tagu, and G. Delp. 1998. Regulation of root and fungal morphogenesis in mycorrhizal symbioses. Plant Physiol. 116:1201–1207.

Benedetto, A., F. Magurno, P. Bonfante, and L. Lanfranco. 2005. Expression profiles of a phosphate transporter gene (*GmosPT*) from the endomycorrhizal fungus *Glomus mosseae*. Mycorrhiza 15:620–627.

Berbee, M.L., and J.W. Taylor. 2001. Fungal molecular evolution: Gene trees and geologic time, p. 229–245. *In* D. McLaughlin, E. McLaughlin, and P. Lemke (eds.) The Mycota VII, Part B, Systematics and evolution. Springer Verlag, Berlin, Germany.

Bergero, R., S. Perotto, M. Girlanda, G. Vidano, and A.M. Luppi. 2000. Ericoid mycorrhizal fungi are common root associates of a Mediterranean ectomycorrhizal plant (*Quercus ilex*). Mol. Ecol. 9:1639–1649.

Bonfante, P. 2001. At the interface between mycorrhizal fungi and plants: The structural organization of cell wall, plasma membrane and cytoskeleton, p. 45–61. *In* K. Esser and B. Hock (eds.) The Mycota IX. Springer-Verlag, Berlin, Germany.

Bonfante, P., R. Balestrini, A. Genre, and L. Lanfranco. 2009. Establishment and functioning of arbuscular mycorrhizas, p. 259–274. *In* H.B. Deising (ed.) Plants relationship. 2nd edn. The Mycota V. Springer Verlag, Berlin, Germany.

Bonfante, P., and A. Genre. 2008. Plants and arbuscular mycorrhizal fungi: An evolutionary-developmental perspective. Trends Plant Sci. 13:492–498.

Bonfante, P., and A. Genre. 2010. Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. Nat. Commun. 1:48. doi:10.1038/ncomms1046.

Brundrett, M.C. 2002. Coevolution of roots and mycorrhizas of land plants. New Phytol. 154:275–304.

Bucher, M. 2007. Functional biology of plant phosphate uptake at root and mycorrhiza interfaces. New Phytol. 173:11–26.

Buée, M., M. Reich, C. Murat, E. Morin, R.H. Nilsson, S. Uroz, and F. Martin. 2009. 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. New Phytol. 184:449–456.

Cameron, D.D., I. Johnson, D.J. Read, and J.R. Leake. 2008. Giving and receiving: Measuring the carbon cost of mycorrhizas in the green orchid, *Goodyera repens*. New Phytol. 180:176–184.

Cappellazzo, G., L. Lanfranco, M. Fitz, D. Wipf, and P. Bonfante. 2008. Characterization of an amino acid permease from the endomycorrhizal fungus *Glomus mosseae*. Plant Physiol. 147:429–437.

Chalot, M., D. Blaudez, and A. Brun. 2006. Ammonia: A candidate for nitrogen transfer at the mycorrhizal interface. Trends Plant Sci. 11:263–266.

Chalot, M., A. Javelle, D. Blaudez, R. Lambilliote, R. Cooke, H. Sentenac, D. Wipf, and B. Botton. 2002. An update on nutrient transport processes in ectomycorrhizas. Plant Soil 244:165–175.

Chambers, S.M., P.G. Williams, R.D. Seppelt, and J.W.G. Cairney. 1999. Molecular identification of *Hymenoscyphus* sp. from rhizoids of the leafy liverwort *Cephaloziella exiliflora* in Australia and Antarctica. Mycol. Res. 103:286–288.

Chivian, D., E.L. Brodie, E.J. Alm, D.E. Culley, P.S. Dehal, T.Z. DeSantis, T.M. Gihring et al. 2008. Environmental genomics reveals a single-species ecosystem deep within earth. Science 322:275–278.

Corradi, N., M. Hijri, L. Fumagalli, and I.R. Sanders. 2004. Arbuscular mycorrhizal fungi (Glomeromycota) harbour ancient fungal tubulin genes that resemble those of the chytrids (Chytridiomycota). Fungal Genet. Biol. 41:1037–1045.

Croll, D., M. Giovannetti, A.M. Koch, C. Sbrana, M. Ehinger, P.J. Lammers, and I.R. Sanders. 2008. Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus *Glomus intraradices*. New Phytol. 181:924–937.

Croll, D., and I.R. Sanders. 2009. Recombination in *Glomus intraradices*, a supposed ancient asexual arbuscular mycorrhizal fungus. BMC Evol. Biol. 9:13.

Cullings, K.W., D.R. Vogler, V.T. Parker, and S.K. Finley. 2000. Ectomycorrhizal specificity patterns in a mixed *Pinus contorta* and *Picea engelmannii* forest in Yellowstone National Park. Appl. Environ. Microbiol. 66:4988–4991.

Da Silva, G.A., E. Lumini, L. Costa Maia, P. Bonfante, and V. Bianciotto. 2006. Phylogenetic analysis of Glomeromycota by partial LSU rDNA sequences. Mycorrhiza 16:183–189.

Ferrol, N., M.J. Pozo, M. Antelo, and C. Azcón-Aguilar. 2002. Arbuscular mycorrhizal symbiosis regulates plasma membrane H$^+$-ATPase gene expression in tomato plants. J. Exp. Bot. 53:1638–1687.

Finlay, R.D. 2008. Ecological aspects of mycorrhizal symbiosis: With special emphasis on the functional diversity of interactions involving the extraradical mycelium. J. Exp. Bot. 59:1115–1126.

Girlanda, M., S. Perotto, and P. Bonfante. 2007. Mycorrhizal fungi: Their habitats and nutritional strategies biology of the fungal cell, p. 223–249. *In* R.J. Howard and N.A.R. Gow (eds.) The Mycota IV. 2nd edn. Springer-Verlag, Berlin, Germany.

Glassop, D., S.E. Smith, and F.W. Smith. 2005. Cereal phosphate transporters associated with the mycorrhizal pathway of phosphate uptake into roots. Planta 222:688–698.

Govindarajulu, M., P.E. Pfeffer, H. Jin, J. Abubaker, D.D. Douds, J.W. Allen, H. Bücking, P.J. Lammers, and Y. Shachar-Hill. 2005. Nitrogen transfer in the arbuscular mycorrhizal symbiosis. Nature 435:819–823.

Grunze, N., M. Willmann, and U. Nehls. 2004. The impact of ectomycorrhiza formation on monosaccharide transporter gene expression in poplar roots. New Phytol. 164:147–155.

Guether, M., R. Balestrini, M. Hannah, J. He, M. Udvardi, and P. Bonfante. 2009a. Genome-wide reprogramming of regulatory networks, transport, cell wall and membrane biogenesis during arbuscular mycorrhizal symbiosis in *Lotus japonicus*. New Phytol. 182:200–212.

Guether, M., B. Neuhaeuser, R. Balestrini, M. Dynowski, U. Ludewig, and P. Bonfante. 2009b. *LjAMT2;2* a mycorrhizal-specific ammonium transporter from *Lotus japonicus* acquires nitrogen released by arbuscular mycorrhizal fungi. Plant Physiol. 150:73–83.

Harrison, M.J., G.R. Dewbre, and J. Liu. 2002. A phosphate transporter from *Medicago truncatula* involved in the acquisition of phosphate released by arbuscular mycorrhizal fungi. Plant Cell 14:2413–2429.

Harrison, M.J., and M.L. van Buuren. 1995. A phosphate transporter from the mycorrhizal fungus *Glomus versiforme*. Nature 378:626–629.

Heckman, D.S., D.M. Geiser, B.R. Eidell, R.L. Stauffer, N.L. Kardos, and S.B. Hedges. 2001. Molecular evidence for the early colonization of land by fungi and plants. Science 293:1129–1133.

Helgason, T., I.J. Watson, and J.P.W. Young. 2003. Phylogeny of the Glomerales and Diversisporales (Fungi: Glomeromycota) from actin and elongation factor 1-alpha sequences. FEMS Microbiol. Lett. 229:127–132.

Hibbett, D.S., L.B. Gilbert, and M. Donoghue. 2000. Evolutionary instability of ectomycorrhizal symbiosis in basidiomycetes. Nature 407:506–508.

Hodge, A., C.D. Campbell, and A.H. Fitter. 2001. An arbuscular mycorrhizal fungus accelerates decomposition and acquires nitrogen directly from organic material. Nature 413:297–299.

Horton, T.R., and T.D. Bruns. 1998. Multiple host fungi are the most frequent and abundant ectomycorrhizal types in a mixed stand of Douglas fir (*Pseudotsuga menziesii* (Mirb.) Franco) and bishop pine (*Pinus muricata* D. Don). New Phytol. 139:331–339.

Horton, T.R., T.D. Bruns, and V.T. Parker. 1999. Ectomycorrhizal fungi associated with *Arctostaphylos* contribute to *Pseudotsuga menziesii* establishment. Can. J. Bot. 77:93–102.

Hugenholtz, P., and G.W. Tyson. 2008. Microbiology: Metagenomics. Nature 455:481–483.

Jakobsen, I., S.E. Smith, and F.A. Smith. 2002. Function and diversity of arbuscular mycorrhizae in carbon and mineral nutrition, p. 75–92. *In* M.G.A. van der Heijden and I.R. Sanders (eds.) Mycorrhizal ecology. Springer-Verlag, Berlin, Germany.

Jargeat, P., D. Rekangalt, M.C. Verner, G. Gay, J.C. Debaud, R. Marmeisse, and L. Fraissinet-Tachet. 2003. Characterisation and expression analysis of a nitrate transporter and nitrite reductase genes, two members of a gene cluster for nitrate assimilation from the symbiotic basidiomycete *Hebeloma cylindrosporum*. Curr. Genet. 43:199–205.

Javelle, A., B. André, A.M. Marini, and M. Chalot. 2003a. High affinity ammonium transporters and nitrogen sensing in mycorrhizas. Trends Microbiol. 11:53–55.

Javelle, A., M. Morel, B.R. Rodriguez-Pastrana, B. Botton, B. Andre, A.M. Marini, A. Brun, and M. Chalot. 2003b. Molecular characterization, function and regulation of ammonium transporters (Amt) and ammonium-metabolizing enzymes (GS, NADP–GDH) in the ectomycorrhizal fungus *Hebeloma cylindrosporum*. Mol. Microbiol. 47:411–430.

Javot, H., N. Pumplin, and M.J. Harrison. 2007. Phosphate in the arbuscular mycorrhizal symbiosis: Transport properties and regulatory roles. Plant Cell Environ. 30:310–322.

Julou, T., B. Burghardt, G. Gebauer, D. Berveiller, C. Damesin, and M.-A. Selosse. 2005. Mixotrophy in orchids: Insights from a comparative study of green individuals and nonphotosynthetic mutants of *Cephalanthera damasonium*. New Phytol. 166:639–653.

Lacourt, I., S. Duplessis, S. Abba, P. Bonfante, and F. Martin. 2002. Isolation and characterization of differentially expressed genes in the mycelium and fruit body of *Tuber borchii*. Appl. Environ. Microb. 68:5788–5788.

Lammers, P.J., J. Jun, J. Abubaker, R. Arreola, A. Gopalan, B. Bago, C. Hernandez-Sebastia, J.W. Allen, D.D. Douds, P.E. Pfeffer, and Y. Shachar-Hill. 2001. The glyoxylate cycle in an arbuscular mycorrhizal fungus: Gene expression and carbon flow. Plant Physiol. 127:1287–1298.

Lee, J., and J.P.W. Young. 2009. The mitochondrial genome sequence of the arbuscular mycorrhizal fungus *Glomus intraradices* isolate 494 and implications for the phylogenetic placement of *Glomus*. New Phytol. 183:200–211.

Leigh, J., A. Hodge, and A.H. Fitter. 2008. Arbuscular mycorrhizal fungi can transfer substantial amounts of nitrogen to their host plant from organic material. New Phytol. 181:199–207.

Lopez-Pedrosa, A., M. Gonzalez-Guerrero, A. Valderas, C. Azcon-Aguilar, and N. Ferrol. 2006. *GintAMT1* encodes a functional high-affinity ammonium transporter that is expressed in the extraradical mycelium of *Glomus intraradices*. Fungal Genet. Biol. 43:102–110.

Lumini, E., A. Orgiazzi, R. Borriello, P. Bonfante, and V. Bianciotto. 2010. Disclosing AMF biodiversity in soil through a land-use gradient using a pyrosequencing. Environ. Microbiol. 12:2165–2179.

Ma, M., T.K. Tan, and S.M. Wong. 2003. Identification and molecular phylogeny of *Epulorhiza* isolates from tropical orchids. Mycol. Res. 107:1041–1049.

Maldonado-Mendoza, I.E., G.R. Dewbre, and M.J. Harrison. 2001. A phosphate transporter gene from the extra-radical mycelium of an arbuscular mycorrhizal fungus *Glomus intraradices* is regulated in response to phosphate in the environment. Mol. Plant-Microbe Interact. 14:1140–1148.

Martin, F. 2007. Fair trade in the underworld: The ectomycorrhizal symbiosis, p. 289–306. *In* R.J. Howard and N.A.R. Gow (eds.) The Mycota VIII. Biology of the fungal cell. Springer-Verlag, Berlin, Germany.

Martin, F., A. Aerts, D. Ahrén, A. Brun, E.G.J. Danchin, F. Duchaussoy, J. Gibon et al. 2008a. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. Nature 452:42–43.

Martin, F., and B. Botton. 1993. Nitrogen metabolism of ectomycorrhizal fungi and ectomycorrhiza. Adv. Plant. Pathol. 9:83–102.

Martin, F., V. Gianinazzi-Pearson, M. Hijri, P. Lammers, N. Requena, I.R. Sanders, Y. Shachar-Hill, H. Shapiro, G.A. Tuskan, and J.P.W. Young. 2008b. The long hard road to a completed *Glomus intraradices* genome. *Glomus* Genome Consortium (GGC) Symposium, Nancy, France, September, 2008. New Phytol. 180:747–750.

Martin, F., A. Kohler, C. Murat, R. Balestrini, P.M. Coutinho, O. Jaillon, B. Montanini et al. 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature 464:1033–1038.

Martin, F., and U. Nehls. 2009. Harnessing ectomycorrhizal genomics for ecological insights. Curr. Opin. Plant Biol. 12:508–515.

Martin, F., S. Perotto, and P. Bonfante. 2007. Mycorrhizal fungi: A fungal community at the interface between soil and roots, p. 201–228. *In* R. Pinton, Z. Varanini, and P. Nannipieri (eds.) The rhizosphere: Biochemistry and organic substances at the soil–plant interface. CRC Press, Boca Raton, FL.

Martino, E., S. Perotto, R. Parsons, and G.M. Gadd. 2003. Solubilization of insoluble inorganic zinc compounds by ericoid mycorrhizal fungi derived from heavy metal polluted sites. Soil Biol. Biochem. 35:133–141.

McLean, C.B., J.H. Cunnington, and A.C. Lawrie. 1999. Molecular diversity within and between ericoid endophytes from the Ericaceae and Epacridaceae. New Phytol. 144:351–358.

Molina, R., H. Massicotte, and J.M. Trappe. 1992. Specificity phenomena in mycorrhizal symbiosis: Community ecological consequences, practical implications, p. 357–423. *In* M.F. Allen (ed.) Mycorrhizal functioning: An integrated plant-fungal process. Chapman and Hall, London, U.K.

Molina, R., and J.M. Trappe. 1982. Patterns of ectomycorrhizal host specificity and potential among Pacific Northwest conifers and fungi. For. Sci. 28:423–458.

Monreal, M., S.M. Berch, and M. Berbee. 1999. Molecular diversity of ericoid mycorrhizal fungi. Can. J. Bot. 77:1580–1594.

Montanini, B., M. Betti, A.J. Marquez, R. Balestrini, P. Bonfante, and S. Ottonello. 2003. Distinctive properties and expression profiles of glutamine synthetase from a plant symbiotic fungus. Biochem. J. 373:357–368.

Montanini, B., A.R. Viscomi, A. Bolchi, Y. Martin, J.M. Siverio, R. Balestrini, P. Bonfante, and S. Ottonello. 2006. Functional properties and differential mode of regulation of the nitrate transporter from a plant symbiotic ascomycete. Biochem. J. 394:125–134.

Nagy, R., V. Karandashov, V. Chague, K. Kalinkevich, M. Tamasloukht, G. Xu, I. Jakobsen, A.A. Levy, N. Amrhein, and M. Bucher. 2005. The characterization of novel mycorrhiza-specific phosphate transporters from *Lycopersicon esculentum* and *Solanum tuberosum* uncovers functional redundancy in symbiotic phosphate transport in solanaceous species. Plant J. 42:236–250.

Näsholm, T., and J. Persson. 2001. Plant acquisition of organic nitrogen in boreal forests. Physiol. Plant. 111:419–426.

Nehls, U. 2008. Mastering ectomycorrhizal symbiosis: The impact of carbohydrates. J. Exp. Bot. 59:1097–1108.

Nehls, U., A. Bock, M. Ecke, and R. Hampp. 2001. Differential expression of hexose-regulated fungal genes *Am-PAL* and *AmMst1* within *Amanita*/*Populus* ectomycorrhizas. New Phytol. 150:583–589.

Nehls, U., M. Ecke, and R. Hampp. 1999. Sugar- and nitrogen-dependent regulation of an *Amanita muscaria* phenylalanine ammonium lyase gene. J. Bacteriol. 181:1931–1933.

Öpik, M., M. Metsis, T.J. Daniell, M. Zobel, and M. Moora. 2009. Large-scale parallel 454 sequencing reveals host ecological group specificity of arbuscular mycorrhizal fungi in a boreonemoral forest. New Phytol. 184:424–437.

Otero, J.T., J.D. Ackerman, and P. Bayman. 2002. Diversity and host specificity of endophytic Rhizoctonia-like fungi from tropical orchids. Am. J. Bot. 89:1852–1858.

Pain, A., and C. Hertz-Fowler. 2008. Genomic adaptation: A fungal perspective. Nat. Rev. Microbiol. 6:572–573.

Parniske, M. 2008. Arbuscular mycorrhiza: The mother of plant root endosymbioses. Nature Rev. Microbiol. 6:763–775.

Paszkowski, U., S. Kroken, C. Roux, and S.P. Briggs. 2002. Rice phosphate transporters include an evolutionarily divergent gene specifically activated in arbuscular mycorrhizal symbiosis. Proc. Natl. Acad. Sci. USA 99:13324–13329.

Perotto, S., M. Girlanda, and E. Martino. 2002. Ericoid mycorrhizal fungi: Some new perspectives on old acquaintances. Plant Soil 244:41–53.

Peter, M., F. Ayer, and S. Egli. 2001. Nitrogen addition in a Norway spruce stand altered macromycete sporocarp production and below-ground ectomycorrhizal species composition. New Phytol. 149:311–325.

Peterson, R.L., H.B. Massicotte, and L.H. Melville. 2004. Mycorrhizas: Anatomy and cell biology. CABI, Toronto, Canada.

Pfeffer, P.E., D.D. Douds, G. Becard, and Y. Shachar-Hill. 1999. Carbon uptake and the metabolism and transport of lipids in arbuscular mycorrhiza. Plant Physiol. 120: 587–598.

Phipps, C.J., and T.N. Taylor. 1996. Mixed arbuscular mycorrhizae from the Triassic of Antarctica. Mycologia 88: 707–714.

Rausch, C., P. Daram, S. Brunner, J. Jansa, M. Laloi, G. Leggewie, N. Amrhein, and M. Bucher. 2001. A phosphate transporter expressed in arbuscule-containing cells in potato. Nature 414:462–466.

Read, D.J. 1991. Mycorrhizas in ecosystems. Experientia 47:376–391.

Read, D.J., J.G. Ducket, R. Francis, R. Ligrone, and A. Russell. 2000. Symbiotic fungal associations in "lower" land plants. Philos. Trans. R. Soc. London Ser. B 355:815–830.

Read, D.J., J.R. Leake, and J. Perez-Moreno. 2004. Mycorrhizal fungi as drivers of ecosystem processes in heathland and boreal forest biomes. Can. J. Bot. 82:1243–1263.

Read, D.J., and J. Perez-Moreno. 2003. Mycorrhizas and nutrient cycling in ecosystems—A journey towards relevance? New Phytol. 157:475–492.

Redecker, D., R. Kodner, and L.E. Graham. 2000. Glomalean fungi from the Ordovician. Science 289:1920–1921.

Roberts, P. 1999. Rhizoctonia-forming fungi. A taxanomic guide. Royal Botanic Gardens, Kew, U.K.

Schüßler, A., H. Martin, D. Cohen, M. Fitz, and D. Wipf. 2006. Characterization of a carbohydrate transporter from symbiotic glomeromycotan fungi. Nature 444:933–936.

Schüßler, A., D. Schwarzott, and C. Walker. 2001. A new fungal phylum, the Glomeromycota: Phylogeny and evolution. Mycol. Res. 105:1413–1421.

Selle, A., M. Willmann, N. Grinze, A. Geßler, M. Weiß, and U. Nehls. 2005. The high-affinity poplar ammonium importer *PttAMT1.2* and its role in ectomycorrhizal symbiosis. New Phytol. 168:697–706.

Selosse, M.A., A. Faccio, G. Scappaticci, and P. Bonfante. 2004. Chlorophyllous and achlorophyllous specimens of *Epipactis microphylla* (Neottieae, Orchidaceae) are associated with ectomycorrhizal septomycetes, including truffles. Microb. Ecol. 47:416–426.

Selosse, M.A., and F. Le Tacon. 1998. The land flora: A phototroph–fungus partnership? Trends Ecol. Evol. 13:15–20.

Selosse, M.A., and M. Roy. 2009. Green plants that feed on fungi: Facts and questions about mixotrophy. Trends Plant Sci. 14:64–70.

Sharples, J.M., S.M. Chambers, A.A. Meharg, and J.W.G. Cairney. 2000. Genetic diversity of root-associated fungal endophytes from *Calluna vulgaris* at contrasting field sites. New Phytol. 148:153–162.

Simard, S.W., D. Durall, and M. Jones. 2002. Carbon and nutrient fluxes within and between mycorrhizal plants, p. 33–74. *In* M.G.A. van der Heijden and I.R. Sanders (eds.) Mycorrhizal ecology. Springer-Verlag, Berlin, Germany.

Smith, S.E., and D.J. Read. 2008. Mycorrhizal symbiosis. 3rd edn. Academic Press, London, U.K.

Smith, S.E., and F.A. Smith. 1990. Structure and function of the interfaces in biotrophic symbioses as they relate to nutrient transport. New Phytol. 114:1–38.

Suarez, J.P., M. Weiss, A. Abele, S. Garnica, F. Oberwinkler, and I. Kottke. 2006. Diverse tulasnelloid fungi form mycorrhizas with epiphytic orchids in an Andean cloud forest. Mycol. Res. 110:1257–1270.

Taylor, T.N., W. Remy, H. Hass, and H. Kerp. 1995. Fossil arbuscular mycorrhizae from the early Devonian. Mycologia 87:560–573.

Taylor, T.N., and E.L. Taylor. 1997. The distribution and interactions of some paleozoic fungi. Rev. Palaeobot. Palynol. 95:83–94.

Tedersoo, L., K. Hansen, B.A. Perry, and R. Kjøller. 2006. Molecular and morphological diversity of pezizalean ectomycorrhiza. New Phytol. 170:581–596.

Tehler, A., J.S. Farris, D.L. Lipscomb, and M. Källersjö. 2000. Phylogenetic analyses of the fungi based on large rDNA data sets. Mycologia 92:459–474.

Trappe, J.M. 1962. Fungus associates of ectotrophic mycorrhizae. Bot. Rev. 28:538–606.

van der Heijden, M.G.A. 2002. Arbuscular mycorrhizal fungi as a determinant of plant diversity: In search for underlying mechanisms and general principles, p. 243–266. *In* M.G.A. van der Heijden and I.R. Sanders (eds.) Mycorrhizal ecology. Springer-Verlag, Berlin, Germany.

van der Heijden, M.G.A., and I.R. Sanders. 2002. Mycorrhizal ecology. Springer-Verlag, Berlin, Germany.

van der Heijden, M.G.A., T.R. Scheublin, and A. Brader. 2004. Taxonomic and functional diversity in arbuscular mycorrhizal fungi—Is there any relationship? New Phytol. 164:201–204.

Vrålstad, T., T. Fossheim, and T. Schumacher. 2000. *Piceirhiza bicolorata*—The ectomycorrhizal expression of the *Hymenoscyphus ericae* aggregate? New Phytol. 145:549–563.

Waterman, R.J., and M.I. Bidartondo. 2008. Deception above, deception below: Linking pollination and mycorrhizal biology of orchids. J. Exp. Bot. 59:1085–1096.

Winther, J.L., and W.E. Friedman. 2008. Arbuscular mycorrhizal associations in Lycopodiaceae. New Phytol. 177:790–801.

Wipf, D., M. Benjdia, M. Tegeder, and W.B. Frommer. 2002. Characterization of a general amino acid permease from *Hebeloma cylindrosporum*. FEBS Lett. 528:119–124.

Young, J.P.W. 2009. Kissing cousins: Mycorrhizal fungi get together. New Phytol. 181:751–753.

25

Soil Fauna

Michael Bonkowski
University of Cologne

M.A. Callaham, Jr.
United States Department of Agriculture

Marianne Clarholm
Swedish University of Agricultural Sciences

David C. Coleman
University of Georgia

D.A. Crossley, Jr.
University of Georgia

Bryan Griffiths
Teagasc, Environment Research Centre

Paul F. Hendrix
University of Georgia

Robert McSorley
University of Florida

Mark G. St. John
Landcare Research

P.C.J. van Vliet
Blgg AgroXpertus

25.1 Protozoa

Bryan Griffiths

Marianne Clarholm

Michael Bonkowski

25.1.1 Introduction

Protozoa belong to the "Protista," which is a paraphyletic group consisting of those eukaryotes that are not animals, true fungi, or green plants. The group is also defined on the basis of the absence of characters (i.e., no complex development from embryos, no extensive cell differentiation) and thus includes widely ranging microbes, including slime molds, algae, and protozoa (Clarholm et al., 2007). Protozoa in soil, which are active in water-filled pores and water films in larger pores, can be generally (or basically) classified in four groups: flagellates, ciliates, and naked and testate amoebae. Provided the appropriate sampling techniques are used, one can find protozoa in virtually every conceivable soil habitat. In harsh environments with low primary production, small forms with low nutrient demands normally dominate, whereas in richer environments their diversity, trophic complexity, and size tend to increase.

25.1.2 Background Biology

Flagellates are evolutionarily the oldest group. Those with one flagellum belong to the evolutionary lineage (unikonts) from which also amoebozoa, fungi, and animals are derived. Those with two (and more) flagella belong to a different evolutionary lineage (bikonts) from which, for example, plants are derived.

Most members of this group are small (4–15 μm body length). The smallest flagellates can squeeze through a 3 μm aperture, as a result of their flexible outer cell membrane.

Flagellates can swim by using up to eight flagella, which are long relative to the length of their body. Flagella may be used not only for swimming or anchoring but also to create a feeding current for catching bacterial cells. Although generally recognized as feeders of suspended bacteria, some important soil species (e.g., cercomonads) need to be surface associated to feed (Kiorboe et al., 2004).

Ciliates are evolutionarily the youngest group of protozoa, and in this group, we find the largest soil protozoa (20–600 μm body length). Ciliates have a more complicated structure than flagellates and have fur-like cilia that coat the whole body surface in distinct patterns that are used in classic taxonomy.

Amoebae have a polyphyletic origin and an uncertain evolutionary history. The general term amoeba, therefore, represents organisms with a similar type of organization of the cell, rather than a taxonomic unit or monophyletic group. Amoebae fall generally into two fundamental different taxonomic lineages: the Amoebozoa with broad pseudopodia and the Rhizaria with fine and slender pseudopodia (i.e., filopodia). Testate amoebae (Figure 25.1) have shells (tests), which are rigid constructions that can be proteinaceous, siliceous, or calcareous. The Arcellinida with broad pseudopodia and tests often composed of inorganic material bound into an organic matrix are generally larger (50–200 μm, and up to 300 μm) whereas the testate amoeba species with filopodia are generally smaller (10–100 μm). Testate amoebae usually have much longer generation times than naked amoebae. The body of the testate amoeba

always stays partly inside its test, whether in a resting cyst or in an active state. When moving around, the active individual stretches its pseudopodia out through the aperture to "walk" and feed on surfaces. Testate amoebae are found in substantial (approximately 30,000 g^{-1} dry wt) numbers in wet and acidic areas like bogs, mires, and the litter layers of coniferous forests. Possibly, the ecological niche for this group has been enlarged in eutrophied, acidified European forest soils, where the pH has decreased quickly with increasing air pollution. In these ecosystems, high numbers of testate amoebae are reported to have a significant impact when modeling nutrient turnover (Schröter et al., 2003). Naked amoebae are commonly 15–100 μm wide but only 1 μm thick when attached to a surface. Active individuals have a loosely shaped, jelly-lump-like body, which is strongly light quenching when observed in a light microscope. Members of this group generally have few characteristic morphological features seen by light microscopy. However, naked amoebae are not shapeless and their various morphotypes are used in classic taxonomy (Smirnov and Brown, 2004). An amoeba will move by changing its cytoplasm between a gel and sol form, "floating" along the surface or around a bacterial colony by pushing out pseudopodia in front while retracting pseudopodia behind it.

25.1.3 Distribution

Many surveys of protozoan morphospecies on a global scale have suggested a ubiquitous dispersal (Finlay, 2002), which would imply that global protozoan species richness would be relatively low, as the diversity of microbial species found in one

HV	Spot	Mag	WD	Det.	8/31/2008	Pressure	400.0 μm
20.0 kV	4.0	200×	13.3 mm	LFD	12:05:38 PM	1.00 mbar	

(a)

HV	Spot	Mag	WD	Det.	8/31/2008	Pressure	100.0 μm
20.0 kV	4.0	840×	13.3 mm	LFD	12:06:48 PM	1.00 mbar	
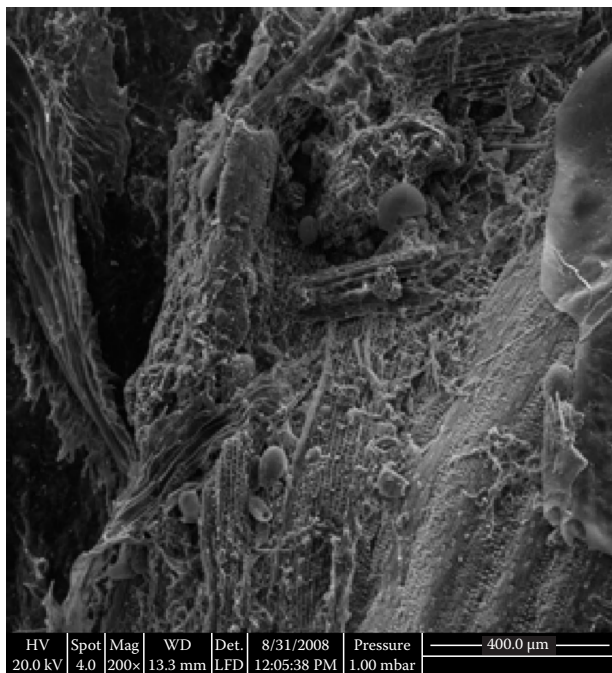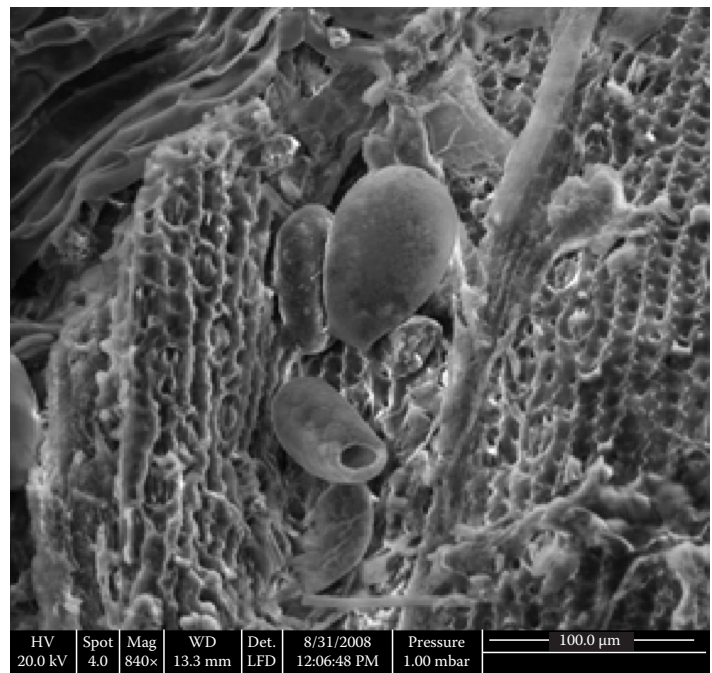
(b)

FIGURE 25.1   Testate amoebae on decomposing leaf litter in overview (a) and in detail (b). Note scale bar at the bottom of each scanning electron micrograph. (Courtesy of Michael Bonkowski.)

locality generally represents global levels of diversity within that habitat. An alternative moderate endemicity model (Foissner, 2006) suggests that some protozoan morphospecies have cosmopolitan distributions, while other protozoa have restricted distributions. A molecular screening study of cercomonads (heterotrophic flagellates, 5–50 μm long, among the most prominent bacterivorous protozoa) confirmed that (1) the diversity of 18S rDNA genotypes (18S types) within morphospecies can be very high, (2) many 18S types have cosmopolitan distributions, and (3) some 18S types within morphospecies have ecologically and/or geographically restricted distributions (Bass et al., 2007). That study also concluded that the number of protozoan species is hugely underestimated. Organisms with restricted distributions may prove to be important indicators of environmental change.

25.1.4 Feeding and Role in Food Webs

Most soil protozoa are bacterial feeders, collecting bacteria using some kind of catching device. This can be a flagellum (in the case of flagellates), an oral groove, or collar of specialized cilia in the ciliates and specific feeding cups in the amoebae. With food in ample supply, the bacterial remains will be excreted after only 30 min. This means that only the most palatable cytoplasmic parts of the bacteria will be used by the protozoa, while the remaining organic matter of good quality is delivered into the soil environment and to take part in the soil nutrient cycle (Griffiths, 1994). Flagellates, ciliates, and naked amoebae all have short generation times, in the order of hours, when food is amply available. Testate amoebae have longer generation times, where a doubling takes several days even with a good food supply (Clarholm et al., 2007).

Protozoa show feeding preferences, preferring bacteria without pigments, poisons, filaments, or external slimes; and, generally, Gram-negative bacteria are preferred to Gram-positive ones. To gain access to other food sources, unicellular yeasts and small fungal spores are phagocytized, while fungal hyphae, large fungal spores, and filamentous algae are enzymatically perforated by some amoebae and ciliates. Some ciliates attach a "feeding tube" to the hypha or spore to make a hole and to suck out the cytoplasm. Predatory protozoa can also feed on other protists and even nematodes, which are not always smaller than themselves.

In soil, protozoa are an important link in the food web (De Ruiter et al., 2002). By feeding on bacteria and, to a lesser extent, fungi, and then being consumed by larger fauna (e.g., individually by nematodes, mites, Collembola; and together with soil and organic matter by enchytraeids and earthworms), protozoa facilitate the flow of carbon and nutrients in the soil. The strength of a trophic interaction (i.e., the effects of the populations on each other) is important to the stability of the food web and interaction strengths for soil protozoa are particularly high (De Ruiter et al., 2002).

Modeling the soil food web can indicate what the most likely outcomes of various scenarios are, as has been applied to the effects of climate change and loss of biodiversity (Hunt and Wall, 2002). In their scenarios, increasing atmospheric CO_2 concentration (which was hypothesized to lead to plant N limitation) led to changes in the soil food web, including protozoa, which resulted in greater N mineralization and no N limitation. When the effects of deleting individual functional groups from the food web was modeled, the removal of amoebae led to a 60% decrease in omnivorous nematodes (a direct effect as they feed on amoebae) and 20%–30% increases in bacterial-feeding nematodes and predatory microarthropods (indirect effects as the nematodes no longer compete with amoebae for prey and the microarthropods feed on the nematodes). These results indicate the importance of compensatory mechanisms operating in the soil food web.

25.1.5 Methodology

All available evidence so far points at an uneven distribution of protozoa in space but also in time, with quickly vanishing "hot spots" of activity and long periods of inactivity as resting cysts. An adequate choice of sampling strategy is therefore crucial to any study and will be strongly dependent on the aim of the study. There is a high level of heterogeneity in most soils. For instance, when samples were taken only 5–10 mm apart from a single sample of sieved, mixed soil (experimentally as homogeneous as possible), amoebal numbers varied between 3000 and 5000 g⁻¹ and their Shannon diversity index varied from 1.8 to 2.6 (Anderson, 2002).

The most probable number (MPN) method is the only practical, quantitative technique available for microscopic organisms normally occurring in numbers of 10^3–10^4 g⁻¹ soil. These numbers are too low to lend protozoa to convenient direct counting. The MPN method has some disadvantages for protozoan enumeration, which may cause a bias in the counting (Adl, 2003). "Morphotype" approaches for the quantitative study of the diversity of flagellates and amoebae have been described (Ekelund et al., 2001; Smirnov and Brown, 2004). Some morphotypes represent taxonomically well-defined species, while others could include whole genera and so would require more detailed observations (i.e., by electron microscopy or sectioning) to differentiate the species. These practical approaches are clearly complementary to emerging molecular methods mentioned later.

Fatty acids can also serve as markers of protozoa. The signature fatty acid 20ω4 is routinely used to detect protozoa in aquatic environments, but since it can occur also in other eukaryotes, it is less easy to connect fatty acid 20ω4 to protozoa in soil, although there was a good correlation in a microcosm study on fresh manure added to an arable soil (Frostegård et al., 1997).

Regarding protozoan taxonomy, rapid progress in molecular methods is being made, and it is here that significant advances are expected. Despite some initial controversies, data from modern morphological approaches, biochemical pathways, and molecular phylogenetics are generally complementary (Adl et al., 2005). It is clear that the phylogenetics of protozoa are not straightforward. The traditional morphology-based classification grouped naked amoebae together in a macrotaxon named Sarcodina, molecular studies have now placed the majority of sarcodinids into two new supergroups: Amoebozoa and Rhizaria (Pawlowski and Burki, 2009). There is still a need to

generate taxon-rich multigene datasets to resolve their phylogenetic relationships and of environmental sequencing to increase taxon sampling in these supergroups. The diversity of 18S rDNA genotypes within traditional morphospecies can be very high, as mentioned above (Bass et al., 2007). Emerging, high-throughput sequencing technologies (pyrosequencing) of both DNA and RNA will, no doubt, reveal much more about the ecology of soil protozoa. Urich et al. (2008), for example, were able to demonstrate relative and quantitative abundances of RNA sequences from active protozoan groups in a grassland soil. Stable isotope probing (SIP), in which the microorganisms involved in the degradation of isotopically labeled substrates can be identified by separation of their isotopically labeled nucleic acids, can reveal the role of protozoa in specific food webs (Lueders et al., 2004).

So far, progress with the novel methods has been most rapid for individual taxa in aquatic environments (Caron et al., 2004), where protozoa can be concentrated by filtration. For example, real-time PCR has been carried out for the detection and quantification of the protozoan oyster parasite *Perkinsus marinus* in natural waters, and that technology is clearly applicable to also assess single species of interest in the soil environment. However, the low "carrying capacity" of soil for protozoa (typically 10^3–$10^4\,g^{-1}$) requires efficient extraction methods for subsequent analyses.

25.1.6 Ecological Roles for Protozoa in Soil

25.1.6.1 Shaping Microbial Community Structure

As a result of their feeding preferences and high consumption rates, protozoa have the capacity to strongly influence the bacterial species composition in soil. The presence of different amoebal and flagellate species resulted in the appearance of different bacterial communities (Rønn et al., 2002).

25.1.6.2 Protection and Distribution of Bacteria

Free-living naked soil amoebae, like *Acanthamoeba*, *Naegleria*, and *Hartmannella*, can act as reservoirs and modes of dispersal for human-pathogenic bacteria. Molecular methods have identified cells of the human pathogens: *Legionella*, *Chlamydia*, *Listeria*, and *Salmonella* inside the amoebae (Greub and Raoult, 2004; Brandl et al., 2005). At outdoor temperatures, the pathogens apparently live as endosymbionts, but at 30°C–37°C they change their lifestyle, becoming lytic and eradicating their hosts. Probably, the coevolution of bacteria inside amoebae has conditioned them to overcome the metazoan immune system, and made them capable of attacking human tissues. When the amoebae encyst, the pathogenic bacteria they harbor will be protected inside the cyst and be able to survive severe disinfection efforts. Although recognized so far as a long-term survival mechanism for pathogenic bacteria, this protection within a cyst should theoretically be operational with other soil bacteria.

25.1.6.3 Influencing Plant Growth

Previous reviews of the role of protozoa in the rhizosphere have tended to concentrate on their contribution to gross flows of carbon and nitrogen (Griffiths, 1994; Zwart et al., 1994) or their role in disease suppression (Curl and Harper, 1990). In the rhizosphere, bacteria are more important decomposer organisms than fungi, because of the large supply of easily decomposable organic matter (Wardle, 2002), so interactions between bacteria and their microfaunal grazers are of more consequence than those of fungi and their grazing animals. Microfaunal grazing in the rhizosphere is particularly important because plant available nutrients will be strongly sequestered during microbial growth (Kaye and Hart, 1997; Wang and Bakken, 1997) and would remain locked up in bacterial biomass if consumption by nematodes and protozoa would not constantly remobilize essential nutrients for plant uptake (Bonkowski et al., 2000). Recent reviews of the role of protozoa (Clarholm et al., 2007; Griffiths et al., 2007) have emphasized the importance of indirect interactions of microfaunal grazing. Protozoan grazing effects stimulating root growth by promoting specific rhizobacteria seem even more important than direct effects due to nutrient release (Bonkowski and Brandt, 2002) and is a development of the microbial-loop hypothesis as applied to the rhizosphere (Clarholm, 1985).

References

Adl, S. 2003. The ecology of soil decomposition. CABI, Wallingford, U.K.

Adl, S.M., A.G.B. Simpson, M.A. Farmer, R.A. Andersen, O.R. Anderson, J.R. Barta, S.S. Bowser et al. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. 52:399–451.

Anderson, O.R. 2002. Laboratory and field-based studies of abundances, small-scale patchiness, and diversity of gymnamoebae in soils of varying porosity and organic content: Evidence of microbiocoenoses. J. Eukaryot. Microbiol. 49:17–23.

Bass, D., T.A. Thomas, L. Matthai, V. Marsh, and T. Cavalier-Smith. 2007. DNA evidence for global dispersal and probable endemicity of protozoa. BMC Evol. Biol. 7:162.

Bonkowski, M., and F. Brandt. 2002. Do soil protozoa enhance plant growth by hormonal effects? Soil Biol. Biochem. 34:1709–1715.

Bonkowski, M., B.S. Griffiths, and C. Scrimgeour. 2000. Substrate heterogeneity and microfauna in soil organic "hotspots" as determinants of nitrogen capture and growth of rye-grass. Appl. Soil Ecol. 14:37–53.

Brandl, M., B. Rosenthal, and A. Haxo. 2005. Enhanced survival of *Salmonella enterica* in vesicles released by a soilborne *Tetrahymena* species. Appl. Environ. Microbiol. 71:1562–1569.

Caron, D., P. Countway, and M. Brown. 2004. The growing contributions of molecular biology and immunology to protistan ecology: Molecular signatures and ecological tools. J. Eukaryot. Microbiol. 51:38–48.

Clarholm, M. 1985. Interactions of bacteria, protozoa and plants leading to mineralization of soil nitrogen. Soil Biol. Biochem. 17:181–187.

Clarholm, M., M. Bonkowski, and B.S. Griffiths. 2007. Protozoa and other protista in soil, p. 147–175. *In* J.D. Van Elsas, J.K. Jansson, and J.T. Trevors (eds.) Modern soil microbiology. 2nd edn. CRC Press, Boca Raton, FL.

Curl, E.A., and J.D. Harper. 1990. Fauna-microflora interactions, p. 369–388. *In* J.M. Lynch (ed.) The rhizosphere. John Wiley & Sons, Chichester, U.K.

De Ruiter, P.C., B.S. Griffiths, and J.C. Moore. 2002. Biodiversity and stability in soil ecosystems: Patterns, processes and the effects of disturbance, p. 102–113. *In* M. Loreau, S. Naeem, and P. Inchausti (eds.) Biodiversity and ecosystem functioning. Oxford University Press, Oxford, U.K.

Ekelund, F., R. Rønn, and B.S. Griffiths. 2001. Quantitative estimation of flagellate community structure and diversity in soil samples. Protist 152:301–314.

Finlay, B.J. 2002. Global dispersal of free-living microbial eukaryote species. Science 296:1061–1063.

Foissner, W. 2006. Biogeography and dispersal of micro-organisms: A review emphasizing protists. Acta Protozool. 45:111–136.

Frostegård, Å., S. Petersen, E. Bååth, and T. Nielsen. 1997. Dynamics of a microbial community associated with manure hot spots as revealed by phospholipid fatty acid analyses. Appl. Environ. Microbiol. 63:2224–2231.

Greub, G., and D. Raoult. 2004. Microorganisms resistant to free-living amoebae. Clin. Microbiol. Rev. 17:413–433.

Griffiths, B.S. 1994. Soil nutrient flow, p. 65–91. *In* J.F. Darbyshire (ed.) Soil protozoa. CABI, Wallingford, U.K.

Griffiths, B.S., S. Christensen, and M. Bonkowski. 2007. Microfaunal interactions in the rhizosphere, how nematodes and protozoa link above- and below-ground processes, p. 57–71. *In* Z. Cardon and J. Whitbeck (eds.) The rhizosphere: An ecological perspective. Elsevier, Amsterdam, the Netherlands.

Hunt, H.W., and D.H. Wall. 2002. Modelling the effects of loss of soil biodiversity on ecosystem function. Global Change Biol. 8:33–50.

Kaye, J.P., and S.C. Hart. 1997. Competition for nitrogen between plants and soil microorganisms. Trends Ecol. Evol. 12:139–143.

Kiorboe, T., H. Grossart, H. Plough, K. Tang, and B. Auer. 2004. Particle associated flagellates: Swimming patterns, colonization rates and grazing of attached bacteria. Aquat. Microb. Ecol. 35:141–152.

Lueders, T., B. Wagner, P. Claus, and M.W. Friedrich. 2004. Stable isotope probing of rRNA and DNA reveals a dynamic methylotroph community and trophic interactions with fungi and protozoa in oxic rice field soil. Environ. Microbiol. 6:60–72.

Pawlowski, J., and F. Burki. 2009. Untangling the phylogeny of amoeboid protists. J. Eukaryot. Microbiol. 56:16–25.

Rønn, R., A. McCaig, B.S. Griffiths, and J. Prosser. 2002. Impact of protozoan grazing on bacterial community structure in soil microcosms. Appl. Environ. Microbiol. 68:6094–6105.

Schröter, D., V. Wolters, and P. De Ruiter. 2003. C and N mineralisation in the decomposer food webs of a European forest transect. Oikos 102:294–308.

Smirnov, A., and S. Brown. 2004. Guide to the study and identification of soil amoebae. Protistology 3:148–190.

Urich, T., A. Lanzén, J. Qi, D.H. Huson, C. Schleper, and C.H. Schuster. 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLoS One 3:e2527.

Wang, J.G., and L.R. Bakken. 1997. Competition for nitrogen during decomposition of plant residues in soil: Effect of spatial placement of N-rich and N-poor plant residues. Soil Biol. Biochem. 29:153–162.

Wardle, D.A. 2002. Communities and ecosystems, linking aboveground and belowground components. Monographs in population ecology 34. Princeton University Press, Princeton, NJ.

Zwart, K.B., P.J. Kuikman, and J.A. van Veen. 1994. Rhizosphere protozoa: Their significance in nutrient dynamics, p. 93–122. *In* J.F. Darbyshire (ed.) Soil protozoa. CABI, Wallingford, U.K.

25.2 Nematodes

Robert McSorley

25.2.1 Introduction

Nematodes, or roundworms, which are common in most habitats, are a diverse group of invertebrates, including forms that inhabit marine, freshwater, or terrestrial habitats, as well as parasites of man and other vertebrates and of insects and other invertebrates. Many different kinds of nematodes inhabit soil, but while often present in large numbers, they often go unnoticed because of their cryptic habits and microscopic size, usually 0.5–2.0 mm in length. Some of the soil-inhabiting forms are plant parasites that are economically important in agriculture (Barker et al., 1998). Over 15,000 species of nematodes have been described (Poinar, 1983), but many more species remain unknown, and it is common to encounter undescribed species in most soil habitats.

25.2.2 Basic Taxonomy and Identification

Nematodes belong to the phylum Nematoda (Poinar, 1983), also referred to by some authors as the phylum Nemata (Thorne, 1961; Maggenti, 1991). Formerly, the Nematoda was considered a class within the phylum Aschelminthes, along with rotifers and related organisms (Barnes, 1965). The phylum Nematoda is divided into two classes: Adenophorea and Secernentea (Thorne, 1961; Poinar, 1983; Freckman and Baldwin, 1990). Poinar (1983) divides the two classes into 16 orders (Table 25.1). Some authors may recognize a slightly different arrangement, separating Triplonchida from Dorylaimida or Diplogasterida from Rhabditida (Andrassy, 1976; Maggenti, 1991; Lorenzen, 1994).

Some orders of nematodes are not common in soil, but certain stages of animal parasites may sometimes be found (Levine, 1968; Anderson, 1992) and marine nematodes may occur in soil

TABLE 25.1 Habitats and Taxonomic References for Nematode Orders, Based on Classification Scheme of Poinar (1983)

Order	Habitats[a]	References to Soil-Inhabiting Forms
Class Adenophorea		
Araeolaimida	Marine, freshwater, soil	Freckman and Baldwin (1990), Andrassy (1976), Ferris et al. (1973), Goodey and Goodey (1963)
Chromadorida	Marine, freshwater, soil	Freckman and Baldwin (1990), Ferris et al. (1973), Goodey and Goodey (1963)
Desmodorida	Marine, freshwater, soil	Freckman and Baldwin (1990), Ferris et al. (1973)
Desmoscolecida	Marine	
Dorylaimida	Soil, freshwater	Ferris (1971), Thorne (1974)
Enoplida	Marine, freshwater, soil	Freckman and Baldwin (1990), Andrassy (1976), Ferris et al. (1973), Goodey and Goodey (1963)
Mermithida	Invertebrates	
Monhysterida	Marine, freshwater, soil	Andrassy (1983)
Mononchida	Soil, freshwater	Jairajpuri and Khan (1982), Jensen and Mulvey (1968)
Class Secernentea		
Aphelenchida	Soil[b], invertebrates	Hunt (1993)
Ascaridida	Vertebrates, invertebrates	
Oxyurida	Invertebrates, vertebrates	
Rhabditida	Soil, freshwater, invertebrates, vertebrates	Andrassy (1983)
Spirurida	Vertebrates, invertebrates	
Strongylida	Vertebrates	
Tylenchida	Soil[b], invertebrates	Mai and Mullin (1996), Siddiqi (2000)

[a] Habitats listed in order of importance.
[b] Some soil-inhabiting forms also live in plant tissues.

from estuaries or other coastal habitats. Many nematodes are associates of insects (referred to as entomogenous nematodes), and are occasionally found in soil (Nickle, 1984; Gaugler and Kaya, 1990). These include insect-parasitic, or entomopathogenic, nematodes. The genera *Steinernema* and *Heterorhabditis* in the order Rhabditida have been used frequently in the biological control of soil insects (Gaugler and Kaya, 1990; Nguyen and Hunt, 2008). Plant-parasitic nematodes are common constituents of the soil fauna. Nonparasites of all types are referred to collectively as "free-living" nematodes.

Specific references are useful for the identification of some of the orders commonly occurring in soil (Table 25.1). More general texts (Thorne, 1961; Goodey and Goodey, 1963; Ferris, 1971; Andrassy, 1976; Freckman and Baldwin, 1990) include these as well as genera from orders that occur less frequently in soil.

In most texts, identification is to genus level; species identification is difficult (Barsi and Luca, 2008) and many species are undescribed.

25.2.3 Biology and Ecology

25.2.3.1 Morphology

Nematodes are unsegmented roundworms and most are worm-like in appearance, although some modifications exist. They have a complete digestive system, nervous, excretory, and reproductive systems, and longitudinal muscles, but lack respiratory and circulatory systems. Sexes are separate, although intersexes occur in some species (Wharton, 1986; Bird and Bird, 1991). Usually reproduction is sexual and amphimictic, although some species reproduce parthenogenetically or hermaphroditically (Poinar, 1983; Wharton, 1986; Bird and Bird, 1991). With substage lighting under a microscope, nematodes appear transparent, so that internal anatomical and morphological features important in taxonomy can be examined. The characteristics of the stoma (mouth region) and esophagus are particularly important in separating some of the nematode orders present in soil. In some groups, the stoma is modified into a needle-like stylet to aid in certain types of feeding.

25.2.3.2 Life Cycle

A typical nematode life cycle consists of the egg, four juvenile stages, and the adult (Wharton, 1986; Bird and Bird, 1991). The juvenile stages differ in size but are similar morphologically. Modifications and exceptions occur in some groups, such as the plant-parasitic genus *Meloidogyne*, in which later juvenile stages become swollen, developing into a saccate female (Poinar, 1983). The lengths of nematode life cycles vary from a few days to more than a year, with those of most common plant parasites averaging 3–6 weeks and those of some free-living forms only 1–2 weeks (Norton, 1978; Bird and Bird, 1991). In all cases, the duration of the life cycle is strongly dependent on temperature (Norton, 1978).

Various types of delayed development, such as the formation of dauer larvae, may occur in some species in response to adverse environmental conditions (Wharton, 1986; Bird and Bird, 1991). In extreme cases, nematodes of some species may enter anabiosis, an extended state of metabolic dormancy that may last for years (Norton, 1978; Wharton, 1986; McSorley, 2003). Anhydrobiosis is the anabiotic response to low soil moisture and desiccation (Wharton, 1986).

25.2.3.3 Nematodes in Belowground Food Webs

Nematodes are an integral part of food webs in soil ecosystems, appearing at several positions and levels due to their diverse feeding habits, which form the basis for subdivisions into a number of trophic groups (Elliott et al., 1988; Moore and de Ruiter, 1991; Yeates et al., 1993). Five of these groups are frequent in soil (Table 25.2): bacterivores feed on soil bacteria; fungivores feed on fungal mycelia; herbivores, or phytophages,

TABLE 25.2 Feeding Habits of Nematodes Found in Soil Ecosystems

Order	Bacterivores	Fungivores	Herbivores	Omnivores	Predators
Class Adenophorea					
Araeolaimida	All soil forms				
Chromadorida[a]	Few				
Desmodorida	Some				
Dorylaimida		Some	Some	Some	Some
Enoplida	Some			Few	Some
Monhysterida	Most soil forms				
Mononchida	Few				Most
Class Secernentea					
Aphelenchida[b]		Many	Some		Few
Rhabditida[b]	Most				Some
Tylenchida[b]		Some	Most		

   [a] Mostly algivores.

   [b] Some members in these orders are insect parasites or insect associates.

are the plant parasites; predators feed on nematodes and other small invertebrates; omnivores feed on a variety of foods such as fungi, algae, small invertebrates, bacteria, etc. (Freckman, 1982; Freckman and Baldwin, 1990). Feeding habits vary within the nematode orders occurring in soil. Because the biology of many nematodes is relatively unstudied, the food habits of some groups are unknown (Yeates et al., 1993). Some of these groups, such as the family Tylenchidae in the order Tylenchida, which are common in soil, are sometimes grouped as "plant associates" since feeding habits remain unclear. Recent work has clarified that several species of *Filenchus*, an abundant genus in this family, feed as fungivores (Okada et al., 2005).

25.2.3.4 Abundance, Distribution, and Diversity

A diverse nematode fauna of typically two or three dozen different genera are present in most soils, with numbers in the range of 10^6–10^7 m^{-2}, as illustrated by results from selected earlier studies (Table 25.3). Subsequent studies typically report 20–40 genera per terrestrial site, with a trend toward greater diversity and numbers of taxa under organic versus conventional management (Yeates and Boag, 2003). Grasslands may show unusually high nematode abundance, with numbers in excess of 10^7 m^{-2} reported in several cases (de Goede and Bongers, 1998). When dry biomass is reported, values in the range of 100–200 mg m^{-2} are common (Sohlenius, 1980) but can be greater in communities containing a high percentage of large dorylaimid omnivores (Wasilewska, 1971; Ferris and Ferris, 1974). The percent composition of the nematode fauna by trophic group varies, depending on the plant species present and the geographic location. Relatively wide ranges in trophic composition or abundance within a site may result from seasonal variation. Enrichment with organic matter stimulates the entire soil food web, increasing both nematode numbers and the variety of taxa present (Ferris et al., 2004; Wang et al., 2004; Ferris and Bongers, 2006). Conversely, nematode abundance may be decreased in polluted sites (Ekschmitt and Korthals, 2006) or in sites disturbed due to

land management (Todd et al., 2006) or some types of agricultural practices (Berkelmans et al., 2003; Yeates and Boag, 2003; Wang et al., 2006).

25.2.4 Importance to Soil Processes

25.2.4.1 Decomposition and Nutrient Cycling

Sohlenius (1980) reviewed the allocation of energy for production or respiration in nematode communities at various sites. Roughly 2/3 of the energy assimilated by nematodes was liberated as respiration energy, with the remainder used for production (Sohlenius, 1980). This energy stored within nematode tissues is then available for cycling to organisms that feed on nematodes. The role of nematodes in energy cycling and flow is significant because their consumption and population turnover rates are high. Estimates of nematode consumption of root biomass from several studies in Colorado and South Dakota prairies ranged from 34.8 to 57 g m^{-2} (more than vertebrate herbivores) and in one study amounted to 5.8%–12.6% of annual net primary production (Ingham and Detling, 1984).

Nematodes occupy several key positions in soil food webs (Moore and de Ruiter, 1991). Since herbivores and microbivores (bacterivores plus fungivores) make up the most abundant nematode groups in most systems, energy flow through the soil nematode community takes two major pathways. Nematode herbivores function only as primary consumers, removing energy directly from plant roots, whereas energy flow from plant to bacterivores or fungivores is indirect. Nematodes do not feed directly on organic matter of plant origin but on the bacteria and fungi that break down this material. Other less abundant nematode groups function as predators that feed on nematodes and other small invertebrates, or as omnivores that feed both as predators and as bacterivores or fungivores. In one study, 34%–44% of the energy in organic matter from crop residues that was passed through bacteria and fungi was eventually consumed by nematodes within a year (Wasilewska and Bienkowski, 1985). Changes in the relative

TABLE 25.3 Examples of Nematode Richness and Abundance from Selected Studies

Ecosystem	Location	Genera per Site	Nematodes[a] m^{-2} ($\times 10^6$)	Reference
Tundra	15 sites		3.5	Sohlenius (1980)
Grassland	20 sites		9.2	Sohlenius (1980)
Grassland	California	26	2.7–3.2	Freckman et al. (1979)
Grassland	New Zealand	29–32	0.8–4.5	Yeates (1981)
Grassland	New Zealand	21–35	2.3	Yeates (1984)
Grassland	South Australia	17–19	0.8–3.4	Yeates and Bird (1994)
Grassland	Poland		1.9–12.0	Wasilewska (1994)
Grassland	Sweden	38–41	7.6	Bostrom and Sohlenius (1986)
Barley field	Sweden	38–41	6.2	Bostrom and Sohlenius (1986)
Soybean field	Florida	26–30	1.7–5.9	McSorley and Frederick (1996)
Cultivated fields	South Australia	10–22	0.6–10.6	Yeates and Bird (1994)
Shrub	South Australia	21–26	0.6–2.8	Yeates and Bird (1994)
Deciduous forest	15 sites		6.3	Sohlenius (1980)
Deciduous forest	Indiana		1.2–2.5	Ferris and Ferris (1974)
Deciduous forest	Tennessee		1.1–6.9	Ferris and Ferris (1974)
Forests, various	Poland		0.5–7.0	Wasilewska (1971)
Pine forest	14 sites		3.3	Sohlenius (1980)
Pine forest	Sweden	23	4.1	Sohlenius (1980)
Pine forest	Florida	34–37	0.4–0.5	McSorley (1993)
Tropical forest	2 sites		1.7	Sohlenius (1980)
Tropical forest	New Caledonia		0.1–0.3	Yeates (1991)
Moss	Signy Island (Antarctica)	9	0.4–0.7	Caldwell (1981)
Bog	9 sites		1.7	Sohlenius (1980)
Desert	2 sites		0.8	Sohlenius (1980)

[a] Mean or range reported, depending on study.

proportions of nematode bacterivores and fungivores over time reflect changes in the pathway of decomposition (Wasilewska and Bienkowski, 1985). Nematodes accelerate the decomposition process by dispersing relatively immobile microflora to new sites and by their feeding, which regulates bacterial growth and decomposition rates (Ingham et al., 1985; Wasilewska and Bienkowski, 1985; Freckman, 1988). Overgrazing of bacteria is avoided as nematodes in turn are regulated by mites and other invertebrate predators (Whitford et al., 1982; Elliott et al., 1988).

The increased decomposition rates resulting from nematode feeding increase the recycling and mineralization of C and other elements (Ingham et al., 1985; Freckman, 1988). Nematodes are particularly important in recycling N, which can become immobilized in bacterial populations during decomposition (Freckman,

1988). Since nematodes have a higher C:N ratio (8:1 to 12:1) than their bacterial food source (3:1 to 4:1) (Wasilewska and Bienkowski, 1985), their feeding results in the excretion of N, mostly as NH_3 (Freckman, 1988). Numerous studies confirm the increase of NH_4^+ and other inorganic N sources in soil with nematodes present, and increased N levels in plant tissue have resulted in some instances (Ingham et al., 1985; Freckman, 1988). Nematodes are also important in the enhanced mineralization of P (Ingham et al., 1985; Freckman, 1988), and even S in some instances (Freckman, 1988).

25.2.4.2 Economic Importance to Agriculture

Nematode herbivores are of particular importance in agriculture because of their potential to damage the roots of crop plants (Norton, 1978). Ectoparasites feed at the root surface but do not enter roots, while endoparasites may feed and live within root tissue. Feeding by large numbers of plant-parasitic nematodes can result in yield losses, wilting, nutrient deficiency symptoms, and other problems associated with a damaged root system. Nematode problems of various crops are already reviewed (Nickle, 1984; Luc et al., 1990; Evans et al., 1993; Barker et al., 1998; Chen et al., 2003); some of the most important nematode pathogens of crops are shown in Table 25.4.

TABLE 25.4 Major Nematode Pathogens of Selected Agricultural Crops

Crop	Nematodes
Alfalfa (*Medicago sativa*)	*Ditylenchus dipsaci, Meloidogyne* spp.
Banana (*Musa* spp.)	*Radopholus similis, Helicotylenchus multicinctus*
Citrus (*Citrus* spp.)	*Tylenchulus semipenetrans, R. similes*
Coconut (*Cocos nucifera*)	*Bursaphelenchus (Rhadinaphelenchus) cocophilus*
Coffee (*Coffea arabica*)	*Meloidogyne* spp., *Pratylenchus* spp.
Cotton (*Gossypium hirsutum*)	*Meloidogyne incognita, Rotylenchulus reniformis*
Grape (*Vitis vinifera*)	*Meloidogyne* spp., *Xiphinema* spp.
Maize (*Zea mays*)	*Pratylenchus* spp., *Belonolaimus longicaudatus*
Peanut (*Arachis hypogaea*)	*Meloidogyne* spp., *Pratylenchus brachyurus*
Pine (*Pinus* spp.)	*Bursaphelenchus xylophilus*
Potato (*Solanum tuberosum*)	*Globodera* spp., *Meloidogyne* spp., *Pratylenchus* spp.
Rice (*Oryza sativa*)	*Hirschmanniella* spp., *Aphelenchoides besseyi*
Soybean (*Glycine max*)	*Heterodera glycines, Meloidogyne* spp.
Sugar beet (*Beta vulgaris*)	*Heterodera schachtii, Meloidogyne* spp.
Sugarcane (*Saccharum* spp.)	*Pratylenchus* spp.
Sweet potato (*Ipomoea batatas*)	*Meloidogyne* spp., *R. reniformis*
Vegetables	
Family: Cruciferae (cabbages, etc.)	*H. schachtii, Meloidogyne* spp.
Family: Leguminosae (beans, peas)[a]	*Meloidogyne* spp., *R. reniformis, Heterodera* spp.
Family: Solanaceae (tomatoes, etc.)[a]	*Meloidogyne* spp.
Wheat (*Triticum aestivum*)	*Heterodera avenae*

[a] See also separate listings for soybean and potato.

25.2.5 Methods

25.2.5.1 Collection, Sampling, and Extraction of Nematodes

Methods for collection of samples and extraction are listed in Table 25.5. The irregular horizontal spatial distribution of nematodes presents particular problems in obtaining precise samples from the field (Barker and Nusbaum, 1971; Barker and Campbell, 1981; Goodell and Ferris, 1981; Goodell, 1982; McSorley, 1987). Therefore, soil cores are typically collected from a number of points within the field and then combined to make a composite sample. The sampling plan is the number of cores needed to form the sample, for example, 20 cores per 1.6 ha in North Carolina (Barker and Nusbaum, 1971; Barker and Campbell, 1981). Some methods for extraction of nematodes from soil depend on nematode activity (active techniques) while others can be used to extract inactive or dead nematodes (passive techniques) (Table 25.5). Since many different methods are used for extracting

nematodes from samples, comparison of counts between laboratories can be difficult and should be standardized (McSorley, 1987). Multiple methods may be needed for efficient extraction of nematode communities (McSorley and Frederick, 2004).

25.2.5.2 Community Structure Analysis and Bioindicators

Data on the population density of a key genus or species are often collected in nematology. Studies have been conducted to assess community structure rather than individual taxa (Table 25.5). Nematode succession and increasing numbers of taxa result as plant communities become more complex or as soil food webs are enriched (Niles and Freckman, 1998; Wang et al., 2004). In the development of the maturity index, Bongers (1990) classified nematodes into five successional classes (cp scale), ranging from families that showed characteristics of good colonizers (cp-1) to those best adapted as persisters (cp-5). These successional classes

TABLE 25.5 Outline of Methods for Sampling and Extraction of Nematodes from Soil and Roots

Procedures	References
Collection of Samples	
Field collection	Barker and Nusbaum (1971), Barker and Campbell (1981), Goodell (1982), McSorley (1987), Freckman and Baldwin (1990)
Sampling tools (cores of soil)	Barker and Campbell (1981), Goodell (1982)
Depth of sampling	Barker and Nusbaum (1971), Norton (1978), Barker and Campbell (1981), McSorley (1987)
Timing of sampling (seasonal)	Barker and Nusbaum (1971), Barker and Campbell (1981), Goodell (1982)
Spatial distribution (clumped)	Barker and Nusbaum (1971), Barker and Campbell (1981), Goodell and Ferris (1981), Goodell (1982), McSorley (1987)
Sampling plans	Barker and Nusbaum (1971), Barker and Campbell (1981), Goodell and Ferris (1981), McSorley (1987), Freckman and Baldwin (1990)
Sampling pattern	Barker and Campbell (1981), Goodell and Ferris (1981), McSorley (1987)
Transport and storage of samples	Barker and Nusbaum (1971), Barker and Campbell (1981), Goodell and Ferris (1981), Goodell (1982), McSorley (1987)
Laboratory Methods	
Extraction from soil	Ayoub (1980), Goodell (1982), Southey (1986), Hooper (1990), Freckman and Baldwin (1990)
Subsample preparation	McSorley (1987), Freckman and Baldwin (1990)
Extraction of vermiform stages	Barker and Nusbaum (1971), Ayoub (1980), Goodell (1982), Southey (1986), McSorley (1987), Hooper (1990), Freckman and Baldwin (1990), McSorley and Walter (1991)
Active techniques	McSorley (1987), Freckman and Baldwin (1990), McSorley and Walter (1991)
Diagram of Baermann method	Ayoub (1980), Southey (1986), Hooper (1990)
Passive techniques	McSorley (1987), McSorley and Walter (1991)
Diagram of sieving	Ayoub (1980), Southey (1986)
Diagram of centrifugation	Ayoub (1980)
Extraction efficiency	McSorley (1987)
Extraction of cysts, anabiotic forms	Freckman et al. (1975), Ayoub (1980), Southey (1986)
Extraction from roots	Ayoub (1980), Southey (1986), Hooper (1990)
Evaluating biological control	Stirling (1991)
Calculations	
Counting extracted nematodes	Southey (1986), Hooper (1990)
Energy and biomass	Klekowski et al. (1972), Yeates (1979), Freckman (1982), Freckman and Baldwin (1990)
Community structure indices	Bostrom and Sohlenius (1986), Bongers (1990), Freckman and Ettema (1993), McSorley (1993), Wasilewska (1994), Neher and Campbell (1994), Yeates (1994), McSorley and Frederick (1996), Bongers and Bongers (1998), Neher (1999), Yeates and Bongers (1999), Ferris et al. (2001)

(cp scale) were coupled with nematode feeding groups (bacterivores, fungivores, etc.) to create a matrix that defined nematode guilds (Bongers and Bongers, 1998; Ferris et al., 2001). Analyses based on nematode guilds have been useful in understanding and interpreting the condition of soil food webs (Ferris et al., 2001; Ferris and Matute, 2003; Ferris and Bongers, 2006).

Environmental changes impact both nematode numbers and community structure (Freckman and Ettema, 1993). The ability of nematodes to respond to changes, as well as their diversity, widespread distribution, size, abundance, and other features make nematodes useful as bioindicators of ecological disturbances (Freckman, 1988; Bongers, 1990; Freckman and Ettema, 1993; Neher and Campbell, 1994). Much work has focused on the utility of various indices of community structure for this purpose (Freckman and Ettema, 1993; Neher and Campbell, 1994; Wasilewska, 1994; Yeates and Bird, 1994; McSorley and Frederick, 1996; Ferris and Matute, 2003; Ferris and Bongers, 2006). Nematodes are particularly useful as bioindicators of organic enrichment and condition of soil food webs (Ferris et al., 2001; Ferris and Bongers, 2006). They have potential as bioindicators of pollution from heavy metals (Ekschmitt and Korthals, 2006), and possibly as bioindicators of historical changes and alterations in land use (Todd et al., 2006).

As global climate changes, geographic distributions of plant-parasitic nematodes could slowly expand into cooler areas when temperatures increase. Yeates and Boag (2003) speculate that plant-parasitic nematodes and their hosts would likely migrate together. Increasing soil temperature affected the nematode community, resulting in shifts to higher percentages of fungivorous and plant-parasitic genera (Ruess et al., 1999). Because the nematode life cycle is temperature-dependent, life cycles would become more rapid, possibly resulting in an extra generation per year in some cases. If agriculture expands into regions that now have short growing seasons, it is likely that the plant-parasitic nematode problems typical of those crops will follow.

References

Anderson, R.C. 1992. Nematode parasites of vertebrates. CABI, Wallingford, U.K.

Andrassy, I. 1976. Evolution as a basis for the systematization of nematodes. Pitman Publishing, San Francisco, CA.

Andrassy, I. 1983. A taxonomic review of the suborder Rhabditina (Nematoda: Secernentia). ORSTOM, Paris, France.

Ayoub, S.M. 1980. Plant nematology, an agricultural training aid. Nema Aid Publications, Sacramento, CA.

Barker, K.R., and C.L. Campbell. 1981. Sampling nematode populations, p. 451–474. *In* B.M. Zuckerman and R.A. Rohde (eds.) Plant parasitic nematodes. Vol. 3. Academic Press, New York.

Barker, K.R., and C.J. Nusbaum. 1971. Diagnostic and advisory programs, p. 303–323. *In* B.M. Zuckerman, W.F. Mai, and R.A. Rohde (eds.) Plant parasitic nematodes. Vol. I. Academic Press, New York.

Barker, K.R., G.A. Pederson, and G.L. Windham. 1998. Plant and nematode interactions. ASA-CSSA-SSSA, Madison, WI.

Barnes, R.D. 1965. Invertebrate zoology. W.B. Saunders Company, Philadelphia, PA.

Barsi, L., and F.D. Luca. 2008. Morphological and molecular characterisation of two putative *Xiphinema americanum*-group species, *X. parasimile* and *X. simile* (Nematoda: Dorylaimida) from Serbia. Nematology 10:15–25.

Berkelmans, R., H. Ferris, M. Tenuta, and A.H.C. van Bruggen. 2003. Effects of long-term crop management on nematode trophic levels other than plant feeders disappear after 1 year of disruptive soil management. Appl. Soil Ecol. 23:223–235.

Bird, A.F., and J. Bird. 1991. The structure of nematodes. 2nd edn. Academic Press, San Diego, CA.

Bongers, T. 1990. The maturity index: An ecological measure of environmental disturbance based on nematode species composition. Oecologia 83:14–19.

Bongers, T., and M. Bongers. 1998. Functional diversity of nematodes. Appl. Soil Ecol. 10:239–251.

Bostrom, S., and B. Sohlenius. 1986. Short-term dynamics of nematode communities in arable soil. Influence of a perennial and an annual cropping system. Pedobiologia 29:345–357.

Caldwell, J.R. 1981. Biomass and respiration of nematode populations in two mass communities at Signy Island, maritime Antarctic. Oikos 37:160–166.

Chen, Z.X., S.Y. Chen, and D.W. Dickson. 2003. Nematology advances and perspectives. CABI Publishing, Wallingford, U.K.

de Goede, R.G.M., and T. Bongers. 1998. Nematode communities of northern temperate grassland ecosystems. Focus Verlag, Giessen, Germany.

Ekschmitt, K., and G.W. Korthals. 2006. Nematodes as sentinels of heavy metals and organic toxicants in the soil. J. Nematol. 38:13–19.

Elliott, E.T., H.W. Hunt, and D.E. Walter. 1988. Detrital food-web interactions in North American grassland ecosystems. Agric. Ecosyst. Environ. 24:41–56.

Evans, K., D.L. Trudgill, and J.M. Webster. 1993. Plant parasitic nematodes in temperate agriculture. CABI, Wallingford, U.K.

Ferris, V.R. 1971. Taxonomy of the Dorylaimida, p. 164–189. *In* B.M. Zuckerman, W.F. Mai, and R.A. Rohde (eds.) Plant parasitic nematodes. Vol. 1. Academic Press, New York.

Ferris, H., and T. Bongers. 2006. Nematode indicators of organic enrichment. J. Nematol. 38:3–12.

Ferris, H., T. Bongers, and R.G.M. de Goede. 2001. A framework for soil food web diagnostics: Extension of the nematode faunal analysis concept. Appl. Soil Ecol. 18:13–29.

Ferris, V.R., and J.M. Ferris. 1974. Inter-relationships between nematode and plant communities in agricultural systems. Agro-Ecosystems 1:275–299.

Ferris, V.R., J.M. Ferris, and J.P. Tjepkema. 1973. Genera of freshwater nematodes (Nematoda) of eastern North America. USEPA, Washington, DC.

Ferris, H., and M.M. Matute. 2003. Structural and functional succession in the nematode fauna of a soil food web. Appl. Soil Ecol. 23:93–110.

Ferris, H., R.C. Venette, and K.M. Scow. 2004. Soil management to enhance bacterivore and fungivore nematode populations and their nitrogen mineralization function. Appl. Soil Ecol. 25:19–35.

Freckman, D.W. 1982. Parameters of the nematode contribution to ecosystems, p. 81–97. *In* D.W. Freckman (ed.) Nematodes in soil ecosystems. University of Texas Press, Austin, TX.

Freckman, D.W. 1988. Bacterivorous nematodes and organic matter decomposition. Agric. Ecosyst. Environ. 24:195–217.

Freckman, D.W., and J.G. Baldwin. 1990. Nematoda, p. 155–200. *In* D.L. Dindal (ed.) Soil biology guide. John Wiley & Sons, New York.

Freckman, D.W., D.A. Duncan, and J.R. Larson. 1979. Nematode density and biomass in an annual grassland system. J. Range Manage. 32:418–421.

Freckman, D.W., and C.H. Ettema. 1993. Assessing nematode communities in agroecosystems of varying human intervention. Agric. Ecosyst. Environ. 45:239–261.

Freckman, D.W., R. Mankau, and H. Ferris. 1975. Nematode community structure in desert soils: Nematode recovery. J. Nematol. 7:343–346.

Gaugler, R., and H.K. Kaya. 1990. Entomopathogenic nematodes in biological control. CRC Press, Boca Raton, FL.

Goodell, P.B. 1982. Soil sampling and processing for detection and quantification of nematode populations for ecological studies, p. 178–198. *In* D.W. Freckman (ed.) Nematodes in soil ecosystems. University of Texas Press, Austin, TX.

Goodell, P.B., and H. Ferris. 1981. Sample optimization for five plant-parasitic nematodes in an alfalfa field. J. Nematol. 13:304–313.

Goodey, T., and J.B. Goodey. 1963. Soil and freshwater nematodes. John Wiley & Sons, New York.

Hooper, D.J. 1990. Extraction and processing of plant and soil nematodes, p. 45–68. *In* M. Luc, R.A. Sikora, and J. Bridge (eds.) Plant parasitic nematodes in subtropical and tropical agriculture. CABI, Wallingford, U.K.

Hunt, D.J. 1993. Aphelenchida, Longidoridae, and Trichodoridae. CABI, Wallingford, U.K.

Ingham, R.E., and J.K. Detling. 1984. Plant-herbivore interactions in a North American mixed-grass prairie. III. Soil nematode populations and root biomass on *Cynomys ludovicianus* colonies and adjacent uncolonized areas. Oecologia 63:307–313.

Ingham, R.E., J.A. Trofymow, E.R. Ingham, and D.C. Coleman. 1985. Interactions of bacteria, fungi, and their nematode grazers: Effects on nutrient cycling and plant growth. Ecol. Monogr. 55:119–140.

Jairajpuri, M.S., and W.U. Khan. 1982. Predatory nematodes (Mononchida). Associated Publishing Co., New Delhi, India.

Jensen, H.J., and R.H. Mulvey. 1968. Predacious nematodes (Mononchidae) of Oregon. Oregon State University Press, Corvallis, OR.

Klekowski, R.Z., L. Wasilewska, and E. Paplinska. 1972. Oxygen consumption by soil inhabiting nematodes. Nematologica 18:391–403.

Levine, N.D. 1968. Nematode parasites of domestic animals and of man. Burgess Publishing Co., Minneapolis, MN.

Lorenzen, S. 1994. The phylogenetic systematics of free-living nematodes. Intercept Ltd., Andover, MA.

Luc, M., R.A. Sikora, and J. Bridge. 1990. Plant parasitic nematodes in subtropical and tropical agriculture. CAB, Wallingford, U.K.

Maggenti, A.R. 1991. Nemata: Higher classification, p. 147–165. *In* W.R. Nickle (ed.) Manual of agricultural nematology. Marcel Dekker, New York.

Mai, W.F., and P.G. Mullin. 1996. Plant parasitic nematodes—A pictorial key to genera. 5th edn. Cornell University, Ithaca, NY.

McSorley, R. 1987. Extraction of nematodes and sampling methods, p. 13–47. *In* R.H. Brown and B.R. Kerry (eds.) Principles and practice of nematode control in crops. Academic Press, Sydney, Australia.

McSorley, R. 1993. Short-term effects of fire on the nematode community in a pine forest. Pedobiologia 37:39–48.

McSorley, R. 2003. Adaptations of nematodes to environmental extremes. Fla. Entomol. 86:138–142.

McSorley, R., and J.J. Frederick. 1996. Nematode community structure in rows and between rows of a soybean field. Fundam. Appl. Nematol. 19:251–260.

McSorley, R., and J.J. Frederick. 2004. Effect of extraction method on perceived composition of the soil nematode community. Appl. Soil Ecol. 27:55–63.

McSorley, R., and D.E. Walter. 1991. Comparison of soil extraction methods for nematodes and microarthropods. Agric. Ecosyst. Environ. 34:201–208.

Moore, J.C., and P.C. de Ruiter. 1991. Temporal and spatial heterogeneity of trophic interactions within below-ground food webs. Agric. Ecosyst. Environ. 34:371–397.

Neher, D.A. 1999. Nematode communities in organically and conventionally managed agricultural soils. J. Nematol. 31:142–154.

Neher, D.A., and C.L. Campbell. 1994. Nematode communities and microbial biomass in soils with annual and perennial crops. Appl. Soil Ecol. 1:17–28.

Nguyen, K.B., and D.J. Hunt. 2008. Entomopathogenic nematodes: Systematics, phylogeny and bacterial symbionts. Brill Academic Publishers, Leiden, the Netherlands.

Nickle, W.R. 1984. Plant and insect nematodes. Marcel Dekker, New York.

Niles, R.K., and D.W. Freckman. 1998. From the ground up: Nematode ecology in bioassessment and ecosystem health, p. 65–85. *In* K.R. Barker, G.A. Pederson, and G.L. Windham (eds.) Plant and nematode interactions. ASA, CSSA, SSSA, Madison, WI.

Norton, D.C. 1978. Ecology of plant-parasitic nematodes. John Wiley & Sons, New York.

Okada, H., H. Harada, and I. Kadota. 2005. Fungal-feeding habits of six nematode isolates in the genus *Filenchus*. Soil Biol. Biochem. 37:1113–1120.

Poinar, G.O., Jr. 1983. The natural history of nematodes. Prentice-Hall, Englewood Cliffs, NJ.

Ruess, L., A. Michelsen, and S. Jonasson. 1999. Simulated climate change in subarctic soils: Responses in nematode species composition and dominance structure. Nematology 1:513–526.

Siddiqi, M.R. 2000. Tylenchida parasites of plants and insects. 2nd edn. CABI, Wallingford, U.K.

Sohlenius, B. 1980. Abundance, biomass and contribution to energy flow by soil nematodes in terrestrial ecosystems. Oikos 34:186–194.

Southey, J.F. 1986. Laboratory methods for work with plant and soil nematodes. Her Majesty's Stationery Office, London, U.K.

Stirling, G. 1991. Biological control of plant parasitic nematodes. CABI, Wallingford, U.K.

Thorne, G. 1961. Principles of nematology. McGraw-Hill Book Co., New York.

Thorne, G. 1974. Nematodes of the northern Great Plains. Part II. Dorylaimoidea in part (Nemata: Adenophorea). South Dakota Agricultural Experiment Station Technical Bulletin 31, Brookings, SD.

Todd, T.C., T.O. Powers, and P.G. Mullin. 2006. Sentinel nematodes of land-use change and restoration in tallgrass prairie. J. Nematol. 38:20–27.

Wang, K.-H., R. McSorley, and N. Kokalis-Burelle. 2006. Effects of cover cropping, solarization, and soil fumigation on nematode communities. Plant Soil 286:229–243.

Wang, K.-H., R. McSorley, A.J. Marshall, and R.N. Gallaher. 2004. Nematode community changes associated with decomposition of *Crotalaria juncea* amendment in litterbags. Appl. Soil Ecol. 27:31–45.

Wasilewska, L. 1971. Nematodes of the dunes in the Kampinos forest. II. Community structure based on numbers of individuals, state of biomass and respiratory metabolism. Ekol. Pol. 19:651–688.

Wasilewska, L. 1994. The effect of age of meadows on succession and diversity in soil nematode communities. Pedobiologia 38:1–11.

Wasilewska, L., and P. Bienkowski. 1985. Experimental study on the occurrence and activity of soil nematodes in decomposition of plant material. Pedobiologia 28:41–57.

Wharton, D.A. 1986. A functional biology of nematodes. The Johns Hopkins University Press, Baltimore, MD.

Whitford, W.G., D.W. Freckman, P.F. Santos, N.Z. Elkins, and L.W. Parker. 1982. The role of nematodes in decomposition in desert ecosystems, p. 98–116. *In* D.W. Freckman (ed.) Nematodes in soil ecosystems. University of Texas Press, Austin, TX.

Yeates, G.W. 1979. Soil nematodes in terrestrial ecosystems. J. Nematol. 11:213–229.

Yeates, G.W. 1981. Populations of nematode genera in soils under pasture. IV. Seasonal dynamics at five North Island sites. N.Z. J. Agric. Res. 24:107–121.

Yeates, G.W. 1984. Variation in soil nematode diversity under pasture with soil and year. Soil Biol. Biochem. 16:95–102.

Yeates, G.W. 1991. Nematode populations at three forest sites in New Caledonia. J. Trop. Ecol. 7:411–413.

Yeates, G.W. 1994. Modification and qualification of the nematode maturity index. Pedobiologia 38:97–101.

Yeates, G.W., and A.F. Bird. 1994. Some observations on the influence of agricultural practices on the nematode faunae of some South Australian soils. Fundam. Appl. Nematol. 17:133–145.

Yeates, G.W., and B. Boag. 2003. Background for nematode ecology in the 21st century, p. 406–437. *In* Z.X. Chen, S.Y. Chen, and D.W. Dickson (eds.) Nematology advances and perspectives. CABI Publishing, Wallingford, U.K.

Yeates, G.W., and T. Bongers. 1999. Nematode diversity in agroecosystems. Agric. Ecosyst. Environ. 74:113–135.

Yeates, G.W., T. Bongers, R.G.M. de Goede, D.W. Freckman, and S.S. Georgieva. 1993. Feeding habits in soil nematode families and genera—An outline for soil ecologists. J. Nematol. 25:315–331.

25.3 Microarthropods

Mark G. St. John

D.A. Crossley, Jr.

David C. Coleman

25.3.1 Introduction

Microarthropods are a major fraction of the soil mesofauna, namely, those arthropods with body widths ranging between approximately 0.05 and 2 mm and body lengths between 0.1 and 10 mm (Figure 25.2). This scheme of classification, although imprecise, is practical, defined by the method of sampling. Microarthropods are sampled by collecting a fragment of habitat (e.g., a soil core) and extracting them from it, while macroarthropods (Section 25.4) are usually hand-collected or trapped. Microarthropods are dominated by two groups: springtails (Collembola, a group of near-insect hexapods) and mites (Acari, a subclass of arachnids).
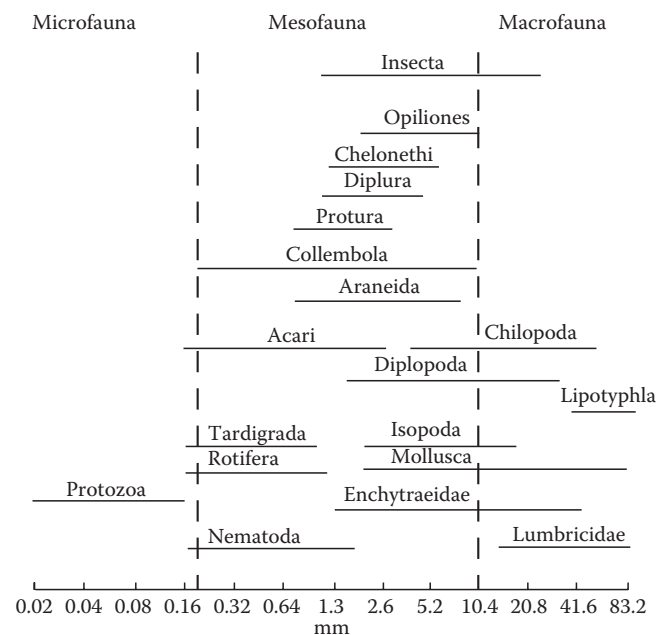
FIGURE 25.2 A generalized classification of soil fauna by body length. Microfauna, mesofauna, and macrofauna plotted on a body length scale from 0.02 to 83.2 mm. (From Wallwork, J.A. 1970. Ecology of soil animals. McGraw Hill, New York.)
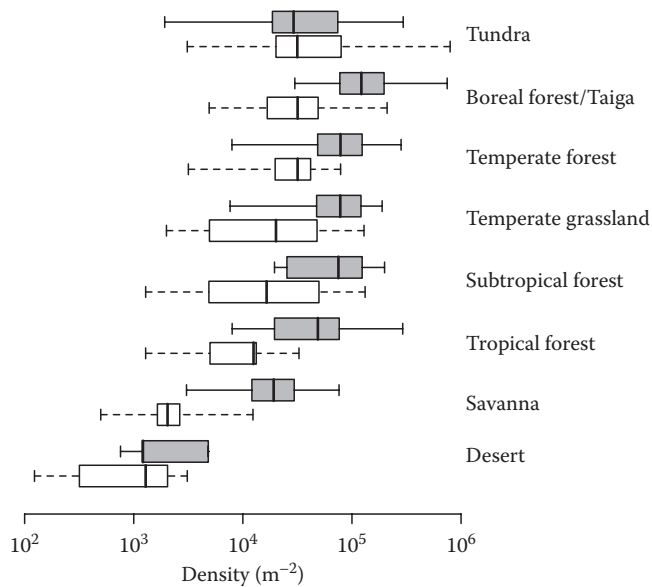
FIGURE 25.3 Mite (gray boxes, solid whiskers) and collembolan (white boxes, dashed whiskers) population densities in eight biomes. The desert biome includes both hot and cold (Arctic and Antarctic) systems. Data are presented as median lines boxed by the interquartile range with the most extreme values as whiskers. (Data from Petersen, H., and M. Luxton. 1982. A comparative analysis of soil fauna populations and their role in decomposition processes. Oikos 39:287–388.)

Together, springtails and mites account for about 90% of the microarthropods in most soil systems. The remaining 10% of soil microarthropods often include Diplura, Protura, Symphyla, Pauropoda, Pseudoscorpiones, fly (Diptera) larvae, thrips (Thysanoptera), and some small spiders (Araneae), true bugs (Hemiptera), and beetles (Coleoptera). Immature stages of many insect orders are collected from soil samples, and some may be considered microarthropods for purposes of a particular study.

Microarthropods are found wherever there is soil, from the Arctic to the tropics to the Antarctic. Densities range widely, with numbers of individuals over 100,000 m^{-2} being typical (Figure 25.3). The assemblage and abundances of species at any one location will be highly dependent on season and depth of sampling. Many studies only consider fauna in the top 5 or 10 cm of soil and assume >80% of the total population is found there; however, microarthropods often migrate to below 10 cm when soils become dry and in desert systems (Petersen and Luxton, 1982). Forested systems generally support higher microarthropod population densities than do grasslands and savannas, deserts, or agricultural systems. Densities tend to be highest in boreal forests (Taiga, temperate coniferous), followed by temperate forests (deciduous, hardwood and mixed) and finally tropical forests. Species richness of microarthropods follows almost the opposite trend, with lower richness at high latitudes than either warm temperate or tropical regions (Maraun et al., 2007). Soils in agroecosystems may have sparse populations, although microarthropod abundance increases under conservation tillage management (Hendrix et al., 1986).

Together with protozoans, nematodes, and other small soil fauna, microarthropods make up a food web of several trophic levels. Their net ecological effects include the breakdown of organic matter and mobilizing nutrient elements. Despite the relatively small biomass of microarthropods in the soil—as little as 10% of the energy flowing through a soil system is assimilated by microarthropods—their indirect action can have large effects (both positive and negative) through control of bacterial and fungal populations (Macfadyen, 1963; Hunt et al., 1987; Huhta, 2007). These influences are strongest in warm temperate and tropical regions (Wall et al., 2008). Microarthropods are also important drivers of soil formation and maintenance, converting plant residues into fine microstructures with their activities and feces (Rusek, 1986). A lesser appreciated role for soil microarthropods is as an essential resource for aboveground predators (Scheu, 2001; Saporito et al., 2007).

Much of our current understanding of the abundance, diversity, and functioning of soil microarthropods (in addition to soil fauna in general) has built on a foundation of work from the International Biological Program (1960s–1970s); a comprehensive resource summarizing this work can be found in Petersen and Luxton (1982). Huhta's (2007) and Coleman's (2008) historical accounts of soil ecological research also provide many essential microarthropod references.

25.3.2 Biology and Ecology

25.3.2.1 Collembola

These minute hexapods are commonly called springtails in reference to an abdominal appendage, the furcula, which enables many species to leap through the air (Figure 25.4). They possess a characteristic ventral tube (collophore, from which their name is derived) with eversible sacs, important in moisture balance. Their mouthparts are internalized like those of the also-wingless Diplura and Protura, collectively earning them the artificial grouping "entognathous hexapods." Arrangement of these groups with the subphylum Hexapoda is still unresolved; however, it is now widely accepted that none of the entognathous hexapods belong to the class Insecta as some authors once believed.

Collembolan numbers in terrestrial ecosystems range between 10^4 and 10^5 m^{-2} and typically constitute 20%–50% of soil microarthropod populations. They are opportunistic species and dominate microarthropod communities where resources are abundant. Some species may be observed as obvious swarms on soil or snow surfaces during population blooms, earning them the name "snowfleas." Their habitat extends from the litter layers of soil down into deeper substrata. Collembolan morphology reflects their habitat in that surface dwellers are larger, often colored, with well-developed furcula and long antennae, while those inhabiting the soil are white, with reduced furcula and shorter antennae (Figure 25.4).

Although most collembolans are fungivores or saprovores, a few species are economic pests, and some have proved to be predaceous. Like other microarthropods, collembolans may feed opportunistically upon nematodes or other periodically abundant resources in the soil. Collembolans appear to be attracted to plant roots and may be important in rhizosphere dynamics. In experiments,

FIGURE 25.5 Scanning electron micrograph of an oribatid mite (*Punctoribates palustris*). (Courtesy of Valerie Behan-Pelletier, Agriculture and Agri-Food Canada.)

FIGURE 25.4 Collembolan life forms. (Syncroscopy images courtesy of Ian Hogg, University of Waikato.)

collembolans grazed selectively on fungal root pathogens, thereby protecting cotton plants (Curl and Truelove, 1986).

For further information, Hopkin (1997) provides an introduction to the Collembola. For more depth, Deharveng (2004) is an excellent resource for both classical and recent collembolan reference works. Additionally, several Internet resources maintained by world authorities are currently available, easily found, and should be consulted for up-to-date information on collembolan literature, biology, and systematics.

25.3.2.2 Soil Mites (Acari or Acarina)

Mites are the second most diverse group of animals on earth, next only to insects, and the vast majority of species live in soil (Figure 25.5). Being members of the class Arachnida, mites are eight-legged, chelicerate relatives of spiders and scorpions. Unlike the typically predatory nature of arachnids, mites have diversified into all manner of additional feeding strategies including plant feeders, parasites, parasitoids, detritivores, fungivores, and scavengers. Mite bodies are unsegmented; however, secondary sutures of their bodies can roughly approximate the separation of tagma (e.g., head, thorax, abdomen, cephalothorax) as found in other arthropods.

Traditionally four suborders of soil mites were recognized (common alternative names are in parentheses): Astigmata (Acaridida), Mesostigmata (Gamasida), Oribatida (Cryptostigmata, Oribatei), and the Prostigmata (Actinedida). Most published works on soil mites subscribed to this taxonomic system; however, decades of investigation (O'Connor, 1984; Norton et al., 1993) have culminated in a new arrangement (Krantz and Walter, 2009). Astigmata is now recognized as a highly derived cohort (Astigmatina) within the Oribatida—but they are so distinct in terms of biology and morphology that it is sensible to continue treating them as separate groups. Mesostigmata has been elevated to the level of order with little change while some members of the Prostigmata have been moved into the new suborders Sphaerolichida (order Trombidiformes) and Endeostigmata (order Sarcoptiformes). These six mite groups are considered in the following sections using the most current systematic understanding, in a manner that allows reconciliation of data reported under past taxonomic schemes.

Each of these groups (as well as mite groups not typically found in soil) is covered in more depth, with illustrated keys to families of the world in relevant chapters of Krantz and Walter (2009). Also see Evans (1992) and Walter and Proctor (1999) for more information on the biology and ecology of mites. For more in-depth coverage, it is recommended to attend the soil mites component of the Acarology Summer Program held at The Ohio State University, Columbus, Ohio. Understanding of this group is in rapid flux and the most current information and materials are disseminated by world experts here, biannually.

25.3.2.2.1 Order Mesostigmata (Parasitiformes)

Mesostigmata are generally flattened, tick-like mites. Although not as numerous as other mite groups, mesostigmatids are

important predators in soil systems, particularly on nematodes and other arthropods. They range in length from 0.2 to 3 mm. Smaller species are mainly nematophagous. Larger ones are active predators of collembolans, small arthropods, and their eggs. A few species within the Uropodina, Digamasellidae, and Ascidae are atypically fungivore–saprovores. In forest floors, litter species tend to be larger and well armored (sclerotized), and mineral soil inhabitants are generally smaller, softer, and colorless. In agricultural soils, Mesostigmata are attracted to decomposing roots. Mesostigmatids are also championed as useful bioindicators in a variety of systems (Beaulieu and Weeks, 2007, and references therein) because of their predatory role and *r*-strategist lifestyle. At least one species, *Geolaelaps* (*Hypoaspis*) *aculeifer*, is a standard organism in ecotoxicological tests.

A large and comprehensive, international literature discusses soil mesostigmatid mites but the group is in need of a comprehensive revision. Useful references include Evans and Till (1979), Krantz and Ainscough (1990), and Karg (1993).

25.3.2.2.2 Suborder Prostigmata (Acariformes: Trombidiformes)

Representatives of this very diverse suborder are often among the most numerous of the soil mites. In general, they are delicate, white to colorless, and subject to desiccation. However, any brightly colored mites encountered are almost certain to be prostigmatids as well. They range widely in size from about 0.1 to several millimeters in length.

Functional generalizations about the Prostigmata are difficult. While many soil-dwelling species are predaceous, some are fungivorous and these species may become abundant in decomposing organic litter. Many Prostigmata are opportunistic species, able to reproduce rapidly when food resources become abundant. Large populations may build in disturbed situations, such as forest clearings, drained marshes, and so forth. Larger Prostigmata are predators of other arthropods, including the bright red velvet mites seen walking on the soil surface in the spring, or following rains in desert systems. Smaller species with piercing stylet chelicerae are generally fungal feeders in soil and litter layers, but some are effective predators on small arthropods or nematodes. Some species may switch between fungal hyphae and nematodes, or perhaps supplement a fungal diet with occasional protein. Additionally, phytophagous species are frequently recovered in soil samples, further weakening generalizations about this suborder. However, prostigmatid behaviors are relatively stable, and morphologies distinct, at the family level. Identification of prostigmatid families (or further) is profitable, if not prudent, in ecological studies where they are abundant.

Kethley (1990) provides a comprehensive treatment of the Prostigmata. He gives extensive tables of abundance and biomass for the Prostigmata in a range of habitats, as well as a key to families commonly found in soil.

25.3.2.2.3 Suborder Sphaerolichida (Acariformes: Trombidiformes)

This suborder is composed of only two monogeneric families: Sphaerolichidae (*Sphaerolichus*) and Lordalycidae (*Hybalicus*).

Sphaerolichida tend to be small, pink, globular mites resembling species of Endeostigmata, of which they were traditionally considered to be members. While of cosmopolitan distribution, species of Sphaerolichida are rarely encountered in great numbers and they have received little study. There is evidence of particulate fungal-feeding in species of both genera although species of *Sphaerolichus* are more likely to be ambush predators.

25.3.2.2.4 Suborder Endeostigmata (Acariformes: Sarcoptiformes)

There are currently 10 recognized families of Endeostigmata, which have been treated as Prostigmata by some authors. They are mostly small (0.2–1.0 mm), globular to cylindrical, pale white to pink, and weakly armored mites, often with highly modified setae (hairs) and/or patterned cuticles; but nematode wormlike forms specializing in deep-soil habitats can also be found. Many appear to be detritivores or omnivores, some have piercing mouthparts, which would suggest fungal-feeding, while others are assumed predaceous on nematodes.

25.3.2.2.5 Suborder Oribatida (Acariformes: Sarcoptiformes), Exclusive of Astigmatina

Oribatid mites (Figure 25.5) are often the most numerous of microarthropods in soils. Most adult oribatids range from 0.2 to 0.7 mm in length, are often brown and beetle-like in form (sometimes called beetle, armored, or moss mites). The armored appearance of adults is usually from sclerotization (and melanization) of adult exoskeletons although some species harden their integuments with Ca deposits (e.g., $CaCO_3$), evidently accumulated from their fungal food sources. Together with snails, millipedes, and isopods, oribatids may play a significant role in Ca metabolism in soil systems. The immature stages of many oribatid mites are typically weakly sclerotized (sometimes being mistaken for Prostigmata or Endeostigmata) and do not resemble adults to the extent that recognition of species based on immatures is often impossible.

Densities of oribatid mites in soils, as with Collembola, typically range from 10^4 to 10^5 m^{-2}. In deciduous forests, peaks of abundance occur in autumn and spring, with numbers remaining high during summer months. Year-to-year variation may be large, but fluctuations are not as wide as with Collembola. Coniferous forest (boreal/Taiga) floors typically support the largest populations of oribatids, followed by temperate deciduous hardwood forest, grasslands/savanna, and tundra. Any disturbance of soil (e.g., cultivation of agricultural soils, introduction of bioturbators) typically reduces oribatid population sizes (Maraun et al., 2003).

Oribatids are similar to Collembola in being primarily particulate-feeding fungivores and saprophages that may ingest nematodes opportunistically; however, their rates of respiration and reproduction are much lower than collembolans. Oribatids maintain similarly high populations by longer survivorship. As a group they are often considered to be generalist feeders, but individual species are quite specific in what they consume (Schneider and Maraun, 2005) and immatures may

feed on different resources than adults (Siepel, 1990). Oribatids are important drivers of decomposition and soil formation. They are able to break down organic litter by fragmenting or tunneling within materials such as leaves and stems with their stout chelate chelicerae while creating fine soil microstructure with their feces. Their activities may stimulate microbial immobilization of nutrient elements, but at the same time, ingestion of fungal hyphae may destroy hyphal bridges (Lussenhop, 1992).

There are approximately 10–100 times more asexual species within the oribatids than is typical for other animals. Asexuality evolved multiple times within Oribatida, in ancient clades, with sexuality re-evolving in some taxa (Domes et al., 2007).

Oribatid mites are often the focus of soil biodiversity studies or used in bioindication (Behan-Pelletier, 1999) because they have a high local (alpha) diversity (an organic temperate forest soil will have upward of 100 species); are found wherever there is soil; are relatively well documented; and their *K*-strategist life history makes them sensitive indicators of soil disturbance and rehabilitation (St. John et al., 2002). The diversity of oribatids is a function of microhabitats, not abundance (Aoki, 1967; Osler and Beattie, 1999; Hansen, 2000). There are over 9000 described species of oribatid mites (Schatz, 2002), which is estimated to be only 20% of the actual world fauna (Behan-Pelletier and Bissett, 1992). The identification of oribatids to genus has been greatly facilitated by Janos Balogh, who has authored a series of well-illustrated works (Balogh and Balogh, 1992).

25.3.2.2.6 *Cohort Astigmatina (Acariformes: Sarcoptiformes: Oribatida)*

This diverse group of mites contains pests of stored products (commonly called cheese mites), and vertebrate commensals and parasites; however, soil-dwelling astigmatids are typically saprophagous or fungivorous. Soil astigmatids are pale white, soft, globular to cylindrical mites ranging approximately 0.2–1.8 mm in length often with prominent, long setae. Immature stages modified for dispersal (hypopus, *pl.* hypopi) are somewhat lens-like in shape (bearing a passing resemblance to mites in prostigmatid family Scutacaridae) with a prominent set of "suckers" on their venter used for affixing themselves to other arthropods (phoresy). Astigmatid adults and immatures are only occasionally found in soil samples but can appear in great numbers under favorable conditions due to their extreme *r*-strategist nature. These mites seem to be associated with highly organic, decomposing materials such as manure. Buried agricultural residues may support astigmatid blooms, and they can become pests of root crops on occasion.

25.3.3 Sampling and Analysis

Microarthropods are not sampled directly; rather, samples of habitat—typically cores 5 cm diameter × 5 cm deep—are collected and microarthropods are extracted from them (Figure 25.6). Extraction can be either "active" by encouraging the animals to move along a humidity and temperature gradient out of the soil (e.g., Macfadyen, 1962; Crossley and Blair, 1991), or "passive" by using physical properties to separate animals from mineral soil (e.g., Walter et al., 1987). See Edwards (1991) for a review of these methods and their variants. An alternate sampling method employs litterbags containing leaf litter, which are placed in the field at the beginning of the season and then retrieved through time (Crossley and Hoglund, 1962). Using this technique, it is possible to identify microarthropod groups associated with various stages of litter decay.
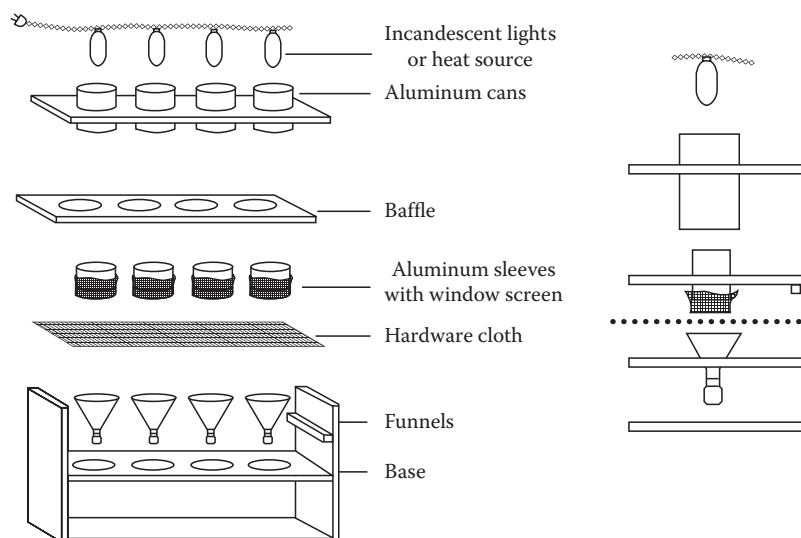
FIGURE 25.6   A high-efficiency extractor for core samples of soil microarthropods. Soil cores contained in aluminum sleeves are inverted and heated from above with 5 W lights. Fauna are collected in funnels below. (Reprinted from Crossley, D.A., and J.M. Blair. 1991. A high-efficiency, low-technology tullgren-type extractor for soil microarthropods. Agric. Ecosyst. Environ. 34:187–192, Copyright (1991), with permission from Elsevier.)

TABLE 25.6 Formulations of Microarthropod Storage Fluids, Clearing Agents, and Slide Mounting Media

Storage fluids	Ethyl alcohol (70%–95%) Oudeman's fluid: glycerin (5 parts); 70% ethyl alcohol (87 parts); and glacial acetic acid (8 parts)
Clearing agents	Lactic acid Nesbitt's fluid: chloral hydrate (40 g); lactic or hydrochloric acid (2.5 mL); and distilled water (25 mL)
Mounting media	Heinze PVA: Stir polyvinyl alcohol powder (10 g) in distilled water (60 mL) under low heat; add lactic acid (35 mL) then glycerol (10 g), stir until smooth; cool until lukewarm; dissolve chloral hydrate (100 g) into mixture; filter or centrifuge to remove impurities

Extracted microarthropods are typically stored in 70% ethanol (Table 25.6) and may be sorted into major taxa with the use of a dissecting microscope at 20 times magnification. Major groups of mites and many families of collembolans are readily identified, once their basic morphology and characteristics are learned. Identification to species requires slide-mounted material. Slide mounts can be made directly from alcohol, although heavily sclerotized specimens may require preliminary clearing (Table 25.6). Identification of North American microarthropod species is difficult because of the large number of poorly described or unnamed species (Behan-Pelletier and Bissett, 1992). While it is essential that identifications be verified by an expert taxonomist specializing in the taxon, enumeration of each of the major taxa (above) is often adequate for comparative studies of soil microarthropods.

The abundance and richness of microarthropods in organic soil and litter samples can be overwhelmingly high for study, necessitating processing strategies that reduce workloads and expert involvement. Sample volume reduction can be used with minimal loss of information (Santos et al., 2008). Alternatively, molecular methods commonly used for bacteria and fungi can "probe" for soil microarthropods in a sample (Gibb et al., 2008; Hamilton et al., 2009; Wu et al., 2009). These methods have the potential to vastly reduce required workloads in soil investigations, but with some important trade-offs. Only species richness estimates (not species identities) are possible unless the analysis is coupled to a database of expert identifications with corresponding molecular profiles, and there is no defensible method to derive abundances from the analysis. Also, the methods often employ sample sizes of soil (1–2 g) that are under-representative of the scale an assemblage of microarthropods exist at. This can be remedied with the use of traditional extraction methods to concentrate microarthropods into a smaller volume of fluid.

It is difficult to answer basic ecological questions about soil microarthropods (e.g., "what do they eat?") because direct observation in situ is practically impossible. Feeding preference studies, observations of gut contents, radioisotope tracers, and enzymatic analyses have all been used to infer ecological roles of a limited set of taxa. Analysis of stable isotopes and fatty acids has recently added resolution and some surprises to our understanding of microarthropods. For example, energy flowing through microarthropods (and the "detrital" food web in general) may originate primarily from roots and root exudates, not litter as has been traditionally assumed (Garrett et al., 2001; Ostle et al., 2007; Pollierer et al., 2009). Also, we now know that supposedly generalist fungivore/detritivore microarthropod species have highly specific, consistent diets (Schneider et al., 2004; Chahartaghi et al., 2005; Illig et al., 2005; Ruess et al., 2005). Low temperature scanning electron microscopy (LTSEM) and x-ray tomography may also prove useful in soil microarthropod ecological research. With LTSEM, researchers are able to observe ecological interactions by literally freezing the animals in situ (Achor et al., 2001), while x-ray tomography and holotomography allows researchers to noninvasively study microarthropods internally in three dimensions (Heethoff and Cloetens, 2008; Heethoff et al., 2008).

References

Achor, D.S., R. Ochoa, E.F. Erbe, H. Aguilar, W.P. Wergin, and C.C. Childers. 2001. Relative advantages of low temperature versus ambient temperature scanning electron microscopy in the study of mite morphology. Int. J. Acarol. 27:3–12.

Aoki, J. 1967. Microhabitats of oribatid mites on a forest floor. Bull. Nat. Sci. Mus. Tokyo 10:133–138.

Balogh, J., and P. Balogh. 1992. The oribatid mites genera of the world. Vol. 1. The Hungarian National Museum Press, Budapest, Hungary.

Beaulieu, F., and A.R. Weeks. 2007. Free-living mesostigmatic mites in Australia: Their roles in biological control and bioindication. Aust. J. Exp. Agric. 47:460–478.

Behan-Pelletier, V.M. 1999. Oribatid mite biodiversity in agroecosystems: Role for bioindication. Agric. Ecosyst. Environ. 74:411–423.

Behan-Pelletier, V.M., and B. Bissett. 1992. Biodiversity of Nearctic soil arthropods. Can. Biodivers. 2:5–14.

Chahartaghi, M., R. Langel, S. Scheu, and L. Ruess. 2005. Feeding guilds in collembola based on nitrogen stable isotope ratios. Soil Biol. Biochem. 37:1718–1725.

Coleman, D.C. 2008. From peds to paradoxes: Linkages between soil biota and their influences on ecological processes. Soil Biol. Biochem. 40:271–289.

Crossley, D.A., and J.M. Blair. 1991. A high-efficiency, low-technology tullgren-type extractor for soil microarthropods. Agric. Ecosyst. Environ. 34:187–192.

Crossley, D.A., and M.P. Hoglund. 1962. A litter-bag method for the study of microarthropods inhabiting leaf litter. Ecology 43:571–573.

Curl, E.A., and B. Truelove. 1986. The rhizosphere. Vol. 15. Advanced series in agricultural sciences. Springer, Berlin, Germany.

Deharveng, L. 2004. Recent advances in collembola systematics. Pedobiologia 48:415–433.

Domes, K., R.A. Norton, M. Maraun, and S. Scheu. 2007. Reevolution of sexuality breaks Dollo's law. Proc. Natl. Acad. Sci. USA 104:7139–7144.

Edwards, C.A. 1991. The assessment of populations of soil-inhabiting invertebrates. Agric. Ecosyst. Environ. 34:145–176.

Evans, G.O. 1992. Principles of acarology. CABI, Wallingford, U.K.

Evans, G.O., and W.M. Till. 1979. Mesostigmatic mites of Britain and Ireland (Chelicerata: Acari-parasitiformes). An introduction to their morphology and classification. Trans. Zool. Soc. Lond. 35:1–39.

Garrett, C.J., D.A. Crossley, D.C. Coleman, P.F. Hendrix, K.W. Kisselle, and R.L. Potter. 2001. Impact of the rhizosphere on soil microarthropods in agroecosystems on the Georgia piedmont. Appl. Soil Ecol. 16:141–148.

Gibb, K., J. Beard, P. O'Reagain, K. Christian, V. Torok, and K. Ophel-Keller. 2008. Assessing the relationship between patch type and soil mites: A molecular approach. Pedobiologia 51:445–461.

Hamilton, M., S. Strickland, K. Wickings, and M.A. Bradford. 2009. Surveying soil faunal communities using a direct molecular approach. Soil Biol. Biochem. 41:1311–1314.

Hansen, R.A. 2000. Effects of habitat complexity and composition on a diverse litter microarthropod assemblage. Ecology 81:1120–1132.

Heethoff, M., and P. Cloetens. 2008. A comparison of synchrotron x-ray phase contrast tomography and holotomography for non-invasive investigations of the internal anatomy of mites. Soil Org. 80:205–215.

Heethoff, M., L. Helfen, and P. Cloetens. 2008. Non-invasive 3D-visualization with sub-micron resolution using synchrotron-x-ray-tomography. J. Vis. Exp. 15:737.

Hendrix, P.F., R.W. Parmelee, D.A. Crossley, Jr., D.C. Coleman, E.P. Odum, and P.M. Groffman. 1986. Detritus food webs in conventional and no-tillage agroecosystems. Bioscience 36:374–380.

Hopkin, S.P. 1997. Biology of the springtails (Insecta: Collembola). Oxford University Press, Oxford, U.K.

Huhta, V. 2007. The role of soil fauna in ecosystems: A historical review. Pedobiologia 50:489–495.

Hunt, H.W., D.C. Coleman, E.R. Ingham, R.E. Ingham, E.T. Elliott, J.C. Moore, S.L. Rose, C.P.P. Reid, and C.R. Morley. 1987. The detrital food web in a shortgrass prairie. Biol. Fertil. Soils 3:57–68.

Illig, J., R. Langel, R.A. Norton, S. Scheu, and M. Maraun. 2005. Where are the decomposers? Uncovering the soil food web of a tropical montane rain forest in southern Ecuador using stable isotopes (^{15}N). J. Trop. Ecol. 21:589–593.

Karg, W. 1993. Acari (Acarina), Milben. Parasitiformes (Anactinochaeta). Cohors Gamasina Leach. Raubmilben. 2. Überarbeitete Auflage. *In* F. Dahl (ed.) Die Tierwelt Deutschlands. Vol. 59. Teil. G. Fischer-Verlag, Jena/Stuttgart/New York.

Kethley, J. 1990. Acarina: Prostigmata (Actinedida), p. 667–778. *In* D.L. Dindal (ed.) Soil biology guide. John Wiley & Sons, New York.

Krantz, G.W., and B.D. Ainscough. 1990. Acarina: Mesostigmata (Gamasida), p. 583–665. *In* D.L. Dindal (ed.) Soil biology guide. John Wiley & Sons, New York.

Krantz, G.W., and D.E. Walter. 2009. A manual of acarology. 3rd edn. Texas Tech University Press, Lubbock, TX.

Lussenhop, J. 1992. Mechanisms of microarthropod microbial interactions in soil. Adv. Ecol. Res. 23:1–33.

Macfadyen, A. 1962. Control of humidity in three funnel-type extractors for soil arthropods, p. 158–168. *In* P. Murphy (ed.) Progress in soil zoology. Butterworths, London, U.K.

Macfadyen, A. 1963. The contribution of the microfauna to total soil metabolism, p. 3–17. *In* J. Doeksen and J. van der Drift (eds.) Soil organisms. North Holland Publishing Company, Amsterdam, the Netherlands.

Maraun, M., J.A. Salamon, K. Schneider, M. Schaefer, and S. Scheu. 2003. Oribatid mite and collembolan diversity, density and community structure in a moder beech forest (*Fagus sylvatica*): Effects of mechanical perturbations. Soil Biol. Biochem. 35:1387–1394.

Maraun, M., H. Schatz, and S. Scheu. 2007. Awesome or ordinary? Global diversity patterns of oribatid mites. Ecography 30:209–216.

Norton, R.A., J.B. Kethley, D.E. Johnston, and B.M. O'Connor. 1993. Phylogenetic perspectives on genetic systems and reproductive modes of mites, p. 8–98. *In* D. Wrensch and M. Ebbert (eds.) Evolution and diversity of sex ratio in insects and mites. Chapman and Hall, New York.

O'Connor, B.M. 1984. Phylogenetic relationships among higher taxa in the Acariformes, with particular reference to the Astigmata. Acarology 6:19–27.

Osler, G.H.R., and A.J. Beattie. 1999. Taxonomic and structural similarities in soil oribatid communities. Ecography 22:567–574.

Ostle, N., M.J.I. Briones, P. Ineson, L. Cole, P. Staddon, and D. Sleep. 2007. Isotopic detection of recent photosynthate carbon flow into grassland rhizosphere fauna. Soil Biol. Biochem. 39:768.

Petersen, H., and M. Luxton. 1982. A comparative analysis of soil fauna populations and their role in decomposition processes. Oikos 39:287–388.

Pollierer, M.M., R. Langel, S. Scheu, and M. Maraun. 2009. Compartmentalization of the soil animal food web as indicated by dual analysis of stable isotope ratios (^{15}N/^{14}N and ^{13}C/^{12}C). Soil Biol. Biochem. 41:1221–1226.

Ruess, L., A. Tiunov, D. Haubert, H.H. Richnow, M.M. Haggblom, and S. Scheu. 2005. Carbon stable isotope fractionation and trophic transfer of fatty acids in fungal based soil food chains. Soil Biol. Biochem. 37:945–953.

Rusek, J. 1986. Soil microstructures—Contributions on specific soil organisms. Quaest. Entomol. 21:497–514.

Santos, E.M.R., E. Franklin, and W.E. Magnusson. 2008. Cost-efficiency of subsampling protocols to evaluate oribatid-mite communities in an Amazonian savanna. Biotropica 40:728–735.

Saporito, R.A., M.A. Donnelly, R.A. Norton, H.M. Garraffo, T.F. Spande, and J.W. Daly. 2007. Oribatid mites as a major dietary source for alkaloids in poison frogs. Proc. Natl. Acad. Sci. USA 104:8885–8890.

Schatz, H. 2002. Die oribatidenliteratur und die beschriebenen oribatidenarten (1758–2001)—Eine analyse. Abh. Ber. Des Naturkundemuseums Görlitz 72:37–45.

Scheu, S. 2001. Plants and generalist predators as links between the below-ground and above-ground system. Basic Appl. Ecol. 2:3–13.

Schneider, K., and M. Maraun. 2005. Feeding preferences among dark pigmented fungal taxa ("Dematiacea") indicate limited trophic niche differentiation of oribatid mites (Oribatida, Acari). Pedobiologia 49:61–67.

Schneider, K., S. Migge, R.A. Norton, S. Scheu, R. Langel, A. Reineking, and M. Maraun. 2004. Trophic niche differentiation in soil microarthropods (Oribatida, Acari): Evidence from stable isotope ratios ($^{15}N/^{14}N$). Soil Biol. Biochem. 36:1769–1774.

Siepel, H. 1990. Niche relationships between two panphytophagous soil mites, *Nothrus silvestris* nicolet (Acari, Oribatida, Nothridae) and *Platynothrus peltifer* (Koch) (Acari, Oribatida, Camisiidae). Biol. Fertil. Soils 9:139–144.

St. John, M.G., G. Bagatto, V. Behan-Pelletier, E.E. Lindquist, J.D. Shorthouse, and I.M. Smith. 2002. Mite (Acari) colonization of vegetated mine tailings near Sudbury, Ontario, Canada. Plant Soil 245:295–305.

Wall, D.H., M.A. Bradford, M.G. St. John, J.A. Trofymow, V. Behan-Pelletier, D.D.E. Bignell, J.M. Dangerfield et al. 2008. Global decomposition experiment shows soil animal impacts on decomposition are climate-dependent. Global Change Biol. 14:2661–2677.

Walter, D.E., J. Kethley, and J.C. Moore. 1987. A heptane flotation method for recovering microarthropods from semiarid soils, with comparison to the Merchant-Crossley high-gradient extraction method and estimates of microarthropod biomass. Pedobiologia 30:221–232.

Walter, D.E., and H.C. Proctor. 1999. Mites: Ecology, evolution and behaviour. CABI, Wallingford, U.K.

Wallwork, J.A. 1970. Ecology of soil animals. McGraw Hill, New York.

Wu, T.H., E. Ayres, G. Li, R.D. Bardgett, D.H. Wall, and J.R. Garey. 2009. Molecular profiling of soil animal diversity in natural ecosystems: Incongruence of molecular and morphological results. Soil Biol. Biochem. 41:849–857.

25.4 Macroarthropods

M.A. Callaham, Jr.

D.A. Crossley, Jr.

David C. Coleman

25.4.1 Introduction

The macroarthropods are those large enough to be sampled as individuals, in contrast to the microarthropods that are sampled by extraction from a fragment of habitat (Section 25.3; Dindal, 1990; Borror et al., 1992; Arnett, 1993). Although smaller macroarthropods overlap in size with the larger microarthropods (Figure 25.2), the distinction between them is a practical one, based on method of sampling. A functional difference lies in their impact on soils. Macroarthropods are capable of restructuring soil profiles or relocating large amounts of soil, whereas microarthropods typically inhabit (and do little to modify) the existing pore spaces in soil (Coleman et al., 2004). Two insect groups, ants and termites, are responsible for major disruptions of soil profiles and have thus been classified as ecosystem engineers (Jones et al., 1997; Jouquet et al., 2006), while other macroarthropods may cause some disturbance. Examples include emergence tunnels of periodical cicadas (Insecta: Homoptera) (Whiles et al., 2001), or chimneys made by terrestrial crayfish (Crustacea: Decapoda) in hydric soils (Welch et al., 2008).

The macroarthropods in soil systems are a highly diverse group. Most terrestrial insect orders contain species that live in the soil at some phase of their life cycle. Transient species (some Lepidoptera), are those that overwinter or pupate in surficial soil layers. Other temporary residents such as midges and other flies spend their immature stages in the soil but emerge as adults to feed elsewhere. Permanent soil residents, such as predaceous beetles, remain in the soil or on soil surfaces. Spiders (Araneae) and centipedes (Chilopoda) are numerous and important predators in soil systems (Kastan, 1978; Camatini, 1979; Kevan and Scudder, 1989; Foelix, 1996). Detrivores include millipedes (Diplopoda) and sowbugs (Isopoda) (Shachak et al., 1976; Camatini, 1979; Snider and Shaddy, 1980). Scorpions (Scorpionida) and windscorpions (Solifugae) are important predators in desert systems (Crawford, 1981; Williams, 1987). Ants fall into several feeding guilds, and are diverse, abundant, and nearly ubiquitous in tropical and temperate ecosystems (Keller and Gordon, 2009). Several groups of macroarthropods (e.g., Coleoptera, Diptera, Lepidoptera) are considered pests in agricultural systems and can cause significant economic impacts either through crop and forage damage, or through expenses associated with their control (e.g., McCracken et al., 1995; Jackson and Klein, 2006; Doğramaci and Tingey, 2009).

25.4.2 Biology and Ecology

The major groups of macroarthropods likely to be found in soil and litter samples are listed in Table 25.7. The list is not inclusive

TABLE 25.7   Major Groups of Macroarthropods

Class	Order	Common Name(s)
Arachnida	Araneae	Spiders
	Scorpiones	Scorpions
	Opiliones	Phalangids, harvestmen
	Pseudoscorpiones	Pseudoscorpions
	Solifugae	Windscorpions
Malacostraca	Isopoda	Sowbugs, pillbugs, woodlice, roly-polies
Diplopoda	Ten orders	Millipedes
Chilopoda	Four orders	Centipedes
Hexapoda	Hymenoptera	Ants, wasps
	Isoptera	Termites
	Coleoptera	Beetles, rove beetles, tiger beetles, white grubs, wireworms
	Diptera	Flies, clusterflies, midges, leatherjackets, maggots
	Lepidoptera	Armyworms, potato tuberworms
	Homoptera	Cicadas, leafhoppers
	Dermaptera	Earwigs
	Diplura	Diplura
	Protura	Protura
	Neuroptera	Antlions
	Twenty-one other orders	

because, aside from representatives of the minor orders of arachnids and insects not listed, samples may also include representatives of major orders not usually considered to be soil fauna. Grasshoppers and crickets are frequently found on the soil surface and in pitfall traps. Even caterpillars that have descended from plant canopies to pupate in soil will be sampled. In fact, nearly every free-living group of terrestrial arthropods may occasionally enter soil food webs as prey items.

The majority of the scientific literature deals with individual taxa in detail, rather than providing overviews of entire macroarthropod faunas. Where general overviews do exist, they attempt broad syntheses, often without detailed information about macroarthropods (Dindal, 1990; Borror et al., 1992). The biology and importance of some ecological groups, such as root feeders, remain poorly known (Hunter, 2001; Blossey and Hunt-Joshi, 2003; Coleman et al., 2004; Johnson and Murray, 2008).

Macroarthropod fauna vary considerably between and within types of ecosystems. Several workers have suggested that macroarthropods can be used as indicators of soil quality in agricultural or pollution contexts (Linden et al., 1994; Kuperman, 1996; Lobry de Bruyn, 1997), and others have indicated that belowground arthropod herbivores have potential for use as biological control agents for invasive plant species, but this potential is largely unexplored relative to aboveground insect herbivores (Blossey and Hunt-Joshi, 2003). Forested ecosystems in general contain macroarthropod fauna dominated by millipedes, spiders, flies, and beetles (Table 25.8). Numbers and biomass tend to be greater in hardwood than in evergreen forests, where microarthropod abundance is high (Section 25.3). Spiders, carabid beetles, and crickets are abundant in pitfall traps placed in agricultural areas (Blumberg and Crossley, 1983; House and Stinner, 1983). Most species of macroarthropods may be more sensitive to cultivation than other soil fauna, and as a result, investigations have neglected the sensitive species in favor of the abundant ones (Wolters and Ekschmitt, 1997). Spiders are the most abundant and probably the most important of the predaceous macroarthropods in terms of their impact on food webs (Ekschmitt et al., 1997; Lawrence and Wise, 2004; Wise, 2004). The ranges of macroarthropod population sizes vary widely. Considerable overlap in abundance of macroarthropod taxa among different ecosystems has been found for millipedes and centipedes, although arable land tends to contain the lower part of the range. The ranges listed in Tables 25.8 and 25.9 illustrate the differences that occur between habitat types and seasons. Adjacent forest, grassland, and agricultural lands often have markedly different species within more general taxonomic groupings such as spiders (Reichert and Lockly, 1984; Draney, 1992). The range in measured abundance is also somewhat affected by method of sampling, and the size-based definition used to delineate "macroarthropod" (Table 25.8, Figure 25.2), and additional difficulty in establishing accurate density estimates arises from the nonrandom, often highly aggregated nature of macroarthropod distribution.

25.4.2.1 Social Insects

Collectively, ants and termites are responsible for major modifications of soil. Termites are typical insects of tropical and subtropical regions and may be dominant in arid or semiarid ecosystems (Lee and Wood, 1971; Bryan, 1978; Hölldobler and Wilson, 1990; Stork and Eggleton, 1992; Bolton, 1994; Arriaga and Maya, 2007). The following soil modifications are brought about by the activities of termites: (1) physical changes of soil profiles, (2) changes in soil structure, (3) changes in the nature and distribution of organic matter, (4) changes in the distribution of plant nutrients, and (5) construction of subterranean galleries (macropores). Ants are more widely distributed than termites, occurring in most terrestrial habitats, and exhibit a variety of trophic behaviors including herbivory (foliar and seed feeding), omnivory, and many are predaceous. Their colonies are smaller than those of termites, but they are responsible for the same kinds of soil modifications. In tropical systems, leaf-cutter ants are among the most important arthropod herbivores with dramatic effects on incorporation of organic materials to belowground pools, and these ants have been implicated as affecting forest plant community dynamics when they reach high abundances (particularly in forest fragments where their predators are absent) (Terborgh et al., 2001). As predators, ants may have a considerable impact on herbivorous insects. The importance of ants as predators of insect pests has been well demonstrated in a variety of stable forest ecosystems. Close, evidently coevolved, relationships exist between some plant and ant species, but in relatively unstable, annual agroecosystems, less is known of the importance of the ants.

TABLE 25.8 Ranges and Mean Abundance (Number m⁻²) of Selected Major Macroarthropod Taxa Reported from Differing Ecosystems Worldwide

Habitat Type	Location[a]	Diplopoda	Diptera (Larvae)	Araneae	Coleoptera (Larvae)	Chilopoda	Method[b]	Sources
Grassland	KS	~5–74	—	~35	~5–30	3–6	HS	Blair et al. (2000), Callaham et al. (2003)
Upland pine	FL	241–276	8–12	12–19	43–82	4–8	TF	Frouz and Ali (2004)
Upland hardwood	FL	23–66	~8	16–31	27–47	4–16	TF	Frouz and Ali (2004)
Spruce forest	FIN	—	—	~700	~371	—		Siira-Pietikäinen et al. (2003)
Mediterranean pine	ISL			40–214			BF	Broza and Izhaki (1997)
Beech forest	FRA	30–40	12–23	16–27	43–52	~46	HS	Aubert et al. (2002)
Mixed forest	FRA	15–17	15–56	15–20	37–41	18–21	HS	Aubert et al. (2002)
Clear-cut with slash	SWE	39	238	289	224	—	TF	Bengtsson et al. (1997)
Clear-cut no slash	SWE	74	106	152	133	—	TF	Bengtsson et al. (1997)
Pasture	FRA	0	—	11	62	16	HS	Decaëns et al. (1998)
44 year fallow	FRA	1	—	13	17	58	HS	Decaëns et al. (1998)
Maple forest	FRA	3	—	6	15	63	HS	Decaëns et al. (1998)
Pine forest	FRA	0	—	7	0.5	39	HS	Decaëns et al. (1998)
Planted birch in forest soil	FIN	0–6	133–208	129–196	338–606	55–95	HSM	Huhta (2002)
Planted birch in agricultural soil	FIN	0–8	126–265	101–156	220–367	2–10	HSM	Huhta (2002)
Natural birch	FIN	0–80	173–360	247–340	220–814	20–91	HSM	Huhta (2002)
Deciduous forest	CAN	139	1003	—	288	—	HSM	Paquin and Coderre (1997)
Mixed forest	CAN	64	299	—	277	—	HSM	Paquin and Coderre (1997)
Coniferous forest	CAN	32	32	—	85	—	HSM	Paquin and Coderre (1997)
Atlantic forest	BRA	~20–320	~100–380	~50–100	~100–230	~10–50	BTF	Pellens and Garay (1999)
Eucalypt plantation	BRA	~0–320	~40–100	~10–120	~100–200	~0–2	BTF	Pellens and Garay (1999)
Acacia plantation	BRA	~0–100	~20–240	~0–100	~230–770	~0–10	BTF	Pellens and Garay (1999)
Norway spruce forest	ITA	24	0	—	835	287	DF	Salmon et al. (2006)
Regenerating spruce	ITA	47–119	287–454	—	239–287	478–957	DF	Salmon et al. (2006)

Notes: When data are presented as a range, this indicates that the authors were reporting mean densities from different sites of similar vegetation or were reporting densities from the same site from different seasons. When data are preceded by "~," this indicates that numbers were estimated (or recalculated) from graphical data and not transcribed directly from tabular data in source materials or derived from multiple sources.

[a] KS, Kansas; FL, Florida; FIN, Finland; ISL, Israel; FRA, France; SWE, Sweden; CAN, Canada; BRA, Brazil; ITA, Italy.

[b] HS, hand sorting; TF, Tullgren funnel; BF, Berlese funnel; HSM, hand sorting under dissecting microscope; BTF, Berlese–Tullgren funnel; DF, dry funnel, all as reported by authors in source materials.

TABLE 25.9 Abundance and Biomass of Macroarthropod Taxa in Three Ecosystems

Taxon	Arable Land N m⁻²	Arable Land mg m⁻²	Temperate Grassland N m⁻²	Temperate Grassland mg m⁻²	Temperate Deciduous Forest N m⁻²	Temperate Deciduous Forest mg m⁻²
Isopoda	5 (0–25)	15	1200 (500–7900)	1600	286 (96–1850)	93
Diplopoda	200 (70–400)	—	— (500–7900)	1250	55 (210–700)	618
Chilopoda	100 (40–220)	—	60 (63–387)	140	187 (50–790)	265

Source: Wolters, V., and K. Ekschmitt. 1997. Gastropods, isopods, diplopods and chilopods: Neglected groups of the decomposer food web, p. 265–306. *In* G. Benckiser (ed.) Fauna in soil ecosystems. Marcel Dekker, New York.

Numbers in parentheses indicate ranges of abundance estimates.

25.4.2.2 Myriapods

The many-legged arthropods (myriapods) are abundant in undisturbed soils of many types, but less abundant in agricultural systems (Table 25.8). Millipedes are major saprovores, feeding upon decomposing organic matter in a variety of ecosystems. Although moisture dependent, they are among the macroarthropods of desert ecosystems (Crawford, 1979). Millipedes, which are important in Ca cycling in forests, have a calcareous exoskeleton and may process 15%–20% of Ca inputs into forest floors (Coleman et al., 2004). Some millipedes are obligate coprophages, feeding on their own microbially enriched fecal matter, while others excrete noxious chemicals as a defense mechanism against predation.

Centipedes are ubiquitous and active predators in soil and litter habitats, able to run rapidly and capture small prey such as microarthropods (especially collembolans). Most are 3–5 cm in length, but tropical centipedes may exceed 30 cm. Typically, centipedes constitute about 20% of the predaceous macroarthropods in temperate forests, but the percentage is lower in subarctic, boreal, and dry forests (Albert, 1979). Species diversity is lower for centipedes than for other predators such as spiders and staphylinid beetles.

25.4.2.3 Spiders

Spiders are the most numerous of the predaceous macroarthropods in ecosystems ranging from forest to grassland to agroecosystems. The taxonomy of soil- and litter-dwelling spiders remains unsettled, especially for the numerous species in the family Linyphiidae, which contains many small soil species. Spiders, which are strictly carnivorous, are generalist feeders that attack insects but also feed upon other invertebrates, including other spiders (Wise, 1993). Despite their numbers, there is no consensus on the abilities of spiders to control insect populations (Riechert and Lockly, 1984). Spiders do not reproduce rapidly enough to keep pace with exploding prey populations. Also, many species are territorial. In forest habitats, spiders may act as a stabilizing influence on populations of forest floor invertebrates by maintaining a continual predation pressure.

25.4.2.4 Beetles

In terms of numbers of species, beetles are the largest order of insects, being found in every habitat except the oceans (Richter, 1958; Thiele, 1977). In soil systems, beetles include species that are phytophagous, saprophagous, and predaceous. Carabid, tenebrionid, and staphylinid beetles are numerous predators in disturbed and undisturbed systems alike. Together with spiders, carabids are the typical ground-surface macroarthropod predators taken in pitfall traps in agroecosystems. Other active predators included the tiger beetles (Cicindelidae) whose larvae construct belowground retreats from which they capture prey. Phytophagous beetles include the Scarabaeidae (June beetles) whose larvae feed extensively on roots. Larvae of elaterid beetles (wireworms) are important root feeders in cropping systems and in forests. Although predaceous beetles

are conspicuous, especially on the soil surface in agricultural fields, the phytophagous species are probably more important. The predaceous carabids may exert some control on caterpillars such as armyworms and similar species. Gypsy moth caterpillars, descending to the soil of the forest to pupate, may fall prey to carabid beetles in large numbers.

25.4.3 Sampling and Analysis

Macroarthropods are sampled in several ways. A good review of the general methodology used for sampling populations of soil macroinvertebrates is given in Edwards (1991). The most basic method involves delineation of an area a square meter or less followed by hand sorting organic horizons and digging to a set depth the mineral horizons of soil and collecting all macroarthropods encountered. Tullgren funnels are useful for extracting arthropods from bulk samples of soil and litter, and this technique relies on heat and light stimuli to drive organisms from the samples (Section 25.4.2.3 and Figure 25.7). It is also possible to collect subterranean macroarthropods by taking soil cores (10–15 cm diameter) and sieving them, either dry or with the use of a wet sieving apparatus. Another technique used to sample macroarthropods involves the separation of organisms from samples by taking advantage of their relatively low specific gravity and the hydrophobic properties of their cuticles. This generally involves some kind of flotation of organisms from samples in a high density liquid (usually salt or sugar solutions; Edwards, 1991).

When investigators are strictly interested in sampling the greatest number of species (i.e., sampling for diversity and not abundance or density), other sampling techniques are useful. For example, Snyder et al. (2006) found that targeted hand collecting of millipedes from sites selected a priori as being likely for encountering specimens (rather than randomly assigned spatially based approaches) produced the largest number of species in the shortest time. Another qualitative method in widespread use involves pitfall traps—cans set flush with the soil surface
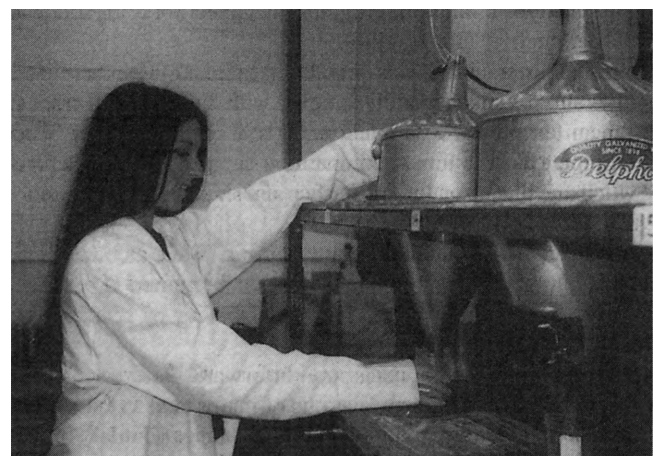
FIGURE 25.7   Large Tulgreen funnels for extraction of macroarthropods from soil cores and litter.
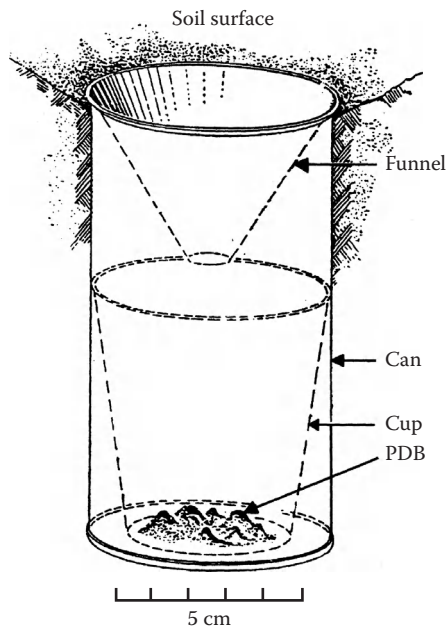
FIGURE 25.8 A pitfall trap for surface active macroarthropods. PDB = P-dichlorobenze preservative. (Reprinted from Reichle, D.E., and D.A. Crossley, Jr. 1965. Radiocesium dispersion in a cryptozoan food web. Health Phys. 11:1375–1384. With permission from Williams and Wilkins, Baltimore, MD.)

and containing a preservative (Figure 25.8). Pitfall traps are an inexpensive sampling technique, but are not entirely quantitative, since captures depend upon the mobility of animals as well as their density. In sampling programs designed for comparison of areas or seasons, pitfall traps are a preferred method (Bater, 1996). Each of these qualitative sampling methods can be rendered semiquantitative if care is taken to standardize the time that searching or trapping is conducted and comparisons may be made on the basis of equal effort collecting for a given habitat type or experimental manipulation.

Because ants and termites are social insects, population estimation for these insects requires special techniques. Species diversity of ants is readily sampled with baited pitfall traps (Romero and Jaffe, 1989), while subterranean termites are sampled using soil cores (Lee and Wood, 1971). Termite mounds may require complete destruction. A comparative sampling technique for desert termites uses rolls of toilet paper, which are placed on the soil surface and shielded from the sun with aluminum foil (Whitford et al., 1982).

Samples may be preserved in 70% alcohol prior to sorting. Although alcoholic storage is satisfactory for storage of specimens, most entomologists prefer that insects be pinned if possible. Different insect orders are pinned in different ways, and detailed instructions for pinning and labeling techniques are given in most entomology textbooks (Borror et al., 1992).

Sorting of adult macroarthropods into major taxa is straightforward. Reference to a general entomology textbook will allow the novice to make identifications to major hexapod taxa to family levels. More detailed sorting of other macroarthropod taxa such as spiders, diplopods, and chilopods may be done with the aid of literature guides (Dindal, 1990; Ubick et al., 2005). Additionally, because samples derived from subterranean sampling will usually include immature stages of Coleoptera, Lepidoptera, and Diptera (among others), special training is usually required for identifications, although a few groups are readily recognized. North American guides for the identification of immature insects including those found predominantly in soils are found in Peterson (1967) and Stehr (2005). Identifications of adult specimens to generic and species levels will also require services of a specialist in the taxonomy of the group. There is some promise in the development of molecular-based techniques for the identification of soil organisms, but to date these have been useful only for establishing the degree of genetic diversity in a sample, and not so useful for providing clues as to the identities of organisms (Wu et al., 2009). A DNA-based technique has been used to identify the gut contents of predatory beetles (Juen and Traugott, 2006), and in the future may prove useful for identification of bulk extracted organisms as well.

Once organisms have been collected and reliably identified, the data typically are reported in terms of presence/absence, relative abundance, or frequency of occurrence (e.g., from pitfall trapped collections), or when sampling is conducted in a spatially explicit way, in terms of density (individuals m^{-2}) or biomass (g ash-free dry mass m^{-2}). It is common for researchers to use abundance and identity data to calculate community measures such as indexes of diversity, evenness, similarity, or rank abundance. This kind of analysis is typically performed when sampling has been conducted in areas that have experienced different land uses or been subjected to experimental treatments that are expected to have influence on macroinvertebrate community structure (e.g., Decaëns et al., 2004; Callaham et al., 2006). Another useful and increasingly common technique for analyzing soil macroinvertebrate community structure is the use of multivariate statistical analyses to describe the dimensionality and variation in assemblages from different experimental treatments or habitats (e.g., Decaëns et al., 1998; Siira-Pietikäinen et al., 2003).

Because the net effect of macroarthropods on ecosystem functions such as decomposition or primary productivity are often dependent in some part on indirect effects through trophic interactions, ecologists are interested in food web structure and energy and nutrient flows through these webs. However, for subterranean linkages in food webs, observations have been difficult to make without fairly intensive disturbances to the system under scrutiny. Relatively recent developments in mass spectroscopy have improved this situation, and made it possible to process a large number of samples at reasonable cost for stable isotopic signatures of invertebrate tissues. This has greatly facilitated the determination of dietary relationships between individual taxa of belowground consumers (e.g., Callaham et al., 2000; Traugott et al., 2008; Seeber et al., 2009) as well as trophic relationships in soil food webs with better resolution than ever before (Halaj et al., 2005; Elfstrand et al., 2008), and this technology provides a fruitful avenue for future research on these organisms.

Finally, further recent advances in the soil food web and trophic interaction arena are due to molecular techniques that can achieve species-specific resolution for the analysis of gut contents (particularly useful in the analysis of gut contents of predators). These molecular approaches are of two basic types—those that utilize monoclonal antibodies and enzyme-linked immunosorbent assays (ELISA) to detect species (or group)-specific proteins from environmental samples (e.g., McKemey et al., 2006; Thomas et al., 2009), and those that utilize DNA-based assays where primers for suspected prey items are used to amplify the DNA of gut contents of predatory macroarthropods (Juen and Traugott, 2007; Kuusk et al., 2008). A good review of the methods for DNA-based approaches to assessing predator–prey interactions, along with a detailed discussion of challenges, is given in King et al. (2008). Both the DNA and ELISA techniques allow a level of resolution that is unprecedented in determining the diets of predators in soil food webs, and the ELISA technique has even been used to estimate densities of soil invertebrates (slugs) with reasonable accuracy at with greater speed than conventional estimation methods (McKemey et al., 2006). Fournier et al. (2008) conducted a comparative study to evaluate the relative sensitivities of ELISA and PCR techniques for detecting a specific prey item in the diets of 30 predator taxa, and found that after initial outlay of time and money for development of the antibody used in the assay, the ELISA technique was faster, more sensitive, and less expensive to process samples. Although these approaches promise major advances in future understanding of soil food webs, and have been used to good effect in aboveground contexts, they are in their infancy as applied to soil systems, and many of the methodological details are still under development (see King et al., 2008; von Berg et al., 2008). Thus, a great deal of work remains to be done in terms of methods development (e.g., production of appropriate primers and antibodies for detection of the full diversity of available food items in soil food webs), and this work must be accomplished before the potential of molecular techniques can be completely realized in soil ecological studies. Nevertheless, with continued growth in the number of investigators using these techniques, and with continued advances in molecular methodology, it is likely that the immense diversity and complexity of the soil biota will be understood with greater detail than had previously been possible.

References

Albert, A.M. 1979. Chilopoda as part of the predatory macroarthropod fauna in forests: Abundance, life-cycle, biomass and metabolism, p. 215–231. *In* M. Camatini (ed.) Myriapod biology. Academic Press, London, U.K.

Arnett, R.H., Jr. 1993. American insects. A handbook of the insects of America north of Mexico. Sandhill Crane Press, Budapest, Hungary.

Arriaga, A., and Y. Maya. 2007. Spatial variability in decomposition rates in a desert scrub of northwestern Mexico. Plant Ecol. 189:213–225.

Aubert, M., M. Hedde, T. Decaëns, F. Bureau, P. Margerie, and D. Alard. 2002. Effects of tree canopy composition on earthworms and other macro-invertebrates in beech forests of upper Normandy (France). Pedobiologia 47:904–912.

Bater, J. 1996. Micro- and macro-arthropods, p. 163–174. *In* G.S. Hall (ed.) Methods for the examination of organismal diversity in soils and sediments. CABI, New York.

Bengtsson, J., T. Persson, and H. Lundkvist. 1997. Long-term effects of logging residue addition and removal on macro-arthropods and enchytraeids. J. Appl. Ecol. 34:1014–1022.

Blair, J.M., T.C. Todd, and M.A. Callaham, Jr. 2000. Responses of grassland soil invertebrates to natural and anthropogenic disturbances, p. 43–71. *In* D.C. Coleman and P.F. Hendrix (eds.) Invertebrates as webmasters in ecosystems. CABI, New York.

Blossey, B., and T.R. Hunt-Joshi. 2003. Belowground herbivory by insects: Influence on plants and aboveground herbivores. Annu. Rev. Entomol. 48:521–547.

Blumberg, A.Y., and D.A. Crossley, Jr. 1983. Comparison of soil surface arthropod populations in conventional tillage, no-tillage and old field systems. Agro-Ecosystems 8:247–253.

Bolton, B. 1994. Identification guide to the ant genera of the world. Harvard University Press, Cambridge, MA.

Borror, D.J., C.A. Triplehorn, and N.F. Johnson. 1992. An introduction to the study of insects. Saunders College Publishers, Fort Worth, TX.

Broza, M., and I. Izhaki. 1997. Post-fire arthropod assemblages in Mediterranean forest soils in Israel. Int. J. Wildland Fire 7:317–325.

Bryan, M.V. 1978. Production ecology of ants and termites. Cambridge University Press, Cambridge, U.K.

Callaham, M.A., Jr., J.M. Blair, T.C. Todd, D.J. Kitchen, and M.R. Whiles. 2003. Macroinvertebrates in North American tallgrass prairie soils: Effects of fire, mowing, and fertilization on density and biomass. Soil Biol. Biochem. 35:1079–1093.

Callaham, M.A., Jr., D.D. Richter, D.C. Coleman, and M. Hofmockel. 2006. Long-term land-use effects on soil invertebrate communities in Southern Piedmont soils, USA. Eur. J. Soil Biol. 42:S150–S156.

Callaham, M.A., Jr., M.R. Whiles, C.K. Meyer, B.L. Brock, and R.E. Charlton. 2000. Feeding ecology and emergence production of annual cicadas (Homoptera: Cicadidae) in tallgrass prairie. Oecologia 123:535–542.

Camatini, M. (ed.). 1979. Myriapod biology. Academic Press, London, U.K.

Coleman, D.C., D.A. Crossley, Jr., and P.F. Hendrix. 2004. Fundamentals of soil ecology. Academic Press, San Diego, CA.

Crawford, C.S. 1979. Desert millipedes: A rationale for their distribution, p. 171–181. *In* M. Camatini (ed.) Myriapod biology. Academic Press, London, U.K.

Crawford, C.S. 1981. Biology of desert invertebrates. Springer-Verlag, Berlin, Germany.

Decaëns, T., T. Dutoit, D. Alard, and P. Lavelle. 1998. Factors influencing soil macrofaunal communities in post-pastoral successions of western France. Appl. Soil Ecol. 9:361–367.

Decaëns, T., J.J. Jiménez, E. Barros, A. Chauvel, E. Blanchart, C. Fragoso, and P. Lavelle. 2004. Soil macrofaunal communities in permanent pastures derived from tropical forest or savanna. Agric. Ecosyst. Environ. 103:301–312.

Dindal, D.L. (ed.). 1990. Soil biology guide. Wiley Interscience, New York.

Doğramaci, M., and W.M. Tingey. 2009. Host resistance and influence of tuber surface on larval performance of potato tuberworm (Lepidoptera: Gelechiidae). J. Pest Sci. 82:109–114.

Draney, M.L. 1992. Biodiversity of ground-layer macroarachnid (Araneae and Opiliones) assemblages from four Piedmont floodplain habitats. University of Georgia, Athens, GA.

Edwards, C.A. 1991. The assessment of populations of soil-inhabiting invertebrates. Agric. Ecosyst. Environ. 34:145–176.

Ekschmitt, K., V. Wolters, and M. Weber. 1997. Spiders, carabids, and staphylinids: The ecological potential of predatory microarthropods, p. 307–362. *In* G. Benckiser (ed.) Fauna in soil ecosystems. Marcel Dekker, New York.

Elfstrand, S., J. Lagerlöf, K. Hedlund, and A. Mårtensson. 2008. Carbon routes from decomposing plant residues and living roots into soil food web assessed with ^{13}C labelling. Soil Biol. Biochem. 40:2530–2539.

Foelix, R.F. 1996. Biology of spiders. Oxford University Press, New York.

Fournier, V., J. Hagler, K. Daane, J. de Leon, and R. Groves. 2008. Identifying the predator complex of *Homalodisca vitripennis* (Hemiptera: *Cicadellidae*): A comparative study of the efficacy of and ELISA and PCR gut content assay. Oecologia 157:629–640.

Frouz, J. and A. Ali. 2004. Soil macroinvertebrates along a successional gradient in central Florida. Fla. Entomol. 87: 386–390.

Halaj, J., R.W. Peck, and C.G. Niwa. 2005. Trophic structure of a macroarthropod litter food web in managed coniferous forest stands: A stable isotope analysis with δ^{15}N and δ^{13}C. Pedobiologia 49:109–118.

Hölldobler, B., and E.O. Wilson. 1990. The ants. Harvard University Press, Cambridge, MA.

House, G.J., and B.R. Stinner. 1983. Arthropods in no-tillage soybean agroecosystems: Community composition and ecosystem interactions. Environ. Manage. 7:23–28.

Huhta, V. 2002. Soil macroarthropod communities in planted birch stands in comparison with natural forests in central Finland. Appl. Soil Ecol. 20:199–209.

Hunter, M.D. 2001. Out of sight, out of mind: The impacts of root-feeding insects in natural and managed systems. Agric. For. Entomol. 3:3–9.

Jackson, T.A., and M.G. Klein. 2006. Scarabs as pests: A continuing problem. Coleopt. Bull. 60:102–119.

Johnson, S.N., and P.J. Murray (eds.). 2008. Root feeders: An ecosystem perspective. CABI, Wallingford, U.K.

Jones, C.G., J.H. Lawton, and M. Shachak. 1997. Positive and negative effects of organisms as physical ecosystem engineers. Ecology 78:1946–1957.

Jouquet, P., J. Dauber, J. Lagerlöf, P. Lavelle, and M. Lepage. 2006. Soil invertebrates as ecosystem engineers: Intended and accidental effects on soil and feedback loops. Appl. Soil Ecol. 32:153–164.

Juen, A., and M. Traugott. 2006. Amplification facilitators and multiplex PCR: Tools to overcome PCR-inhibition in DNA-gut-content analysis of soil-living invertebrates. Soil Biol. Biochem. 38:1872–1879.

Juen, A., and M. Traugott. 2007. Revealing species specific trophic links in soil food webs: Molecular identification of scarab predators. Mol. Ecol. 16:1545–1557.

Kastan, B.J. 1978. How to know the spiders. William C. Brown, Dubuque, IA.

Keller, L., and E. Gordon. 2009. The lives of ants. Oxford University Press, New York.

Kevan, D.K. McE., and G.G.E. Scudder. 1989. Illustrated keys to the families of terrestrial arthropods of Canada. 1. Myriapods (Millipedes, Centipedes, etc.). Biological Survey of Canada Taxonomic Series 1. Biological survey of Canada, Ottawa, Canada.

King, R.A., D.S. Read, M. Traugott, and W.O.C. Symondson. 2008. Molecular analysis of predation: A review of best practice for DNA-based approaches. Mol. Ecol. 17:947–963.

Kuperman, R. 1996. Relationships between soil properties and community structure of soil macroinvertebrates in oak-hickory forests along an acidic deposition gradient. Appl. Soil Ecol. 4:125–137.

Kuusk, A.-K., A. Cassel-Lundhagen, A. Kvarnheden, and B. Ekbom. 2008. Tracking aphid predation by lycosid spiders in spring-sown cereals using PCR-based gut-content analysis. Basic Appl. Ecol. 9:718–725.

Lawrence, K.L., and D.H. Wise. 2004. Unexpected indirect effect of spiders on the rate of litter disappearance in a deciduous forest. Pedobiologia 48:149–157.

Lee, K.E., and T.G. Wood. 1971. Termites and soils. Academic Press, London, U.K.

Linden, D.R., P.F. Hendrix, D.C. Coleman, and P.C.J. van Vliet. 1994. Faunal indicators of soil quality, p. 91–106. *In* J.W. Doran, D.C. Coleman, D.F. Bezdicek, and B.A. Stewart (eds.) Defining soil quality for a sustainable environment. SSSA Special Publication No. 35. SSSA, Madison, WI.

Lobry de Bruyn, L.A. 1997. The status of soil macrofauna as indicators of soil health to monitor the sustainability of Australian agricultural soils. Ecol. Econ. 23:167–178.

McCracken, D.I., G.N. Foster, and A. Kelly. 1995. Factors affecting the size of leatherjacket (Diptera: Tipulidae) populations in pastures in the west of Scotland. Appl. Soil Ecol. 2:203–213.

McKemey, A.R., D.M. Glen, C.W. Wiltshire, and W.O.C. Symondson. 2006. Molecular quantification of slug density in the soil using monoclonal antibodies. Soil Biol. Biochem. 38:2903–2909.

Paquin, P., and D. Coderre. 1997. Changes in soil macroarthropod communities in relation to forest maturation through three successional stages in the Canadian boreal forest. Oecologia 112:104–111.

Pellens, R., and I. Garay. 1999. Edaphic macroarthropod communities in fast-growing plantations of *Eucalyptus grandis* Hill ex Maid (Myrtaceae) and *Acacia mangium* Wild (Leguminosae) in Brazil. Eur. J. Soil Biol. 35:77–89.

Peterson, A. 1967. Larvae of insects: An introduction to Nearctic species. 6th edn. Edwards and Bothers, Ann Arbor, MI.

Reichert, S.E., and T. Lockly. 1984. Spiders as biological control agents. Annu. Rev. Entomol. 29:299–320.

Reichle, D.E., and D.A. Crossley, Jr. 1965. Radiocesium dispersion in a cryptozoan food web. Health Phys. 11:1375–1384.

Richter, P.O. 1958. Biology of Scarabaeidae. Annu. Rev. Entomol. 3:25–58.

Romero, H., and K. Jaffe. 1989. A comparison of methods for sampling ants (Hymenoptera, Formicidae) in savannas. Biotropica 21:348–352.

Salmon, S., J. Mantel, L. Frizzera, and A. Zanella. 2006. Changes in humus forms and soil animal communities in two developmental phases of Norway spruce on an acidic substrate. For. Ecol. Manage. 237:47–56.

Seeber, J., R. Langel, E. Meyer, and M. Traugott. 2009. Dwarf shrub litter as a food source for macro-decomposers in alpine pastureland. Appl. Soil Ecol. 41:178–184.

Shachak, M.E., A. Chapman, and Y. Steinberger. 1976. Feeding, energy flow and soil turnover in the desert isopod, *Hemilepistus reamuri*. Oecologia 24:57–69.

Siira-Pietikäinen, A., J. Haimi, and J. Siitonen. 2003. Short-term responses of soil macroarthropod community to clear felling and alternative forest regeneration methods. For. Ecol. Manage. 172:339–353.

Snider, R.M., and J.H. Shaddy. 1980. The ecobiology of *Trachelipus rathkei* (Isopoda). Pedobiologia 20:394–410.

Snyder, B.A., M.L. Draney, and P. Sierwald. 2006. Development of an optimal sampling protocol for millipedes (Diplopoda). J. Insect Conserv. 10:277–288.

Stehr, F.W. 2005. Immature insects. 2nd edn. (two volumes). Kendall/Hunt Publishing, Dubuque, IA.

Stork, N.E., and P. Eggleton. 1992. Invertebrates as determinants of soil quality. Am. J. Altern. Agric. 7:38–47.

Terborgh, J., L. Lopez, P. Nuñez V., M. Rao, G. Shahabuddin, G. Orihuela, M. Riveros, R. Ascanio, G.H. Adler, T.D. Lambert, and L. Balbas. 2001. Ecological meltdown in predator-free forest fragments. Science 294:1923–1926.

Thiele, H.U. 1977. Carabid beetles in their environments. Springer-Verlag, Berlin, Germany.

Thomas, R.S., J.D. Harwood, D.M. Glen, and W.O.C. Symondson. 2009. Tracking predator density dependence and subterranean predation by carabid larvae on slugs using monoclonal antibodies. Ecol. Entomol. 34:569–579.

Traugott, M., N. Schallhart, R. Kaufmann, and A. Juen. 2008. The feeding ecology of elaterid larvae in central European arable land: New perspectives based on naturally occurring stable isotopes. Soil Biol. Biochem. 40:342–349.

Ubick, D., P. Paquin, P.E. Cushing, and V. Roth (eds.) 2005. Spiders of North America: An identification manual. American Arachnological Society, Washington, DC.

von Berg, K., M. Traugott, W.O.C. Symondson, and S. Scheu. 2008. The effects of temperature on detection of prey DNA in two species of carabid beetle. Bull. Entomol. Res. 98:263–269.

Welch, S.M., J.L. Waldron, A.G. Eversole, and J.C. Simoes. 2008. Seasonal variation and ecological effects of Camp Shelby burrowing crayfish (*Fallicambarus gordoni*) burrows. Am. Midl. Nat. 159:378–384.

Whiles, M.R., M.A. Callaham, Jr., C.E. Meyer, B.L. Brock, and R.E. Charlton. 2001. Emergence of periodical cicadas (*Magicicada cassini*) from a Kansas riparian forest: Densities, biomass and nitrogen flux. Am. Midl. Nat. 145:176–187.

Whitford, W.G., Y. Steinberger, and G. Ettershank. 1982. Contributions of subterranean termites to the "economy" of Chihuahuan desert ecosystems. Oecologia 55:298–302.

Williams, S.C. 1987. Scorpion bionomics. Annu. Rev. Entomol. 32:275–295.

Wise, D.H. 1993. Spiders in ecological webs. Cambridge University Press, Cambridge, U.K.

Wise, D.H. 2004. Wandering spiders limit densities of a major microbe-detritivore in the forest-floor food web. Pedobiologia 48:181–188.

Wolters, V., and K. Ekschmitt. 1997. Gastropods, isopods, diplopods and chilopods: Neglected groups of the decomposer food web, p. 265–306. *In* G. Benckiser (ed.) Fauna in soil ecosystems. Marcel Dekker, New York.

Wu, T., E. Ayers, G. Li, R.D. Bardgett, D.H. Wall, and J.R. Garey. 2009. Molecular profiling of soil animal diversity in natural ecosystems: Incongruence of molecular and morphological results. Soil Biol. Biochem. 41:849–857.

25.5 Enchytraeidae—Oligochaeta

P.C.J. van Vliet

Paul F. Hendrix

25.5.1 Introduction

The Enchytraeidae is a family of Oligochaetes that occur in terrestrial, littoral, and aquatic habitats. A total of some 600 species are now known worldwide, from the tropics to polar latitudes (Dash, 1990; Vaculik et al., 2004; Christensen and Dozsa-Farkas, 2006). Enchytraeids are mostly pale-colored and are anatomically similar to earthworms (only smaller). Their length ranges from 5 cm to less than 1 mm. Larger enchytraeids (up to 60 mm long) have been found in subarctic soils from the unglaciated portion of the northern Yukon (Canada) (Smith et al., 1990).

Of the Oligochaeta, earthworms (Lumbricidae) have been the subject of most studies. Enchytraeidae (also known as "potworms") have been studied less frequently although they are distributed throughout the world. The biology and ecology of enchytraeids have become somewhat better known during the

last 30 years, especially in Europe. This intensification was partly due to the systematic revision of this family for European species (Nielsen and Christensen, 1959, 1961, 1963), the development of efficient methods of extraction of enchytraeids from soil (Nielsen, 1953; O'Connor, 1955), and the fact that they were very abundant in certain ecosystems and were expected to have an important role in nutrient cycling and soil structure formation.

In the past two decades, enchytraeids have become widely used as test organisms for ecotoxicological evaluation of soil contaminants (Didden and Römbke, 2001; Römbke, 2003), and recent studies suggest that they may be sensitive indicators of effects of climate change (Briones et al., 2007; Maraldo et al., 2008).

25.5.2 Taxonomy and Morphology

Currently the Nielsen and Christensen (1959, 1961, 1963) monographs for European species are used for the identification of enchytraeids, although many more genera and species have been found since. In 2003, Schmelz (2003) described the taxonomy of the dominant *Fridericia* genus. Schmelz has also submitted a new key to European terrestrial and freshwater species of enchytraeids (Schmelz and Collado, 2010). A simple key to the most common genera of Enchytraeidae can be found in the *Soil Biology Guide* (Dash, 1990). Taxonomic literature on North American Enchytraeidae is fairly sparse and no good key is available.

The easiest way to observe and identify enchytraeids is by getting them to clear their gut in a Petri dish in a cool place. A sexually mature worm (alive) can then be transferred to a drop of water on a slide and covered with a cover slip. Most of the structures, shown in Figure 25.9, are fairly transparent in live worms and can be examined at a 100–400× magnification.

The known genera of enchytraeids are mentioned in Table 25.10. Two-thirds of the genera can be found in soil. Information (when available) about the habitats of the other genera is mentioned as well.

25.5.3 Collection and Extraction Methods

For quantitative sampling purposes, enchytraeids can best be sampled with a cylindrical soil corer, which will keep the soil intact. Different studies have shown that the optimum sampler size is 5–7.5 cm in diameter (Lal et al., 1981; Didden et al., 1995). Because enchytraeids are clustered in soil (see later), sufficient replicates need to be taken, to get a good estimate of population size and composition. However, the size and number of the sampling units are mostly chosen as a compromise between accuracy of the abundance estimates and the amount of work involved (Didden, 1993).

Hand sorting of enchytraeids out of soil is possible, but the efficiency of this labor-intensive method is very low. Enchytraeids are often covered with small sand grains and organic particles (Ponge, 1984), making them hard to distinguish from soil material. Small animals are also hard to recover with the hand sorting method.
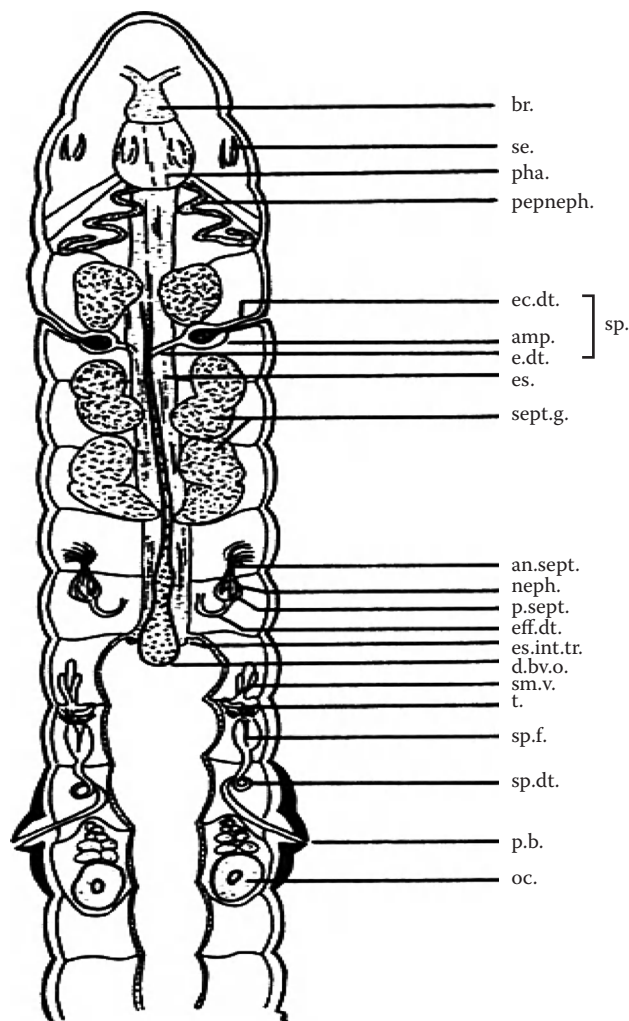
FIGURE 25.9 Morphological characters of an enchytraeid worm. amp., ampulla; an.sept., ante-septal; br., brain; d.bv.o., dorsal blood vessel origin; ec.dt., ectal duct; e.dt. ental duct; eff.dt., efferent duct; es., esophagus; es.int.tr., esophageal intestinal transition; neph., nephridia; oc., oocyte; pha., pharynx; p.b., penial bulb; pepneph., peptonephridia; p.sept., postseptal; se., setae; sept.g., septal gland; sm.v., seminal vesicle; sp. spermatheca; sp.dt., sperm duct; sp.f., sperm funnel; t., testes. (Reprinted from Dindal, D.L. (ed.), *Soil Biology Guide*, Wiley Interscience, New York, 1990.)

The most common method used to extract enchytraeids from soil is the O'Connor wet funnel method (O'Connor, 1955). In this method, a soil slice (maximum 3 cm thick), resting on a sieve in a funnel filled with water, is exposed to light and heat. In 3 h, the light intensity is increased gradually until the soil surface has reached a temperature of 45°C. Enchytraeids respond to the light and heat by moving away from the source and pass through the sieve into the water below. Figure 25.10 shows a modified O'Connor extractor apparatus. It consists of a box containing 32 funnels above which 25 W bulbs are generating enough heat in 3 h to get the soil surface to the correct temperature.

Recently, a modified extraction method has been proposed (Graefe, 1973; Schauermann, 1983). In this method

TABLE 25.10 Enchytraeid Genera and Their Occurrence in Soil

Genera Occurring in Soil	Other Genera	Environment
Achaeta	*Aspidodrilus*	Epizoic on earthworms
Bryodrilus	*Barbidrilus*	Freshwater
Buchholzia	*Enchylea*	Only found in enchytraeid culture
Cernosvitoviella	*Enchytraeina*	Marine
Cognettia	*Grania*	Marine
Enchytraeus	*Pelmatodrilus*	Epizoic on earthworms
Enchytronia	*Propappus*	Freshwater
Fridericia	*Randidrilus*	Marine
Guaranidrilus	*Stephensoniella*	Marine
Hemienchytraeus		
Hemifridericia		
Henlea		
Isosetosa		
Lumbricillus		
Marionina		
Mesenchytraeus		
Oconnorella		
Stercutus		
Tupidrilus		

Source: Nielsen, C.O., and B. Christensen. 1959. The Enchytraeidae: Critical revision and taxonomy of European species. Natura Jutl. 8–9:1–160; Römbke, J. 1992. New taxa since 1985. *In* J. Römbke (ed.) Newsletter on Enchytraeidae No. 3. Institut für Angewandte Bodenbiologie, Hamburg, Germany; Römbke, J., and K. Dósza-Farkas. 1996. New taxa since 1992, p. 5–9. *In* K. Dósza-Farkas (ed.) Newsletter on Enchytraeidae No. 5. Eötvös Loránd University, Budapest, Hungary; Dósza-Farkas, personal communication.
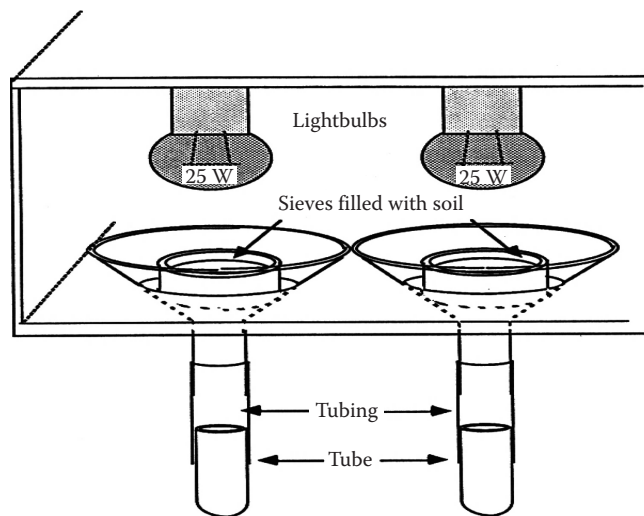
FIGURE 25.10 Enchytraeid extractor after O'Connor (1955). The sieves with soil are completely submerged in water; after 4 h in the dark the light intensity is increased in 3 h from very dim to full bright.
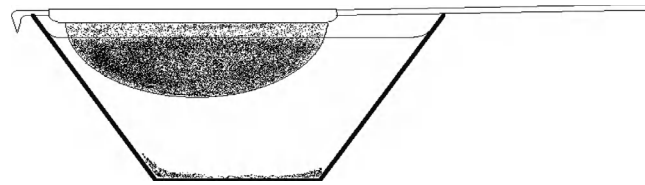
FIGURE 25.11 Part of an enchytraeid extractor after consisting of a sieve (1 mm mesh) held in a bowl filled with water, developed by Graefe (1973).

(Figure 25.11), enchytraeids are extracted from soil without heat. The total extraction time is extended from 3 h to several days for soils rich in organic matter and up to 2 weeks for mineral soils. The length of the extraction period is determined by the risk of an oxygen deficit and of dying of larger individuals.

A comparative study of the two methods (Didden et al., 1995) has shown that a better extraction is obtained without heating. The length of the cold extraction period and the total extraction time had a significant positive influence on the extraction efficiency. Also, the Graefe method is cheap and easy to set up. However, due to its long extraction time, the method is not very appropriate for studies where large numbers of samples are taken or where samples are taken at short time intervals. In those cases, the faster O'Connor method (maybe modified with a longer settling time before the heated extraction starts) might be more appropriate (Didden et al., 1995).

25.5.4 Life History: Reproduction

Several types of reproduction strategies are known in enchytraeids (Table 25.11). Reproduction occurs mostly through the formation of cocoons. A cocoon contains one or more eggs, depending upon the individual and the species. Some enchytraeids cover their cocoons with a layer of organic debris and

TABLE 25.11 Reproduction Strategies in Enchytraeids

Reproduction	Description	Genera/Species
Only sexually, all during the year possible	Irregular reproduction periods, specimens with eggs can be found any time of the year	Many *Fridericia* sp., most *Enchytraeus* sp., *Henlea* sp., *Enchytronia* sp., *Marionina* sp., *Achaeta* sp.
Only sexually, become mature once a year	Thereafter, they regress and stay in rejuvenile state till the next reproduction period	*Mesenchytraeus* sp., *S. niveus*
Parthenogenesis	Egg development without fertilization	*Lumbricillus* sp. and *Fridericia* sp.
Fragmentation		*Buchholzia* sp. and *Cognettia* sp.
Self-fertilization		*E. buchholzi, E. bulbosis*

Source: After Dósza-Farkas, K. 1996. Reproduction strategies in some enchytraeid species, p. 25–33. *In* K. Dósza-Farkas (ed.) Newsletter on Enchytraeidae No. 5. Eötvös Loránd University, Budapest, Hungary.

sand grains, which makes them hard to find by predators in the soil (Christensen, 1956). Covering the cocoons might also increase drought resistance of the cocoons (Dash, 1990). Cocoon production is more rapid in young mature than older worms. Depending on the species and the temperature, the duration from cocoon hatching to maturity lasts about 65–120 days. The maximum lifespan of enchytraeids is assumed to be 1 year (Dash, 1990).

Parthenogenesis, self-fertilization, and fragmentation are reproduction strategies often used by cosmopolite and widespread species. These strategies allow animals to survive under particularly unfavorable conditions and enhance their chances to successfully colonize new habitats (Dósza-Farkas, 1996).

25.5.5 Food Preferences

It is known that enchytraeids ingest mineral particles, leaf litter, fungi, bacteria, oats and yeast, seaweeds, and sewage. Dash et al. (1981) and Urbášek and Chalupský (1991) found that enchytraeids were capable of digesting di- and some polysaccharides in food, but that most complex plant substances would become available through microbial decomposition. Related to their food source, different species have different amounts of certain enzymes. For instance, litter-inhabiting species such as *Fridericia galba* and *Cognettia sphagnetorum* manifested higher activities of enzymes to break down polysaccharides, while *Enchytraeus albidus*, which prefers mostly compost, had higher proteolytic activity (Urbášek and Chalupský, 1991). In food preference experiments, Dósza-Farkas (1982) showed that enchytraeids preferred old and microbially decomposed leaves over fresh leaves. However, consumption of the partially decomposed leaves differed by species. According to Toutain et al. (1982), enchytraeids only consumed the epidermis and parenchyma cells of leaves, while avoiding the veins. After passage through their guts, the structure of the material had not changed much; only bacteria and fungi earlier present on the leaves were digested. Brockmeyer et al. (1990) used radiolabeled food sources to determine food preferences of two enchytraeid species and found that enchytraeids were not able to break down the cell walls of the fresh plant material, while the muramic acid cell walls of the microorganisms were no barrier to the worms' digestive capabilities. Using ^{14}C dating techniques, Briones and Ineson (2002) found that enchytraeids mostly assimilated soil carbon that was 5–10 years old regardless of the depth at which the enchytraeids were collected.

Looking at the feeding habits of *C. sphagnetorum*, Persson et al. (1980) considered enchytraeid communities to be 50% saprovorous, 25% fungivorous, and 25% bacterivorous. This differentiation is often used in production ecological studies. It seems however that microbivory is underestimated, because some species are well adapted to digest microbes, while no species have been found that produce sufficient amounts of enzymes to decompose complex plant compounds. Therefore, it seems to be more realistic to assume that an enchytraeid community contains 80% microbivorous and 20% saprovorous feeders (Didden, 1993).

25.5.6 Enchytraeidae Abundances and Distributions

Enchytraeidae are common in cultivated soil but reach their highest abundances in acid soils with a high organic matter content (e.g., O'Connor, 1957; Peachey, 1963). Large populations have been found in cold and temperate moist habitats, but this may be more a function of the greater amount of data available from Europe than of environmental conditions. Table 25.12 gives an overview of average annual abundances of enchytraeids in various habitats throughout the world. Didden (1991) found no correlations between enchytraeid densities on one hand and annual precipitation, annual temperature, and acidity of the soil on the other.

It is well documented that enchytraeids occur in aggregated distribution patterns (Nielsen, 1954; Peachey, 1963; O'Connor, 1967; Abrahamsen, 1969; Abrahamsen and Strand, 1970; Nakamura, 1979, 1982). Chalupský and Lepš (1985) demonstrated that enchytraeids occur in multispecies aggregation centers. It seems therefore that aggregations are not due to sudden breeding from cocoons but are environmentally conditioned (e.g., physical conditions, food availability).

The vertical distribution of enchytraeids in a soil is mostly related to the organic material distribution in the soil profile. In most natural habitats, up to 90% of the population occurs in the upper layers where most of the organic matter is located. Persson et al. (1980) found 57% of the enchytraeid biomass in the organic layers in a Scots pine forest. O'Connor (1957) found that more than 50% of the enchytraeids occurred in the humus layer of a Douglas fir coniferous forest.

The importance of organic matter for the vertical distribution of enchytraeids can easily be demonstrated in agricultural ecosystems. In pastures and no-tillage field, where the organic matter is not plowed in, enchytraeids occur mostly in the top layer. On the other hand, in tilled agricultural fields, where the organic matter is mixed throughout the soil profile, enchytraeids are found evenly through the whole plow layer (Lagerlöf et al., 1989; Didden, 1991; van Vliet et al., 1997).

The vertical distribution of enchytraeids is also influenced by changes in soil moisture content and temperature. Enchytraeids are known to react rapidly to changes in soil moisture (Nielsen, 1955a; O'Connor, 1967). In low-moisture soils, enchytraeids lose water quickly, which can be prevented by deeper migration (Nielsen, 1955b). Springett et al. (1970) showed that enchytraeids were able to move vertically as the soil water content changed, covering up to 6 cm in a few hours. Apart from the distribution of organic material and environmental variables, the vertical distribution of enchytraeids may also be species specific.

TABLE 25.12 Enchytraeid Abundances (Annual Average Number m^{-2}) in Different
Biotypes and Countries

Biotope	Country	Abundance (Number m^{-2})	Reference
Forest			
Douglas fir	Wales	134,300	O'Connor (1967)
Pinus radiata 50 stems ha^{-1}	New Zealand	64,002	Yeates (1988)
Pacific silver fir, mature stand	Washington	49,400	Piper (1982)
Pinus radiata 200 stems ha^{-1}	New Zealand	39,270	Yeates (1988)
Spruce	Norway	34,700	Abrahamsen (1972)
Rhododendron-oak 1160 m altitude	North Carolina	32,630	van Vliet et al. (1995)
Rhododendron-oak 750 m altitude	North Carolina	26,811	van Vliet et al. (1995)
Pine	Norway	22,900	Abrahamsen (1972)
Pinus radiata 100 stems ha^{-1}	New Zealand	21,391	Yeates (1988)
Scots pine forest	Sweden	16,200	
Deciduous forest	United Kingdom	14,590	Phillipson et al. (1979)
Spruce	South Finland	13,400	Huhta and Koskenniemi (1975)
Pacific silver fir, young stand	Washington	11,400	Piper (1982)
Pinus radiata 0 stems ha^{-1}	New Zealand	10,647	Yeates (1988)
Spruce	South Finland	8,200	Huhta and Koskenniemi (1975)
Spruce	North Finland	4,000	Huhta and Koskenniemi (1975)
Arable Land			
Sugar beet	The Netherlands	30,000	Didden (1991)
Winterwheat	The Netherlands	19,437	Didden (1991)
NT corn-clover	Georgia	16,830	van Vliet et al. (1995)
CT corn-clover	Georgia	15,270	van Vliet et al. (1995)
Potato field	Poland	13,200	Ryl (1977)
Barley, no N	Sweden	10,000	Lagerlöf et al. (1989)
Rye field	Poland	9,800	Ryl (1977)
Barley, 120 kg N	Sweden	8,100	Lagerlöf et al. (1989)
Rice/wheat/barley (organic)	Japan	4,940	Nakamura (1989)
Rice/wheat/barley (conven.)	Japan	525	Nakamura (1989)
Moor			
Juncus peat	United Kingdom	145,000	Peachey (1963)
Nardus	United Kingdom	71,000	Peachey (1963)
Blanket bog	United Kingdom	40,000	Standen (1973)
Fen	Canada	5,600	Dash and Cragg (1972)
Grassland			
Grassland soil	Sweden	24,000	Persson and Lohm (1977)
Lucerne ley	Sweden	9,900	Lagerlöf et al. (1989)
Grassland 10 sheep ha^{-1}	Australia (NSW)	6,000	King and Hutchinson (1976)
Grassland 30 sheep ha^{-1}	Australia (NSW)	2,300	King and Hutchinson (1976)

Dósza-Farkas (1973) findings for a Hungarian forest revealed that *Stercutus niveus* Michaelsen, 1888 had a pattern of distribution and migration differing substantially from that of other species present, pointing out the importance of knowing species composition and behavioral characteristics of the enchytraeid community.

Nielsen (1955a) determined that the seasonal trends in population densities of enchytraeids are caused directly or indirectly by external factors and do not arise from any inherent rhythmic fluctuation. According to O'Connor (1957), seasonal trends in enchytraeid numbers can be predicted from climatological data. In a temperate oceanic climate, soil moisture would be sufficient for breeding during the complete year, resulting in a generally high density of enchytraeids, which would be decreasing in winter, due to temperatures too low for breeding. A south continental climate would have low abundances in summer and fall when temperature is too high and moisture too low to be suitable for breeding. Maximum densities would occur in winter and spring. Yeates (1986) found a significant correlation between enchytraeid

abundances in a pasture in New Zealand and soil moisture and temperature at the sampling occasion or with these environmental data from 3 weeks earlier. Abrahamsen (1972) found that enchytraeid abundances were correlated with soil moisture but also with vegetation type. Environmental conditions can cause changes in the active state of the animals (e.g., cocoons, diapause) and/or accelerate or retardate the reproductive cycle (Didden, 1993).

25.5.7 Enchytraeids and Organic Matter Decomposition

Abrahamsen (1990) studied the influence of *C. sphagnetorum* on N mineralization in homogenized mor humus. He found that temperature, moisture, and the presence of enchytraeids significantly influenced N mineralization. The effect of *Cognettia* on N mineralization was most important in the very moist soil where the species reached its greatest abundance. Presence of enchytraeids in the systems increased the level of NH_4^+ and NO_3^- by 18%. Almost 40% of this difference was explained by the decomposition of dead enchytraeids. Didden (1995) also found that the decomposition of litter and nitrogen mobilization was enhanced in the presence of enchytraeids. Enchytraeids have also been shown to stimulate nitrogenase activity in soil (Šimek et al., 1991). Cole et al. (2002) estimated that enchytraeids accounted for up to 26% of dissolved organic carbon (DOC) in blanket peat bogs; a positive relationship between soil temperature and enchytraeid abundance suggested that climate warming of 2.5°C could result in an 11% increase in enchytraeid-induced release of DOC.

Enchytraeids can influence decomposition processes by affecting soil microorganisms in several ways. Grazing might lead to a younger microbial population, that is, a population with a higher metabolic activity (Sparling et al., 1981). Also due to the microbivorous and saprovorous feeding of enchytraeids, the release of nutrients from the microbial biomass is promoted. Williams and Griffiths (1989) showed that enchytraeids enhanced the leaching of nutrients out of columns by reducing nutrient immobilization in the microbial biomass. The activities of enchytraeids may also change the species composition of the microbial community and may inhibit the growth of fungal hyphae. In a laboratory experiment, Forster et al. (1995) showed that enchytraeids increased the CO_2 production and decreased microbial biomass. Apparently, the metabolic quotient of the microorganisms was increased by the enchytraeids.

Enchytraeids seem to be more efficient in promoting mineralization than microarthropods. Sulkava et al. (1996) found in microcosms with microorganisms and enchytraeids and with and without microarthropods that the amount of NH_4^+ mineralized was strongly correlated with enchytraeid biomass.

Wolters (1988) demonstrated that the enchytraeid species *Mesenchytraeus glandulosus* Levinsen, 1884 had a large influence on litter decomposition processes in beech forest soil. On the one hand, the enchytraeid promoted decomposition, for example, through an increase in the inoculation capacity of the microflora and through alteration of the physiochemical conditions of the substrate. On the other hand, *M. glandulosus* was capable of decreasing mineralization rates, for example, by reducing the number of microbial inhabitants in the environment and probably by occluding the products synthesized by soil fungi in their feces. Therefore, the net effect of the enchytraeids on litter decomposition must be regarded as the result of both promotory and inhibitory influences on the soil microflora.

Golebiowska and Ryszkowski (1977) found that tillage shifted the proportional importance of the Annelida from earthworms to the metabolically more active enchytraeids. They concluded that in conventionally tilled fields Enchytraeidae were important in the process of organic matter mineralization due to their high respiration. Hendrix et al. (1986) came to a similar conclusion: they hypothesized that enchytraeids were more important in the detritus food web in conventionally tilled than in no-tillage fields.

Kasprzak (1982) concluded that in cultivated fields in Poland the contribution of enchytraeids to the process of organic matter mineralization was much greater than that of earthworms. The percentage of total soil respiration that can be attributed to enchytraeid activity varied from 0.4% in a potato field in Poland (Golebiowska et al., 1974) to 3.2% in a sugar beet field in the Netherlands (Didden, 1991).

So, enchytraeids can influence the decomposition of organic matter by comminution of the material, which will enhance decomposition by the microbial biomass, by their selective grazing of the microbial population, by incorporating soil organic matter (SOM) into soil and soil aggregates (see in the following section) (e.g., Toutain et al., 1982; Ponge, 1984; Toutain, 1987; Wolters, 1988; Dash, 1990), by the dispersal of spores and fungal propagules and by immobilization of nutrients in their biomass (Zachariae, 1964; Abrahamsen, 1990; Didden, 1993).

25.5.8 Enchytraeids and Soil Structure

Enchytraeids excrete fecal pellets consisting of fine particles, with little cellulose plant residues. In coniferous forests, these pellets contain a considerable portion of humus (Zachariae, 1963), while in mineral soils the fecal pellets are more sponge-like structures in which humus and loamy parts are combined (Kasprzak, 1982). Fecal pellets produced by enchytraeids have been described by several authors (Zachariae, 1964; Babel, 1968; Rusek 1975, 1985). Babel (1968) measured the longitudinal axis of excrements as being between 40 and 100 μm (average 60 μm) long. Didden (1990) concluded that enchytraeids have a positive effect on the aggregate stability in the 600–1000 μm fraction. These aggregates might have been partly composed of enchytraeid excrements with a size of about 200 μm. However, he observed no significant effects of enchytraeids on the distribution of water stable aggregates. Vaculik et al. (2004) found that enchytraeids consumed cyanobacteria in biological crusts on granite outcrops in tropical forests and savannas and formed a humus layer consisting of fecal pellets below the crust; they concluded that this accumulated organic

matter facilitates plant succession. In the O horizon of forest soils, enchytraeids are secondary decomposers of the larger excrements of soil macrofauna (Rusek, 1985). They make narrow channels in the large spongy earthworm excrements (Zachariae, 1964) and mix organic aggregates into mineral soil (Karroum et al., 2005). A number of researchers (e.g., Zachariae, 1964; Babel, 1968; Pawluk, 1987) have pointed to the occasionally large numbers of enchytraeid fecal pellets in mor and moder humus soils. When high densities of enchytraeids are present in soil ecosystems, their fecal matter may be important for SOM dynamics and changes in soil structure. Encapsulation of organic substances into fecal pellets will protect this material from extensive decomposition. Also these microaggregates may in turn serve as building blocks for macroaggregates (Tisdall and Oades, 1982).

Many enchytraeid researchers (e.g., O'Connor, 1967; Ponge, 1984) have mentioned the transport of mineral particles by enchytraeids, either by ingestion or attached to the body surface. van Vliet et al. (1995) calculated the amount of soil ingested by enchytraeids in an agricultural field and in rhododendron-oak forest ecosystems in the United States. In the agricultural field, enchytraeids ingested 2180 g m^{-2} year^{-1} of mineral soil. In the forests, only about 400 g m^{-2} year^{-1} of soil was ingested. Based on these estimates, enchytraeids transported 1% and 0.5% of the soil volume (0–15 cm) in the agricultural field and forest, respectively. The enchytraeid community in the agricultural field was composed for more than 50% of large *Fridericia* enchytraeids, which can contain large quantities of soil in their guts. It seems that, due to the species-specific differences between the forest and the agricultural field, the effect of enchytraeids on soil structure will be greater in agricultural fields. However, Didden (1990) found that the amount of soil transported by *Enchytraeus buchholzi* through ingestions was very low, about 0.001%–0.01% of the total soil volume. The large difference between these two measurements is mainly due to different methods of determining the amount of ingested soil, but also due to species-specific differences.

The burrowing activity of enchytraeids in soil is poorly known. Górny (1984) stated that enchytraeids were unable to burrow channels and did not use earthworm channels for penetrating deeper layers of the soil; for vertical migration, enchytraeids were dependent upon the smaller pores present in the soil. Alternatively, Jegen (1920) and Rusek (1985) concluded that enchytraeids can make active microtunnels in the soil matrix. Standen (1984) suggested that large enchytraeid species, like *F. galba*, formed burrows and transported or deposited material as feces to a depth of 3.5–4.0 cm. Species of other genera did not form visible burrows but material was transported 0.5 cm in some cases. Didden (1990) concluded that *Enchytraeous bucholzi* Vejdovsky, 1879 increased pore continuity and raised the volume of pores corresponding to their body size (size 50–200 μm). His analysis of thin sections led to the conclusion that enchytraeids removed pore necks.

Observations of enchytraeids kept in small microcosms constructed of a thin layer of soil sandwiched between glass plates have led to the conclusion that enchytraeids do make channels in soil (van Vliet et al., 1993). A later experiment (van Vliet et al., 1997) showed that in the presence of incorporated organic matter, enchytraeids increased the hydraulic conductivity of a soil column. The impact of enchytraeids on hydraulic conductivity, however, depended not only on the incorporation of organic matter in the soil but also on the number of enchytraeids present and the duration of the experiment.

Overall, it seems reasonable to assume that enchytraeids affect soil structure through their fecal pellets and their burrowing capacity. However, the total effect of enchytraeids on soil structure in an ecosystem depends on the species composition of the enchytraeid community, enchytraeid abundances, and on environmental conditions.

References

Abrahamsen, G. 1969. Sampling design in studies of population densities in Enchytraeidae (Oligochaeta). Oikos 20:45–66.

Abrahamsen, G. 1972. Ecological study of Enchytraeidae (Oligochaeta) in Norwegian coniferous forest soils. Pedobiologia 12:26–82.

Abrahamsen, G. 1990. Influence of *Cognettia sphagnetorum* (Oligochaeta) on nitrogen mineralization in homogenized mor humus. Biol. Fertil. Soils 9:159–162.

Abrahamsen, G., and L. Strand. 1970. Statistical analysis of population density data of soil animals, with particular reference to Enchytraeidae (Oligochaeta). Oikos 21:276–284.

Babel, U. 1968. Enchytraeen-Losungsgefüge in Löss. Geoderma 2:57–63.

Briones, M.J.I., and P. Ineson. 2002. Use ^{14}C carbon dating to determine feeding behaviour of enchytraeids. Soil Biol. Biochem. 34:881–884.

Briones, M.J.I., P. Ineson, and A. Heinemeyer. 2007. Predicting potential impacts of climate change on the geographical distribution of enchytraeids: A meta-analysis approach. Global Change Biol. 13:2252–2269.

Brockmeyer, V., R. Schmid, and W. Westheide. 1990. Quantitative investigations of the food of two terrestrial enchytraeid species (Oligochaeta). Pedobiologia 34:151–156.

Chalupský, J., Jr., and J. Lepš. 1985. The spatial pattern of Enchytraeidae (Oligochaeta). Oecologia 68:153–157.

Christensen, B. 1956. Studies on Enchytraeidae 6. Technique for culturing Enchytraeidae, with notes on cocoon types. Oikos 7:303–307.

Christensen, B., and K. Dozsa-Farkas. 2006. Invasion of terrestrial enchytraeids into two postglacial tundras: Northeastern Greenland and the Arctic Archipelago of Canada (Enchytraeidae, Oligochaeta). Polar Biol. 29:454–466.

Cole, L., R.D. Bardgett, P. Ineson, and J.K. Adamson. 2002. Relationships between enchytraeid worms (Oligochaeta), climate change, and the release of dissolved organic carbon from blanket peat in northern England. Soil Biol. Biochem. 34:599–607.

Dash, M.C. 1990. Enchytraeidae, p. 311–340. *In* D.L. Dindal (ed.) Soil biology guide. John Wiley & Sons, New York.

Dash, M.C., and J.B. Cragg. 1972. Ecology of Enchytraeidae (Oligochaeta) in Canadian Rocky Mountain soils. Pedobiologia 12:323–335.

Dash, M.C., B. Nanda, and P.C. Mishra. 1981. Digestive enzymes in three species of Enchytraeidae (Oligochaeta). Oikos 36:316–318.

Didden, W.A.M. 1990. Involvement of Enchytraeidae (Oligochaeta) in soil structure evolution in agricultural fields. Biol. Fertil. Soils 9:152–158.

Didden, W.A.M. 1991. Population ecology and functioning of Enchytraeidae in some arable farming systems. Ph.D. Dissertation. Agricultural University Wageningen. Wageningen, the Netherlands.

Didden, W.A.M. 1993. Ecology of Enchytraeidae. Pedobiology 37:2–29.

Didden, W. 1995. The effect of nitrogen deposition on enchytraeid-mediated decomposition and mobilization—A laboratory experiment. Acta Zool. Fennica 196:60–64.

Didden, W., H. Born, H. Domm, U. Graefe, M. Heck, J. Kühle, A. Mellin, and J. Römbke. 1995. The relative efficiency of wet funnel techniques for the extraction of Enchytraeidae. Pedobiology 39:52–57.

Didden, W., and J. Römbke. 2001. Enchytraeids as indicator organisms for chemical stress in terrestrial ecosystems. Ecotoxicol. Environ. Saf. 50:25–43.

Dósza-Farkas, K. 1973. Saisondynamische Untersuchungen des Enchytraeiden-Besatzes im Boden eines ungarischen Quercetum petraeae cerris. Pedobiologia 13:361–367.

Dósza-Farkas, K. 1982. Konsum verschiedener Laubarten durch Enchytraeiden (Oligochaeta). Pedobiology 23:145–156.

Dósza-Farkas, K. 1996. Reproduction strategies in some enchytraeid species, p. 25–33. *In* K. Dósza-Farkas (ed.) Newsletter on Enchytraeidae No. 5. Eötvös Loránd University, Budapest, Hungary.

Forster, B., J. Römbke, T. Knacker, and E. Morgan. 1995. Microcosm study of the interactions between microorganisms and enchytraeid worms in grassland soil and litter. Eur. J. Soil Biol. 31:21–27.

Golebiowska, J., Z. Margowski, and L. Ryszkowski. 1974. An attempt to estimate the energy and matter economy in the agroeconoses, p. 19–40. *In* L. Ryszkowski (ed.) Ecological effects of intensive agriculture. Polish Scientific Publishers, Warszawa, Poland.

Golebiowska, J., and L. Ryszkowski. 1977. Energy and carbon fluxes in soil compartments of agroecosystems. *In* U. Lohm and T. Persson (eds.) Soil organisms as components of ecosystems. Ecol. Bull. (Stockholm) 25:274–283.

Górny, M. 1984. Studies on the relationship between enchytraeids and earthworms, p. 769–776. *In* J. Szegi (ed.) Soil biology and conservation of the biosphere. Akademiai Kiado, Budapest, Hungary.

Graefe, U. 1973. Systematische Untersuchungen an der Gattung Achaeta (Enchytraeidae, Oligochaeta). Universität Hamburg, Diplomarbeit, Germany.

Hendrix, P.F., R.W. Parmelee, D.A. Crossley, Jr., D.C. Coleman, E.P. Odum, and P.M. Groffman. 1986. Detritus food webs in conventional and no-tillage agroecosystems. Bioscience 36:372–380.

Huhta, V., and A. Koskenniemi. 1975. Numbers, biomass and community respiration of soil invertebrates in spruce forests at two latitudes in Finland. Ann. Zool. Fennici 12:164–182.

Jegen, G. 1920. Die Bedeutung der Enchytraeiden fur die Humusbildung. Landwirtsch. Jarhb. Schweiz 34:55–71.

Karroum, M., B. Guillet, F. Laggoun-Defarge, J.-R. Disnar, N. Lottier, G. Villemin, and F. Toutain. 2005. Morphological evolution of beech litter (*Fagus sylvatica* L.) and biopolymer transformation (lignin, polysaccharides) in a mull and a moder, under temperate climate (Fougeres forest, Britany, France). Can. J. Soil Sci. 85:405–416.

Kasprzak, K. 1982. Review of enchytraeid (Oligochaeta, Enchytraeidae) community structure and function in agricultural ecosystems. Pedobiology 23:217–232.

King, K.L., and K.J. Hutchinson. 1976. The effects of sheep stocking intensity on the abundance and distribution of mesofauna in pastures. J. Appl. Ecol. 13:41–55.

Lagerlöf, J., O. Andrén, and K. Paustian. 1989. Dynamics and contribution to carbon flows of Enchytraeidae (Oligochaeta) under four cropping systems. J. Appl. Ecol. 26:183–199.

Lal, V.B., J. Singh, and B. Prasad. 1981. A comparison of different size samplers for collection of Enchytraeidae (Oligochaeta) in tropical soil. J. Soil Biol. Ecol. 1:21–26.

Maraldo, K., I.K. Schmidt, C. Beier, and M. Holmstrup. 2008. Can field populations of the enchytraeid, *Cognettia sphagnetorum*, adapt to increased drought stress? Soil Biol. Biochem. 40:1765–1771.

Nakamura, Y. 1979. Microdistribution of Enchytraeidae in *Zoysia* grassland (Studies on Japanese enchytraeids I). Bull. Nat. Grassl. Res. Inst. 14:21–27.

Nakamura, Y. 1982. Distribution pattern of *Achaeta camerani* (Cognetti) in grazing and cutting grasslands (studies on Japanese enchytraeids III). Edaphologia 25–26:41–47.

Nakamura, Y. 1989. Oribatids and enchytraeids in ecofarmed and conventionally farmed dryland grainfields of central Japan. Pedobiologia 33:389–398.

Nielsen, C.O. 1953. Studies on Enchytraeidae. 1. A technique for extracting Enchytraeidae from soil samples. Oikos 4:187–196.

Nielsen, C.O. 1954. Studies on Enchytraeidae. 3. The microdistribution of Enchytraeidae. Oikos 5:167–178.

Nielsen, C.O. 1955a. Studies on Enchytraeidae. 5. Factors causing seasonal fluctuations in numbers. Oikos 6:153–169.

Nielsen, C.O. 1955b. Studies on Enchytraeidae. 2. Field studies. Natura Jutl. 4–5:1–58.

Nielsen, C.O., and B. Christensen. 1959. The Enchytraeidae: Critical revision and taxonomy of European species. Natura Jutl. 8–9:1–160.

Nielsen, C.O., and B. Christensen. 1961. The Enchytraeidae: Critical revision and taxonomy of European species. Natura Jutl. (Suppl. 1) 10:1–23.

Nielsen, C.O., and B. Christensen. 1963. The Enchytraeidae: Critical revision and taxonomy of European species. Natura Jutl. (Suppl. 2) 10:1–19.

O'Connor, F.B. 1955. Extraction of enchytraeid worms from a coniferous forest soil. Nature 175:815–816.

O'Connor, F.B. 1957. An ecological study of the enchytraeid worms from a coniferous forest soil. Oikos 8:161–169.

O'Connor, F.B. 1967. The Enchytraeidae, p. 213–257. *In* A. Burgess and F. Raw (eds.) Soil biology. Academic Press, New York.

Pawluk, S. 1987. Faunal micromorphical features in moder humus of some western Canadian soils. Geoderma 40:3–16.

Peachey, J.E. 1963. Studies on Enchytraeidae (Oligochaeta) of moorland soils. Pedobiologia 2:81–95.

Persson, T., E. Bååth, M. Clarholm, H. Lundkvist, B.E. Söderström, and B. Sohlenius. 1980. Trophic structure, biomass dynamics and carbon metabolism of soil organisms in Scots pine forest. *In* T. Persson (ed.) Structure and function of northern coniferous forests. Ecol. Bull. 32:419–459.

Persson, T., and U. Lohm. 1977. Energetical significance of the annelids and arthropods in a Swedish grassland soil. Ecol. Bull. 23:1–211.

Phillipson, J., R. Abel, J. Steel, and S.R.J. Woodell. 1979. Enchytraeid numbers, biomass and respiration metabolism in a Beech woodland—Wytham Woods, Oxford. Oecologia 43:173–193.

Piper, S.R. 1982. Enchytraeid (Oligochaeta) worms: Species composition, distribution, abundance, respiration and production in two Pacific Silver fir stands. M.Sc. Thesis. University of Washington. Seattle, WA.

Ponge, J.F. 1984. Étude écologique d'un humus forestier par l'observation d'un petit volume, premiers résultats. I. La couche L1 d'un moder sous pin sylvestre. Rev. Écol. Biol. Sol. 21:161–187.

Römbke, J. 1992. New taxa since 1985. *In* J. Römbke (ed.) Newsletter on Enchytraeidae No. 3. Institut für Angewandte Bodenbiologie, Hamburg, Germany.

Römbke, J. 2003. Ecotoxicological laboratory tests with enchytraeids: A review. Pedobiologia 47:607–616.

Römbke, J., and K. Dósza-Farkas. 1996. New taxa since 1992, p. 5–9. *In* K. Dósza-Farkas (ed.) Newsletter on Enchytraeidae No. 5. Eötvös Loránd University, Budapest, Hungary.

Rusek, J. 1975. Die bodembildende funktion von Collembolen und Acarina. Pedobiology 15:299–308.

Rusek, J. 1985. Soil microstructures—Contributions on specific soil organisms. Quaest. Entomol. 21:497–514.

Ryl, B. 1977. Enchytraeids (Oligochaeta, Enchytraeidae) on rye and potato fields in Turew. Ekol. Pol. 25:519–529.

Schauermann, J. 1983. Eine Verbesserung der Extraktionsmethode für terrestrische Enchytraeiden, p. 669–670. *In* Ph. Lebrun, H.M. André, A. De Medts, C. Grégoire-Wibo, and G. Wauthy (eds.) New trends in soil biology. Dieu-Brichart, Ottignies-Louvain-la-Neuve, Belgium.

Schmelz, R.M. 2003. Taxonomy of *Fridericia* (Oligochaeta, Enchytraeidae). Revision of species with morphological and biochemical methods. Hubert and Co., Göttingen, Germany.

Schmelz, R.M. and R. Collado, 2010. A guide to European terrestrial and freshwater species of Enchytraeidae (Oligochaeta). Soil Org. 81(19), 1–176.

Smith, C.A.S., A.D. Tomlin, J.J. Miller, L.V. Moore, M.J. Tynen, and K.A. Coates. 1990. Large enchytraeid (Annelida: Oligochaeta) worms and associated fauna from unglaciated soils of the northern Yukon, Canada. Geoderma 47:17–32.

Šimek, M., V. Pizl, and J. Chalupský. 1991. The effect of some terrestrial oligochaeta on nitrogenase activity in the soil. Plant Soil 137:161–165.

Sparling, G., B.G. Ord, and D. Vaughan. 1981. Microbial biomass and activity in soils amended with glucose. Soil Biol. Biochem. 13:99–104.

Springett, J.A., J.E. Brittain, and B.P. Springett. 1970. Vertical movement of Enchytraeidae (Oligochaeta) in moorland soils. Oikos 21:16–21.

Standen, V. 1973. The production and respiration of an enchytraeid population in blanket bog. J. Anim. Ecol. 42:219–245.

Standen, V. 1984. Production and diversity of enchytraeids, earthworms and plants in fertilized hay meadow plots. J. Appl. Ecol. 21:293–312.

Sulkava, P., V. Huhta, and J. Laakso. 1996. Impact of soil fauna structure on decomposition and N-mineralisation in relation to temperature and moisture in forest soil. Pedobiologia 40:505–513.

Tisdall, J.M., and J.M. Oades. 1982. Organic matter and water stable aggregates in soils. J. Soil Sci. 33:141–163.

Toutain, F. 1987. Activité biologique des sols, modalités et litho-dépendence. Biol. Fertil. Soils 3:31–38.

Toutain, F., G. Villemin, A. Albrecht, and O. Reisinger. 1982. Etude ultrastructurale des processes de biodégradation. II. Modèle enchytraeides-litière de feuillus. Pedobiologia 23:145–156.

Urbášek, F., and J. Chalupský. 1991. Activity of digestive enzymes in 4 species of Enchytraeidae (Oligochaeta). Rev. Ecol. Biol. Sol. 28:145–154.

Vaculik, A., C. Kounda-Kiki, C. Sarthou, and J.F. Ponge. 2004. Soil invertebrate activity in biological crusts on tropical inselbergs. Eur. J. Soil Sci. 55:539–549.

van Vliet, P.C.J., M.H. Beare, and D.C. Coleman. 1995. A comparison of the population dynamics and functional roles of Enchytraeidae (Oligochaeta) in hardwood forest and agricultural ecosystems, p. 237–245. *In* H.P. Collins, G.P. Robertson, and M.J. Klug (eds.) The significance and regulation of soil biodiversity. Kluwer Academic, Dordrecht, the Netherlands.

van Vliet, P.C.J., D.C. Coleman, and P.F. Hendrix 1997. Population dynamics of Enchytraeidae (Oligochaeta) in different agricultural systems. Biol. Fertil. Soils 25:123–129.

van Vliet, P.C.J., L.T. West, P.F. Hendrix, and D.C. Coleman. 1993. The influence of Enchytraeidae (Oligochaeta) on the soil porosity of small microcosms. Geoderma 56:287–299.

Williams, B.L., and B.S. Griffiths. 1989. Enhanced nutrient mineralization and leaching from decomposing Sitka spruce litter by enchytraeid worms. Soil Biol. Biochem. 21:183–188.

Wolters, V. 1988. Effects of *Mesenchytraeus glandulosus* (Oligochaeta, Enchytraeidae) on decomposition processes. Pedobiologia 32:387–398.

Yeates, G. 1986. Enchytraeidae—Some population estimates for grasslands and a New Zealand bibliography. New Zealand Soil Bureau Scientific Report 77. Lower Hutt, New Zealand.

Yeates, G. 1988. Earthworm and enchytraeid populations in a 13-year-old agroforestry system. N.Z. J. For. Sci. 18:304–310.

Zachariae, G. 1963. Was leisten Collembolen für den Waldhumus? p. 57–68. *In* J. Doeksen and J. van der Drift (eds.) Soil organisms. North-Holland, Amsterdam, the Netherlands.

Zachariae, G. 1964. Welche Bedeutung haben Enchytraen im Waldboden? p. 57–67. *In* A. Jongerius (ed.) Soil micromorphology. Elsevier Publishing, Amsterdam, the Netherlands.

25.6 Earthworms

P.C.J. van Vliet

Paul F. Hendrix

M.A. Callaham, Jr.

25.6.1 Introduction

Among the soil fauna, earthworms are perhaps the most widely recognized and, along with ants and termites, function as ecosystem engineers with significant effects on soil structure and processes (Lavelle and Spain, 2001; Wardle, 2002). For these reasons, earthworms have been intensively studied for their potential benefits in agriculture, waste management, and land reclamation. The scientific literature on earthworms dates back over 200 years to the taxonomic description of *Lumbricus terrestris* by Linnaeus (1758). The modern era of earthworm research, in the context of soil science, began with Darwin (1881) and a vast literature has accumulated since then. Reviews of the literature from the past several decades can be found in Satchell (1983), Lee (1985), Dindal (1990), Curry (1994), Hendrix (1995), Edwards and Bohlen (1996), Lavelle et al. (1999), Lavelle and Spain (2001), and Edwards (2004). This chapter draws from these and other works to give a brief overview of earthworm biology, ecology, methods of collection, and analyses of earthworm tissues in food-web studies or for advanced systematic and taxonomic work.

25.6.2 Basic Taxonomy

Earthworms are classified within the phylum Annelida, class Clitellata, and subclass Oligochaeta (increasingly, Crassiclitellata). Seventeen families are usually recognized worldwide, making up the semiaquatic and terrestrial forms commonly known as earthworms (Jamieson et al., 2002). Approximately 3700 species of these megadrile oligochaetes have been described, and it is estimated that total global species richness may exceed 7000 (Reynolds, 1994; Fragoso et al., 1999; Lavelle and Lapied, 2003). Earthworm families along with their biogeographic origins are listed in Table 25.13. This table also includes the family Enchytraeidae, which is discussed in Section 25.5. Although the taxonomy and systematics of earthworms is

TABLE 25.13 Classification and Regions of Origin of Major Families of the Terrestrial Oligochaetes

Phylum: Annelida		
Class: Clitellata		
Subclass: Oligochaeta (Crassiclitellata)		
Order: Haplotaxida		
Suborder: Enchytraeina		
Family:	Enchytraeidae	NH, SH
Suborder: Lumbricina		
Family:	Lumbricidae	NH—NA, EU
	Komarekionidae	NH—NA
	Sparganophilidae	NH—NA
	Lutodrilidae	NH—NA
	Megascolecidae	NH, SH—NA, SA, OC, AS
	Glossoscolecidae	SH—SA
	Eudrilidae	SH—AF
	Acanthodrilidae[a]	SH—AS, SA, AF
	Octochaetidae[a]	SH—OC
	Ocnerodrilidae	SH—SA, AF, AS, MA
	Ailoscolecidae	NH—EU
	Hormogastridae	NH—ME
	Kynotidae	SH—MA
	Microchaetidae	SH—AF
	Almidae	SH—SA, AF, AS
	Biwadrilidae	NH—JA

Source: Summarized from Wallwork, J.A. 1983. Earthworm biology. Studies in biology No. 161, Institute of Biology, Camelot Press, Southampton, U.K.; Sims, R.W. 1980. A classification and the distribution of earthworms, suborder Lumbricina (Haplotaxida: Oligochaeta). Bull. Br. Mus. Nat. Hist. Zool. 39:103–124; Jamieson, B.G.M. 1981. Historical biogeography of the Australian Oligochaeta, p. 887–921. *In* A. Keast (ed.) Ecological biogeography of Australia. W. Junk, The Hague, the Netherlands; Reynolds, J.W., and D.G. Cook. 1993. Nomenclatura Oligochaetologica, Supplementum Tertium: A catalogue of names, descriptions and type specimens of the Oligochaeta. New Brunswick Museum Monograph Series No. 9. (Nat. Sci.). New Brunswick, NJ; Jamieson, B.G.M., S. Tillier, A. Tillier, J.-L. Justine, E. Ling et al. 2002. Phylogeny of the Megascolecidae and Crassiclitellata (Annelida, Oligochaeta): Combined versus partitioned analysis using nuclear (28S) and mitochondrial (12S, 16S) rDNA. Zoosystema 24:707–734.

NH, Northern Hemisphere; SH, Southern Hemisphere; AF, Africa; AS, Asia; EU, Europe; JA, Japan; MA, Madagascar; ME, Mediterranean; NA, North America; OC, Oceania; SA, South America.

[a] Taxonomic status in question (Jamieson et al., 2002, Jamieson, 2006).

now reasonably stable at the family and generic level, there has long been confusion and controversy associated with certain groups in terms of species names. This situation results in part from the remarkable variability in external physical characters for some groups (e.g., the lumbricid genus *Aporrectodea*) that are reliable for taxonomic diagnoses for other groups (e.g., the lumbricid genus *Lumbricus*). Likewise, the frequent occurrence of parthenogenetic (asexual) reproduction in some groups, with variable reduction of male reproductive parts that are also used for taxonomic diagnosis, has resulted in different morphotypes of a single species being named as several species (e.g., the megascolecid genus *Amynthas*; Gates, 1972). It is to be hoped that advances in genetic and molecular techniques will provide resolution for many of these problems, but this work is just under way (see Briones et al., 2009; Chang et al., 2009; Dupont, 2009). Further information on earthworm taxonomy and biogeography can be found in Reynolds (1977, 1994, 1995), Sims (1980), Jamieson (1981), Gates (1982), Hendrix (1995), James 1995, Edwards and Bohlen (1996), Sims and Gerard (1999), Omodeo (2000), Jamieson (2006), and Hendrix et al. (2008).

The families Lumbricidae and Megascolecidae are ecologically the most important in North America, Europe, Australia, and Asia, while the families Glossoscolecidae and Eudrilidae are prevalent in South America and sub-Saharan Africa, respectively. Species from several of these families have been introduced worldwide by human activities and now dominate the earthworm fauna in many areas. Such "peregrine" or "anthropochorous" species are highly successful in agricultural or otherwise disturbed areas and often show significant effects on soil processes (Lee, 1985; Edwards and Bohlen, 1996; Hendrix and Bohlen, 2002). Whether introduced earthworms displace native species or occupy areas devoid of native species due to disturbance is a subject of debate (Lee 1961; Kalisz and Wood, 1995). Earthworm invasion biology and ecology have become topics of keen interest recently (Bohlen et al., 2004; Hendrix, 2006; Hendrix et al., 2008).

25.6.3 Biology and Ecology

Earthworms are elongated, cylindrical, segmented invertebrates, ranging in length from a few millimeters to 1.4 m, such as the giant Australian *Megascolides australis*. They consist of a relatively simple, tube-within-a-tube body plan, the internal tube comprising the alimentary canal. The body segments are separated by septa and are filled with coelomic fluid that provides a dynamic, hydrostatic "skeleton" for locomotion. When fully hydrated in free water, earthworm body weight may consist of 80%–90% water, but under ideal soil moisture conditions water content is typically 65%–75% (Lee, 1985). Water is lost from the body as mucus secretions onto the body surface from epidermal gland cells. Respiration occurs through the moist integument, where blood in subcuticular capillaries absorbs oxygen that is transported throughout the body in a closed vascular system driven by a series of muscular heart-like structures. Earthworms are hermaphroditic, each individual carrying male and female reproductive organs, but a number of species display

parthenogenesis. During amphimictic reproduction, sperm is exchanged between two individuals, stored in spermathecae, and later released along with eggs into cocoons secreted by the glandular clitellum, a characteristic thickening along several anterior segments. Single or multiple hatchlings emerge from cocoons after a period of embryological development determined by species and prevailing environmental conditions. Further details of earthworm biology can be found in Wallwork (1983), Lee (1985), Edwards and Bohlen (1996), and Sims and Gerard (1999).

Earthworms occur worldwide in most areas where climatic conditions are favorable for at least part of the year (all but desert and polar conditions); temperature is the main controlling factor globally, whereas soil moisture strongly influences local patterns of abundance and distribution (Lee, 1985; Curry, 2004). Across this range of habitats, earthworms display a wide array of morphological, physiological, and behavioral adaptations to environmental conditions. During unfavorable periods (e.g., drought), many species are able to enter a temporary dormant state (aestivation or diapause) or produce resistant cocoons that can hatch when conditions improve.

The abundance of earthworms across habitats is highly variable depending on climatic and edaphic conditions, ecosystem type, and the degree to which the habitat has been altered, for example, by agriculture. Under otherwise suitable conditions, soil C concentration has been shown to be highly correlated with earthworm population density and biomass (Edwards, 1983; Hendrix et al., 1992). Earthworm density and biomass in a variety of habitats worldwide are presented in Table 25.14. Densities range from <10 to >2000 m^{-2}, with the highest values occurring in grasslands (especially fertilized pastures) and the lowest in acid or arid soils (coniferous or sclerophyllous forest). Typical densities from temperate deciduous or tropical forests and certain arable systems range from ca. 100 to 400 m^{-2}, depending on intensity and nature of disturbance. Earthworm biomass tends to track density, but biomass comparisons can be problematic due to different methods used by various investigators (Section 25.5.1). Because earthworm populations often show seasonal variation in abundance (especially in temperate regions), time of sampling also affects density and biomass estimates.

Within habitats, earthworms often show heterogeneous spatial distributions. "Single-tree-influence," spatial distributions of other controlling environmental factors (e.g., soil texture, moisture, or organic matter content), or intrinsic population characteristics (e.g., fecundity, body size, and dispersal ability) often result in aggregation or clumped distribution patterns of earthworm populations across landscapes. In addition, local habitat and feeding preferences of various earthworm species dictate their vertical distributions within the soil profile (Boetcher and Kalisz, 1991; Barois et al., 1999; Rossi et al., 2006). A categorization of earthworm life forms or functional types based on habitat and feeding ecology is presented in Table 25.15. These categories describe niche separation of earthworm species within a soil volume. Polyhumic, epigeic, and epiendogeic species utilize litter and organically enriched surface layers; poly-, meso-, and oligohumic endogeic species inhabit mineral soil

TABLE 25.14 Abundance and Biomass of Earthworms in Selected Habitats from Various Parts of the World

Habitat	Location	Collection Method	Earthworm Taxa	Abundance No. m^{-2}	Biomass g m^{-2} Fresh wt
Sown pastures	New Zealand	Hand sorting	Lumbricidae	208–775	60–241
				740–1235	146–303
				690–2020	305 (mean)
Sown pastures	South Australia	Hand sorting	Lumbricidae	460–625	62–78
Sown pastures	South Africa	Hand sorting	Lumbricidae	72–1112	—
Fertilized pasture	Argentina	Hand sorting	Lumbricidae, Megascolecidae, and Glossoscolecidae	27	—
Pastures with heavy rates of fertilizers	Ireland	Hand sorting	Lumbricidae	400–500	100–200
Old pasture	Sweden	Hand sorting	Lumbricidae	109	59
Old pasture	England	Hand sorting	Lumbricidae	390–470	52–110
Old pasture	Wales	Hand sorting	Lumbricidae	646	149
Old pasture	France	Hand sorting	Lumbricidae	288	125
Fallow	South Australia	Hand sorting	Lumbricidae	210–460	16–76
Fallow	Wales	Hand sorting	Lumbricidae	226	79
Cropland	South Australia	Hand sorting	Lumbricidae	20–25	2–2.5
Cropland	Romania	Hand sorting	Lumbricidae	5–100	0.5–20
Natural grassland	Romania	Hand sorting	Lumbricidae	200 (mean)	10–60
Natural grassland	Wales	Hand sorting	Lumbricidae	22	8
Natural grassland	Tennessee	Hand sorting	Lumbricidae	13–41	3.2–7.5
Natural grassland	South Africa	Hand sorting	Glossoscolecidae	74	96
Natural grassland	India	Hand sorting	Megascolecidae and Ocnerodrilidae	64–800	6–60
Natural grassland	New Zealand	Hand sorting	Megascolecidae	250–750	—
Tropical savannas	Ivory Coast	Hand sorting and washing/sieving	Megascolecidae and Eudrilidae	230	49
Orchard	Netherlands	Hand sorting	Lumbricidae	300–500	75–122
Orchards	Australia	Hand sorting	Lumbricidae	150	—
Mulched and irrigated orchards	Australia	Hand sorting	Lumbricidae	2000	—
Garden	Egypt	Hand sorting	Megascolecidae	420	153
Gardens	Argentina	Hand sorting	Lumbricidae, Megascolecidae, and Glossoscolecidae	73	—
Taiga	Finland	Hand sorting	Lumbricidae	17.4	2.8
	Siberia			23.0	8.4
	USSR			3–7	—
Northern European and Asian coniferous forests	Finland	Hand sorting	Lumbricidae	14–68	—
	Sweden			103–167	30–35
	USSR			12	—
	Japan			27–72	—
Spruce forest with lime topdressing	USSR	Hand sorting	Lumbricidae	1000	—
European deciduous forests	England	Hand sorting	Lumbricidae	118–138	—
	USSR			136	68.3
	Czechoslovakia			106	98.1
North American deciduous forests	Canada	Hand sorting	Lumbricidae	240–780	38–109
	Tennessee			2–96	1.3–14
	Indiana			14–124	26.3–280.3
Dry schlerophyll forest	Australia	Hand sorting	Megascolecidae	7–38	1.3–25.5
Wet schlerophyll forest				34–76	12.3–47.9
Subalpine woodland				15–106	5.7–35.7

(continued)

TABLE 25.14 (continued)   Abundance and Biomass of Earthworms in Selected Habitats from Various Parts of the World

Habitat	Location	Collection Method	Earthworm Taxa	Abundance No. m^{-2}	Biomass g m^{-2} Fresh wt
Gallery forests	Ivory Coast	Hand sorting and washing/sieving	Megascolecidae and Eudrilidae	70–103	3.4–6.8
Tropical forest	Nigeria		Eudrilidae	34	10.2
Tropical forest	Nigeria	Hand sorting	Eudrilidae	61.7	2.5
Lowland dipterocarp forest	Sarawak	Hand sorting	Moniligastridae and Megascolecidae	37–92	0.7–1.3
Lower montane forest			Moniligastridae and Megascolecidae	55	3.1
Upper montane forest			Moniligastridae and Megascolecidae	47–108	1.8–2.7
Upper montane low forest			Megascolecidae	2–24	0.2–2.1

Source: Reproduced with permission from Lee, K.E. 1985. Earthworms: Their ecology and relationships with soils and land use. Academic Press, Sydney, Australia.

TABLE 25.15   Ecological Strategies of Earthworms

Epigeic (litter dweller)—Mesophage; detritivore

Lives in and consumes litter; small size; uniformly pigmented

(e.g., *L. rubellus*, *Bimastos* spp., *Dendrobaena octaedra*, *Dendrobaena rubida*, *Eisenia foetida*, *Amynthas* spp.)

Endogeic (subsoil dweller)—Microphage; geophage; (epiendogeic or hypoendogeic; oligohumic, mesohumic, or polyhumic)

Lives in horizontal, branching burrows in organomineral layer; consumes soil; small to large in size; weakly pigmented

(e.g., *Aporrectodea caliginosa*, *Octolasion cyaneum*, *Diplocardia* spp., *Pontoscolex corethrurus*)

Anecic (topsoil dweller)—Macrophage; detritivore

Lives in deep vertical burrows, casting on surface; emerges at night to draw down organic matter (plant residue, etc.); large as adults (200–1100 mm); brown pigment anteriorly and dorsally

(e.g., *L. terrestris*, *Allolobophora longa*)

Source: Lee, K.E. 1959. The earthworm fauna of New Zealand. Department of Scientific and Industrial Research Bulletin No. 130. Wellington, New Zealand; Bouché, M.B. 1977. Stratégies lombriciennes. *In* U. Lohm and T. Persson (eds.) Soil organisms as components of ecosystems. Ecol. Bull. (Stockholm) 25:122–133; Lavelle, P. 1983. The structure of earthworm communities, p. 449–466. *In* J.E. Satchell (ed.) Earthworm ecology: From Darwin to vermiculture. Chapman and Hall, London, U.K.; Barois, I., M. Brossard, P. Lavelle, J. Tondoh, J. Kanyonyo, A. Martínez, J. Jiménez et al. 1999. Ecology of earthworm species with large environmental tolerance and/or extended distribution, p. 57–85. *In* P. Lavelle, L. Broussssard, and P. Hendrix (eds.) Earthworms management in tropical agroecosystems. CABI Publishing, New York.

within the rhizosphere and beyond; and anecic species exploit both the surface litter as a source of food and the mineral soil as a refuge. Lee (1985) summarizes data showing that within a particular soil, commonly less than a half-dozen earthworm species are found. The species in a given earthworm association often effectively partition the soil volume according to the functional categories mentioned above. Furthermore, the activities of earthworms within these categories influence biogeochemical processes in various ways. For example, epigeic species facilitate the breakdown and mineralization of surface litter, whereas anecic species incorporate organic matter deeper into the soil profile and enhance aeration and water infiltration through burrow formation (Lee, 1985; Shipitalo and Le Bayon, 2004).

For management of earthworms in agroecosystems, Lee (1991, 1995) recommends that "target earthworm communities" consist of one or more anecic/epigeic species that make deep vertical burrows and that cast on the surface and bury residues, and one or more endogeic species that feed belowground on dead roots and organic matter and that make horizontal burrows. Diverse assemblages of earthworms may more effectively exploit soil resources and influence a wider array of processes, such as organic matter turnover, in addition to soil structural properties, than a single species.

25.6.4 Importance to Soil Processes

Where earthworms are abundant, they can exert significant influence on soil processes through effects on organic matter and nutrient cycling, and on soil structure. These topics are reviewed in Lee (1985), Hendrix (1995), Edwards and Bohlen (1996), Lavelle et al. (1999), Lavelle and Spain (2001), and Edwards (2004).

25.6.4.1 Organic Matter Dynamics and Nutrient Cycling

Effects of earthworms on organic matter and nutrient cycling are closely linked with the life form and feeding ecology of earthworms (Table 25.15). Epigeic species typically live in the O and upper A soil horizons where, through feeding and casting activities, they mix mineral soil and plant litter, fragment organic particles, inoculate them with microbes, and thereby accelerate organic matter decomposition rates. Anecic forms pull surface litter into their burrows, thus transporting organic material deeper into the soil profile. They cast on the soil surface, mixing organic and mineral particles in the litter layer. The activities of both epigeic and anecic earthworms produce "mull" soil horizons, defined as those in which organic matter is intimately incorporated into the upper mineral soil of a well-developed A horizon overlain with litter or humus layers <2 cm thick. The extreme case is termed "vermimull," in which the Ah horizon is granular and characterized by strong organomineral complexes consisting of earthworm casts (Green et al., 1993). Endogeic earthworms feed within the soil on organic matter and microbes associated with the rhizosphere or mineral soil. As mentioned previously, they are termed oligohumic, mesohumic, or polyhumic,

depending on the level of organic enrichment of their substrate. Casts and burrows of endogeic earthworms are also sites of increased microbial activity and organic matter decomposition, and the presence of these worms has been shown to positively influence the availability of nutrients in some soils (e.g., Callaham and Hendrix, 1997; Chapuis-Lardy et al., 2009). Mineralization of organic matter in earthworm casts and burrow linings produces zones of nutrient enrichment compared to bulk soil. These "hot spots" (the "drilosphere") are often sites of enhanced activity of plant roots and other soil biota (Beare et al., 1995; Lavelle and Spain, 2001). Indeed, most of the effects that earthworms have on soil organic matter and nutrient dynamics in ecosystems are mediated through their interactions with other soil biota and in particular soil microbiota, as earthworms have been shown to influence the size and composition of microbial communities (e.g., Svensson and Friberg, 2007; Pawlett et al., 2009), as well as the activity of microbial communities in terms of both total soil respiration and gaseous losses of nitrogen (Speratti et al., 2007).

25.6.4.2 Soil Structure

Soil structure is affected by earthworms principally through production of casts, which form stable aggregates upon and within the soil, and formation of burrows, which produce macropores that may increase water infiltration and aeration within the soil. Casts are produced by ingestion of mineral and organic particles, mixing, organic enrichment, and microbial stimulation in the gut, and egestion of the material as a slurry or as discrete pellets (depending on earthworm species), which harden into stable aggregates. Mechanisms of cast stabilization include organic bonding of particles by polymers secreted by earthworms and microbes, mechanical stabilization by plant fibers and fungal hyphae, and stabilization due to wetting and drying cycles and age-hardening/thixotrophic effects (Tomlin et al., 1995). Earthworm casts are usually enriched with plant available nutrients and thus may enhance soil fertility.

Earthworms create burrows of various sizes, depths, and orientations, depending on species and soil type. Burrows tend to be similar in diameter to that of the body, ranging from 1 to >10 mm diameter and constituting among the largest of soil pores (Lee, 1985; Tomlin et al., 1995; Shipitalo and Le Bayon, 2004). Geophagous species (Table 25.15) may form networks of variously oriented macropores, as the earthworms consume the soil and cast behind them as they burrow. Although such networks may form continuous pores for some depth, casting within the burrows may impede free water movement. Anecic earthworms may create vertical burrows that can form continuous macropores to depths of >1 m. Such burrows are often highly stable because their walls are lined with organic matter drawn in or secreted by earthworms, and they tend to have higher bulk density than surrounding soil. Continuous macropores resulting from earthworm burrowing may greatly enhance water infiltration by functioning as by-pass flow pathways through saturated soils (Lee, 1985; Tomlin et al., 1995; Shipitalo and Le Bayon, 2004). These pores may or may not be important in solute transport depending on antecedent soil water, nature of the solute, and exchange properties of the burrow linings (Edwards et al., 1990, 1993).

25.6.5 Methods

25.6.5.1 Earthworm Collection

Techniques for field sampling of earthworms are reviewed in Lee (1985) and Edwards and Bohlen (1996). Unless otherwise given, methodological details and specific reference citations can be found in these works. Collection techniques are passive, behavioral, and indirect (Table 25.16).

Hand digging and sorting, is the most commonly used and probably the most reliable method for quantitative sampling of earthworms. The technique involves digging pits of known dimensions (e.g., $25 \times 25 \times 25$ cm), breaking the soil by hand, and collecting all earthworms found. Often the collected specimens are immediately preserved in 70% ethanol or 5% formalin for later counting and identification (see Fender and McKey-Fender, 1990; Schwert, 1990, for details of preservation and preparation of specimens for identification).

Washing and sieving is an elaboration of hand sorting, in that the soil is dispersed in water (or a dispersing agent), poured through a sieve or nest of sieves, and the earthworms and cocoons hand picked from the sieve contents. Mechanized approaches to washing and sieving are described by Bouché and Beugnot (1972). Flotation of sieve contents in a high density solution, such as 1.16–1.20 SG $MgSO_4$, is an additional means of separating earthworms and other soil fauna from more dense soil particles.

A number of factors influence efficiency of hand sorting, including species and body size of earthworms (i.e., seasonal phenology and population demography), root density (especially in grasslands), soil type, and a "human factor," related to training of personnel and time spent on each sample (Schmidt, 2001a; Jiménez et al., 2006).

Several approaches have been taken to extracting earthworms from soil based on their behavioral response to certain stimuli. A number of chemical irritants have been used, including $HgCl_2$, $KMnO_4$, formalin and, more recently, mustard. Aqueous solutions of 0.165%–0.550% formalin have been used commonly and shown to be effective on *L. terrestris* when applied in three sequential doses of 18 L m^{-2} but formalin may be less effective on other species (Satchell, 1969; Callaham and Hendrix, 1997; Schmidt, 2001b). Aqueous extract of mustard has shown earthworm extraction efficiency similar to that of other chemical extractants and has come into favor because of its minimal effects on human health and low phytotoxicity compared to formalin (Gunn, 1992). Effectiveness of any chemical extractant varies with earthworm species and activity, temperature, soil porosity, and soil water content, saturated soils being less likely to transmit extractant solutions deep into the soil. Comparisons with hand sorting should be done before adopting chemical extraction techniques for quantitative sampling.

TABLE 25.16 Descriptions of Methods for Collecting Earthworms

Method	Description	Advantages	Disadvantages
Passive			
Hand sorting	Known volume of soil cut with spade or corer, broken and worms removed by hand	Simple, reliable in the field; low cost	Laborious; may not collect deep burrowing species, small earthworms and cocoons
Washing and sieving	Known volume of soil cut with spade or corer, soaked in dispersant/preservative, and washed through sieve(s) by hand or mechanical device	Higher recovery of cocoons and small individuals	Laborious; may not collect deep burrowing species
Flotation	Material from hand sorting or washing/sieving floated in high density solution (e.g., $MgSO_4$)	Separates earthworms from soil and plant debris; cocoons and small individuals collected	Laborious; may not collect deep burrowing species
Behavioral			
Chemical extraction	Soil saturated with chemical irritant (e.g., 0.2% formalin) causing earthworms to emerge onto soil surface	Simple; effective on deep burrowing anecic species	Not effective on all species, in all soils or under all conditions
Heat extraction	Soil blocks or cores suspended under heat lamps in water into which earthworms migrate	Effective on dense root mats	Not effective on all species; inconvenient for field use
Electrical extraction	Metal rods inserted into soil and connected to AC electrical source	Useful for selective or comparative sampling	Highly variable; not convenient in the field; dangerous
Mechanical vibration	Stake or rod inserted into soil and vibrated with bow or flat iron	Simple; useful for selective or comparative sampling	Not effective on all species
Trapping	Pitfall or baited traps placed in soil and sampled at desired intervals	Simple; useful for selective or comparative sampling	Not effective on all species
Mark–recapture	Individuals tagged, released, and population sampled at intervals	Useful for estimating population density, dispersal, and mortality	Laborious
Indirect			
Cast counting	Surface castings enumerated and identified	Simple	Not a quantitative estimate of population density

Sources: Summarized from Lee, K.E. 1985. Earthworms: Their ecology and relationships with soils and land use. Academic Press, Sydney, Australia. Edwards, C.A., and P.J. Bohlen. 1996. Biology and ecology of earthworms. 3rd edn. Chapman and Hall, New York.

The heat extraction method is a modification of that used for enchytraeids (see previous section). Soil cores or blocks are placed in pans of water, exposed to heat from overhead light-bulbs, and earthworms are collected from the water after several hours. This technique was more effective than hand sorting or formalin extraction on small earthworms in dense root mats (Satchell, 1969). As with hand sorting, it is not effective on deep burrowing, anecic species such as *L. terrestris*.

Mechanical vibration employs a rod or wooden stake driven into the soil, vibration for a few minutes with a bow or flat piece of metal (e.g., an automobile leaf spring), and collection of earthworms that emerge onto the soil surface. *Diplocardia mississippiensis* are routinely collected by fishermen in north Florida, with this method (termed "grunting") and Hendrix et al. (1994) used it for sampling *Diplocardia* populations in that region. Mitra et al. (2008) analyzed seismic signal characteristics of vibrations generated by the technique and found that numbers of worms emerging were positively correlated with signal strength. In general, vibration techniques may not be effective on many groups of earthworms, for example, lumbricids (Reynolds, 1973) and are probably not suited to quantitative measurements of population density. However, they may be useful for selective or comparative sampling of certain earthworm populations.

Electrical extraction of earthworms involves inserting metal rods into the soil, connecting them to a source of alternating current and collecting earthworms that come to the soil surface. Different voltages and amperages have been used with varying degrees of success; effectiveness of the technique is highly dependent on soil water content, electrolyte concentration, and temperature (Lee, 1985). As with mechanical vibration, the soil volume sampled is not known and therefore this method has been considered best suited for qualitative or comparative sampling. However, a comparative analysis by Schmidt (2001b) found that the "octet method" of Thielemann (1986) extracted higher earthworm numbers than the formalin method and gave community size and species composition estimates comparable to that of the hand sorting method. Schmidt (2001b) concluded that the octet method may be reliable and especially useful in situations requiring minimal soil disturbance. It must be cautioned that all electrical methods are potentially very dangerous and should only be used with extreme care.

Two earthworm-trapping techniques have been described. Pitfall traps (open-top containers buried level with the soil surface and containing a fixative solution, such as picric acid; see Section 25.4) may be useful for sampling surface-active species in diurnal or seasonal studies (Callaham et al., 2003b). Arrays of

traps are installed and sampled at 12, 24 h, or longer intervals. Baited traps, such as perforated clay pots containing manure or other attractants and inserted into the soil may also be useful for collecting certain species. As with other behavioral methods, trapping is probably highly selective and best suited for qualitative or comparative sampling.

Mark, release, and recapture techniques have been widely used to study population dynamics of animals, including earthworms. Large numbers of individuals of desired species are collected, marked (e.g., with brands or nontoxic dyes), and released into the population of interest. Sampling over time and distance from the target site, and enumeration of tagged relative to untagged individuals, yields information on dispersal, mortality, and population density. Dyes (Mazaud and Bouché, 1980) and radioisotopes (Bastardie et al., 2003) have been employed to mark earthworms. González et al. (2006) recently evaluated the use of a fluorescent elastomer injected into *Pontoscolex corethrurus* and were able to trace the marker for four months in populations incubated in field enclosures. Likewise, other studies (Butt and Lowe, 2006; Butt et al., 2009) showed that these elastomers did not affect the growth rates or cocoon production in *L. terrestris* and could be detected in the coelomic cavities of seven different earthworm species for up to 2 years (but sometimes was detected only after dissection). For earthworm species that cast on the soil surface, such as *Aporrectodea longa*, numbers and identity of castings may be a useful index of population activity. Because casting is dependent on soil temperature and moisture, this technique is highly variable and not a quantitative estimate of population density.

In addition to measurements of earthworm density (i.e., numbers per unit area), it is desirable to have estimates of earthworm biomass in most quantitative studies. Specimens from field collections contain varying amounts of gut contents that affect body mass measurements and therefore it is necessary to remove this material either by allowing live earthworms to void their guts (e.g., overnight in moist paper towels) or by dissecting and removing gut contents from preserved specimens. "Fresh weight" or "wet weight" measurements face two further problems: First, earthworms under field conditions contain varying amounts of water, and second, preserved specimens lose differing amounts of fresh mass in different preservative fluids (e.g., profuse mucus secretions in formalin; Lee, 1985). These factors must be considered in comparative studies across sites or seasons, or when different preservations techniques are used. For most purposes, it is best to convert live weight into dry weight, which may be accomplished by directly drying gut-voided specimens (e.g., freeze drying) or by using allometric relationships or regressions that relate body length or wet weight to dry weight (Lee, 1985; Hale et al., 2004), or oven dried weight to ash-free dry weight (Callaham et al., 2003a). Isotopic studies of earthworm feeding ecology require measurement of carbon mass, while also maintaining specimens for taxonomic identification. Live specimens can be killed instantly in boiling water and divided into a posterior half for gut clearance and freeze drying for chemical analysis, and an anterior half for preservation in formalin or ethanol for morphological taxonomic study (e.g., Hendrix et al., 1999).

Finally, for those interested in molecular techniques for evaluating the presence of earthworms in environmental samples, new techniques are rapidly being developed for detection of earthworm DNA or other proteins, and these hold great promise for rapid analysis of samples with high resolution information for diversity of organisms in samples. These techniques include using whole soil DNA extractions followed by sequencing and comparison against clone libraries (as in Wu et al., 2009), as well as more specific applications such as those in Juen and Traugott (2006) where DNA was used to identify the different prey items (including earthworms) in the gut of a soil predator.

In summary, digging and hand sorting or washing are probably the most reliable means of sampling earthworms. However, no single method will be adequate to sample earthworm populations in all situations. Combinations of methods will probably achieve reasonable results. For example, formalin or mustard solution can be applied to the bottom of pits previously excavated for hand sorting, to extract deep burrowing anecic forms not sampled by digging (Edwards and Bohlen, 1996). Combinations of various methods may be useful in other situations.

25.6.5.2 Identification

Many earthworm species in the family Lumbricidae can be identified from external body characteristics if the specimens are sexually mature. Taxonomic keys by Reynolds (1977), Schwert (1990), and Sims and Gerard (1999) are useful for the common lumbricids found worldwide. Reynolds (1977) also includes a key to North American Sparganophilidae, a limicolous or semi-aquatic group.

Most earthworms other than the Lumbricidae require dissection for accurate taxonomic identification. Knowledge of position and characteristics of sexual organs, the gut and associated glands, and other structures is required. The procedures must be done carefully and require a degree of skill and practice. At the family level, several keys are available: Jamieson (1988, 2000) contain a key and diagrammatic comparison of characteristic internal structures of most families, including the Megascolecidae of Australia; Edwards and Bohlen (1996) review major characteristics of the families; Sims and Gerard (1999) provide keys and species descriptions for seven families found in Great Britain; Fender and McKey-Fender (1990) give keys to the families of North America and the genera of Megascolecidae from western North America; James (1990) provides a key to the genus *Diplocardia*, a group of megascolecids found mostly in eastern North America, including, Mexico and the Caribbean. Righi (1971) describes the Glossoscolecidae in Central and South America. Many of the works cited in Gates (1982) include keys and species descriptions for families and genera found in North and Central America.

The taxonomy and systematics of earthworms, as with many other taxa, is currently undergoing considerable revision as advances in molecular biology prompt reevaluation of accepted

phylogenies and taxonomic relationships (e.g., Jamieson et al., 2002; Marotta et al., 2008). Although these approaches have great promise for providing clarity to taxonomic problems that have plagued earthworm nomenclature for decades, it is clear that much work remains and that even with DNA-based approaches, some controversy will remain (see Chang et al., 2008). Internet resources provide the best means of keeping track of these revisions and the development of new taxonomic keys (e.g., http://zipcodezoo.com/animals, http://www.discoverlife.org, http://species.wikimedia.org/wiki/Main_Page, and http://tolweb.org/tree).

References

Barois, I., M. Brossard, P. Lavelle, J. Tondoh, J. Kanyonyo, A. Martínez, J. Jiménez et al. 1999. Ecology of earthworm species with large environmental tolerance and/or extended distribution, p. 57–85. *In* P. Lavelle, L. Brousssard, and P. Hendrix (eds.) Earthworms management in tropical agroecosystems. CABI Publishing, New York.

Bastardie, F., Y. Capowiez, and D. Cluzeau. 2003. Burrowing behaviour of radio-labeled earthworms revealed by analysis of 3D-trajectories in artificial soil cores. Pedobiologia 7:554–559.

Beare, M.H., D.C. Coleman, D.A. Crossley, Jr., P.F. Hendrix, and E.P. Odum. 1995. A hierarchical approach to evaluating the significance of soil biodiversity to biogeochemical cycling. Plant Soil 170:5–22.

Boetcher, S.E., and P.J. Kalisz. 1991. Single-tree influence on earthworms in forest soils in eastern Kentucky. Soil Sci. Soc. Am. J. 55:862–865.

Bohlen, P.J., S. Scheu, C.M. Hale, M.A. McLean, S. Migge, P.M. Groffman, and D. Parkinson. 2004. Non-native invasive earthworms as agents of change in northern temperate forests. Front. Ecol. Environ. 2:427–435.

Bouché, M.B. 1977. Stratégies lombriciennes. *In* U. Lohm and T. Persson (eds.) Soil organisms as components of ecosystems. Ecol. Bull. (Stockholm) 25:122–133.

Bouché, M.B., and M. Beugnot. 1972. Contribution à l'approche méthodologique de l'étude des biocénoses. II. L'extraction des macroéléments du sol par lavage-tamisage. Annu. Zool. Ecol. Anim. 4:537–544.

Briones, M.J.I., P. Morán, and D. Posada. 2009. Are the sexual, somatic and genetic characters enough to solve nomenclatural problems in lumbricid taxonomy? Soil Biol. Biochem. 41:2257–2271.

Butt, K.R., M.J.I. Briones, and C.N. Lowe. 2009. Is tagging with visual implant elastomer a reliable technique for marking earthworms? Pesq. Agropec. Bras. 44:969–974.

Butt, K.R., and C.N. Lowe. 2006. A viable technique for tagging earthworms using visible implant elastomer. Appl. Soil Ecol. 35:454–457.

Callaham, M.A., Jr., J.M. Blair, T.C. Todd, D.J. Kitchen, and M.R. Whiles. 2003a. Macroinvertebrates in North American tallgrass prairie soils: Effects of fire, mowing, and fertilization on density and biomass. Soil Biol. Biochem. 35:1079–1093.

Callaham, M.A., Jr., and P.F. Hendrix. 1997. Relative abundance and seasonal activity of earthworms (Lumbricidae and Megascolecidae) as determined by hand-sorting and formalin extraction in forest soils on the southern Appalachian Piedmont. Soil Biol. Biochem. 29:317–322.

Callaham, M.A., Jr., P.F. Hendrix, and R.J. Phillips. 2003b. Occurrence of an exotic earthworm (*Amynthas agrestis*) in undisturbed soils of the southern Appalachian Mountains, USA. Pedobiologia 47:466–470.

Chang, C.-H., S.-M. Lin, and J.-H. Chen. 2008. Molecular systematics and phylogeography of the gigantic earthworms of the *Metaphire formosae* species group (Clitellata, Megascolecidae). Mol. Phylogenet. Evol. 49:958–968.

Chang, C.-H., R. Rougerie, and J.-H. Chen. 2009. Identifying earthworms through DNA barcodes: Pitfalls and promise. Pedobiologia 52:171–180.

Chapuis-Lardy, L., R.S. Ramiandrisoa, L. Randriamanantsoa, C. Morel, L. Rabeharisoa, and E. Blanchart. 2009. Modification of P availability by endogeic earthworms (Glossoscolecidae) in ferralsols of the Malagasy Highlands. Biol. Fertil. Soils 45:415–422.

Curry, J.P. 1994. Grassland invertebrates. Ecology, influence on soil fertility and effects on plant growth. Chapman and Hall, London, U.K.

Curry, J.P. 2004. Factors affecting the abundance of earthworms in soils, p. 91–113. *In* C. Edwards (ed.) Earthworm ecology. CRC Press, Boca Raton, FL.

Darwin, C.R. 1881. The formation of vegetable mould through the action of worms, with observations on their habits. Murray, London, U.K.

Dindal, D.L. (ed.). 1990. Soil biology guide. John Wiley & Sons, New York.

Edwards, C.A. 1983. Earthworm ecology in cultivated soil, p. 123–138. *In* J.E. Satchell (ed.) Earthworm ecology from Darwin to vermiculture. Chapman and Hall, London, U.K.

Edwards, C.A. (ed.). 2004. Earthworm ecology. CRC Press, Boca Raton, FL.

Edwards, C.A., and P.J. Bohlen. 1996. Biology and ecology of earthworms. 3rd edn. Chapman and Hall, New York.

Edwards, W.M., M.J. Shipitalo, L.B. Owens, and W.A. Dick. 1993. Factors affecting preferential flow of water and atrazine through earthworm burrows under continuous no-till corn. J. Environ. Qual. 22:453–457.

Edwards, W.M., M.J. Shipitalo, L.B. Owens, and L.D. Norton. 1990. Effect of *Lumbricus terrestris* L. burrows on hydrology of continuous no-till corn fields. Geoderma 46:73–84.

Fender, W.F., and D. McKey-Fender. 1990. Oligochaeta: Megascolecidae and other earthworms from Western North America, p. 357–378. *In* D.L. Dindal (ed.) Soil biology guide. John Wiley & Sons, New York.

Fragoso, C., J. Kanyonyo, A.L. Moreno, B.K. Senapati, E. Blanchart, and C. Rodriguez. 1999. A survey of tropical earthworms: Taxonomy, biogeography and environmental

plasticity, p. 1–26. *In* P. Lavelle, L. Brussaard, and P. Hendrix (eds.) Earthworm management in tropical agroecosystems. CABI Publishing, New York.

Gates, G.E. 1972. Burmese earthworms: An introduction to the systematics and biology of megadrile oligochaetes with special reference to Southeast Asia. Trans. Am. Philos. Soc. 62:1–326.

Gates, G.E. 1982. Farewell to North America megadriles. Megadrilogica 4:12–77.

González, G., E. Espinosa, Z. Liu, and X. Zou. 2006. A fluorescent marking and re-count technique using the invasive earthworm, *Pontoscolex corethrurus* (Annelida: Oligochaeta). Carib. J. Sci. 42:371–379.

Green, R.N., R.L. Trowbridge, and K. Klinka. 1993. Towards a taxonomic classification of humus forms. Forest Science Monograph 29. Society of American Foresters, Bethesda, MD.

Gunn, A. 1992. The use of mustard to estimate earthworm populations. Pedobiologia 36:65–67.

Hale, C., P. Reich, and L. Frelich. 2004. Allometric equations for estimation of ash-free dry mass from length measurements for selected European earthworm species (Lumbricidae) in the western Great Lakes region. Am. Midl. Nat. 151:179–185.

Hendrix, P.F. (ed.). 1995. Earthworm ecology and biogeography in North America. Lewis Publisher, Boca Raton, FL.

Hendrix, P.F. (ed.). 2006. Biological invasions belowground: Earthworms as invasive species. Springer-Verlag, Amsterdam, the Netherlands.

Hendrix, P.F., and P.J. Bohlen. 2002. Exotic earthworm invasions in North America: Ecological and policy implications. Bioscience 52:801–811.

Hendrix, P., M. Callaham, Jr., J. Drake, C. Huang, S. James, B. Snyder, and W. Zhang. 2008. Pandora's box contained bait: The global problem of introduced earthworms. Annu. Rev. Ecol. Evol. Syst. 39:593–613.

Hendrix, P.F., M.A. Callaham, Jr., and L. Kirn. 1994. Ecological studies of nearctic earthworms in the southern USA. II. Effects of bait harvesting on *Diplocardia* populations in Apalachicola National Forest in north Florida. Megadrilogica 5:73–76.

Hendrix, P.F., S.L. Lachnicht, M.A. Callaham, Jr., and X. Zou. 1999. Stable isotopic studies of earthworm feeding ecology in tropical ecosystems of Puerto Rico. Rapid Commun. Mass Spectrom. 13:1295–1299.

Hendrix, P.F., B.R. Mueller, R.R. Bruce, G.W. Langdale, and R.W. Parmelee. 1992. Abundance and distribution of earthworms in relation to landscape factors on the Georgia Piedmont, USA. Soil Biol. Biochem. 24:1357–1361.

James, S.W. 1990. Oligochaeta: Megascolecidae and other earthworms from Southern and Midwestern North America, p. 379–386. *In* D.L. Dindal (ed.) Soil biology guide. Wiley, New York.

James, S.W. 1995. Systematics, biogeography and ecology of nearctic earthworms from eastern, central, southern and southwestern United States, p. 29–52. *In* P.F. Hendrix (ed.) Earthworm ecology and biogeography in North America. Lewis Publishers, Boca Raton, FL.

Jamieson, B.G.M. 1981. Historical biogeography of the Australian Oligochaeta, p. 887–921. *In* A. Keast (ed.) Ecological biogeography of Australia. W. Junk, The Hague, the Netherlands.

Jamieson, B.G.M. 1988. On the phylogeny and higher classification of the Oligochaeta. Cladistics 4:367–401.

Jamieson, B.G.M. 2000. Native earthworms of Australia (Megascolecidae, Megascolecinae). Science Publishers, Enfield, NH.

Jamieson, B.G.M. 2006. Non-leech clitellata. With contributions by M. Ferraguti, p. 235–392. *In* G. Rouse and F. Pleijel (eds.) Reproductive biology and phylogeny of annelida. Science Publishers, Enfield, NH.

Jamieson, B.G.M., S. Tillier, A. Tillier, J.-L. Justine, E. Ling et al. 2002. Phylogeny of the Megascolecidae and Crassiclitellata (Annelida, Oligochaeta): Combined versus partitioned analysis using nuclear (28S) and mitochondrial (12S, 16S) rDNA. Zoosystema 24:707–734.

Jiménez, J.J., P. Lavelle, and T. Decaëns. 2006. The efficiency of soil hand-sorting in assessing the abundance and biomass of earthworm communities. Its usefulness in population dynamics and cohort analysis studies. Eur. J. Soil Biol. 42:S225–S230.

Juen, A., and M. Traugott. 2006. Amplification facilitators and multiplex PCR: Tools to overcome PCR-inhibition in DNA-gut-content analysis of soil-living invertebrates. Soil Biol. Biochem. 38:1872–1879.

Kalisz, P.J., and H.B. Wood. 1995. Native and exotic earthworms in wildland ecosystems, p. 117–126. *In* P.F. Hendrix (ed.) Earthworm ecology and biogeography in North America. Lewis Publishers, Boca Raton, FL.

Lavelle, P. 1983. The structure of earthworm communities, p. 449–466. *In* J.E. Satchell (ed.) Earthworm ecology: From Darwin to vermiculture. Chapman and Hall, London, U.K.

Lavelle, P., L. Brussaard, and P. Hendrix (eds.). 1999. Earthworm management in tropical agroecosystems. CABI, New York.

Lavelle, P., and E. Lapied. 2003. Endangered earthworms of Amazonia: An homage to Gilberto Righi. Pedobiologia 47:419–427.

Lavelle, P., and A. Spain. 2001. Soil ecology. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Lee, K.E. 1959. The earthworm fauna of New Zealand. Department of Scientific and Industrial Research Bulletin No. 130. Wellington, New Zealand.

Lee, K.E. 1961. Interactions between native and introduced earthworms. Proc. N.Z. Ecol. Soc. 8:60–62.

Lee, K.E. 1985. Earthworms: Their ecology and relationships with soils and land use. Academic Press, Sydney, Australia.

Lee, K.E. 1991. The diversity of soil organisms, p. 73–87. *In* D.L. Hawksworth (ed.) The biodiversity of microorganisms and invertebrates: Its role in sustainable agriculture. CABI Publishing, New York.

Lee, K.E. 1995. Earthworms and sustainable land use, p. 215–234. *In* P.F. Hendrix (ed.) Earthworm ecology and biogeography in North America. Lewis Publishers, Boca Raton, FL.

Linnaeus, C. (1758). *Systema naturae* per regna tria naturae, Secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis. Locis. Tomus I. Editio decima, reformata. Editio decima revisa. Vol. 1 pp. [i–iv], [1]-823. Holmiae [Stockholm]: impensis direct. Laurentii Salvii.

Marotta, R., M. Ferraguti, C. Erseus, and L.M. Gustavsson. 2008. Combined-data phylogenetics and character evolution of Clitellata (Annelida) using 18S rDNA and morphology. Zool. J. Linn. Soc. 154:1–26.

Mazaud, D., and M. Bouché. 1980. Introductions sur population et migrations of lombriciens marques, p. 687–701. *In* D. Dindal (ed.) Soil biology as related to land use practices. Proc. 7th Intl Soil Zool. Coll., Syracuse, NY. EPA, Washington, DC. 29 July–3 August, 1979.

Mitra, O., M.A. Callaham, Jr., M.L. Smith, and J.E. Yack. 2008. Grunting for worms: Seismic vibrations cause *Diplocardia* earthworms to emerge from the soil. Biol. Lett. 5:16–19.

Omodeo, P. 2000. Evolution and biogeography of megadriles (Annelida: Clitellata). Ital. J. Zool. 67:179–201.

Pawlett, M., D.W. Hopkins, B.F. Moffett, and J.A. Harris. 2009. The effect of earthworms and liming on soil microbial communities. Biol. Fertil. Soils 45:361–369.

Reynolds, J.W. 1973. Earthworm (Annelida: Oligochaeta) ecology and systematics, p. 95–120. *In* D.L. Dindal (ed.) Proc. 1st Soil Microcommun. Conf., Syracuse, NY. U.S. Atomic Energy Commission, Washington, DC.

Reynolds, J.W. 1977. The earthworms (Lumbricidae and Sparganophilidae) of Ontario. Royal Ontario Museum Life Sciences Miscellaneous Publication, Toronto, Canada.

Reynolds, J.W. 1994. Earthworms of the world. Global Biodivers. 4:11–16.

Reynolds, J.W. 1995. Status of exotic earthworm systematics and biogeography in North America, p. 1–28. *In* P.F. Hendrix (ed.) Earthworm ecology and biogeography in North America. Lewis Publishers, Boca Raton, FL.

Reynolds, J.W., and D.G. Cook. 1993. Nomenclatura Oligochaetologica, Supplementum Tertium: A catalogue of names, descriptions and type specimens of the Oligochaeta. New Brunswick Museum Monograph Series No. 9. (Nat. Sci.). New Brunswick, NJ.

Righi, G. 1971. Sobre a familia Glossoscolecidae (Oligochaeta) no Brasil. Arquivos de Zoología S. Paulo 20:1–95.

Rossi, J.-P., H. Huerta, C. Fragoso, and P. Lavelle. 2006. Soil properties inside earthworm patches and gaps in a tropical grassland (la Mancha, Veracruz, Mexico). Eur. J. Soil Biol. 42:S284–S288.

Satchell, J.E. 1969. Studies on methodical and taxonomical questions. Pedobiologia 9:20–25.

Satchell, J.E. (ed.). 1983. Earthworm ecology: From Darwin to vermiculture. Chapman and Hall, London, U.K.

Schmidt, O. 2001a. Time-limited soil sorting for long-term monitoring of earthworm populations. Pedobiologia 45:69–83.

Schmidt, O. 2001b. Appraisal of the electrical octet method for estimating earthworm populations in arable land. Annu. Appl. Biol. 138:231–241.

Schwert, D.P. 1990. Oligochaeta: Lumbricidae, p. 341–356. *In* D.L. Dindal (ed.) Soil biology guide. John Wiley & Sons, New York.

Shipitalo, M.J., and R.-C. Le Bayon. 2004. Quantifying the effects of earthworms on soil aggregation and porosity, p. 183–200. *In* C.A. Edwards (ed.) Earthworm ecology. CRC Press, Boca Raton, FL.

Sims, R.W. 1980. A classification and the distribution of earthworms, suborder Lumbricina (Haplotaxida: Oligochaeta). Bull. Br. Mus. Nat. Hist. Zool. 39:103–124.

Sims, R.W., and B.M. Gerard. 1999. Earthworms. Notes for the identification of British species. Synop. Br. Fauna new Ser. 31:1–169.

Speratti, A.B., J.K. Whalen, and P. Rochette. 2007. Earthworm influence on carbon dioxide and nitrous oxide fluxes from an unfertilized corn agroecosystem. Biol. Fertil. Soils 44:405–410.

Svensson, K., and H. Friberg. 2007. Changes in active microbial biomass by earthworms and grass amendments in agricultural soil. Biol. Fertil. Soils 44:223–228.

Thielemann, U. 1986. Elektrischer Regenwurmfang mit der Oktett-Methode. Pedobiologia 29:296–302.

Tomlin, A., M. Shipitalo, W. Edwards, and R. Protz. 1995. Earthworms and their influence on soil structure and infiltration, p. 159–184. *In* P.F. Hendrix (ed.) Earthworm ecology and biogeography in North America. Lewis Publishers, Boca Raton, FL.

Wallwork, J.A. 1983. Earthworm biology. Studies in biology No. 161, Institute of Biology, Camelot Press, Southampton, U.K.

Wardle, D. 2002. Communities and ecosystems: Linking aboveground and belowground components. Princeton University Press, Princeton, NJ.

Wu, T., E. Ayres, G. Li, R.D. Bardgett, D.H. Wall, and J.R. Garey. 2009. Molecular profiling of soil animal diversity in natural ecosystems: Incongruence of molecular and morphological results. Soil Biol. Biochem. 41:849–857.

26

Microbially Mediated Processes

Susumu Asakawa
Nagoya University

Else K. Bünemann
*Eidgenössische Technische
Hochschule Zürich*

Emmanuel Frossard
*Eidgenössische Technische
Hochschule Zürich*

E.G. Gregorich
Agriculture and Agri-Food Canada

Jan Jansa
*Eidgenössische Technische
Hochschule Zürich*

H.H. Janzen
Agriculture and Agri-Food Canada

Michael A. Kertesz
The University of Sydney

Makoto Kimura
Nagoya University

Loretta Landi
University of Florence

David Long
Michigan State University

Terence L. Marsh
Michigan State University

Paolo Nannipieri
University of Florence

Astrid Oberson
*Eidgenössische Technische
Hochschule Zürich*

Giancarlo Renella
University of Florence

Thomas Voice
Michigan State University

26.1 Phosphorus and Sulfur in Soil

Emmanuel Frossard

Else K. Bünemann

Astrid Oberson

Jan Jansa

Michael A. Kertesz

26.1.1 Introduction

26.1.1.1 Phosphorus and Sulfur as Indispensable Elements for Life

Both phosphorus (P) and sulfur (S) are indispensable for living organisms. Phosphorus is contained in DNA and RNA and as such is important with respect to storing and reading genetic information. Phosphorus is also found in ATP and its derivatives, and is vital for energy storage and transfer in the cell. Finally, phosphorus is present in phospholipids that are important components of cell membranes.

Sulfur is part of some amino acids (cysteine and methionine) and of proteins. The presence of cysteine in proteins allows the formation of disulfide bonds, which are involved in the tertiary structure of proteins. Sulfur plays an important role in enzymes as a metal-binding element. It is present in the coenzyme A that is a key intermediate in lipid synthesis and in the vitamins biotin and thiamin. Sulfur is also contained in specific molecules such as glutathione that play an important role for the transport and detoxification of metals or pesticides in the cell.

Given their importance for plants, phosphorus and sulfur are widely used as fertilizers in agro-ecosystems. However, mineable reserves of phosphorus are finite and might be sufficient for little more than 100 years (Cordell et al., 2009). Thereafter, phosphorus fertilizer prices will increase strongly. For sulfur, the reserves are probably much larger. The sulfur removed from fuel supplies and smoke stacks provides an extensive reserve.

26.1.1.2 The Cycles of Phosphorus and Sulfur in Soils

The concentrations of total phosphorus and sulfur in soil range between less than 100 and more than 1000 mg kg^{-1}. Young soils located on phosphorus- or sulfur-rich parent material present higher total phosphorus and sulfur contents (>1000 mg kg^{-1}) while strongly weathered soils developed on a parent material with low phosphorus and sulfur contents can have very low total phosphorus and sulfur contents (<100 mg kg^{-1}).

Most of the soil phosphorus is present as phosphate. Soils contain between 20% and 80% of the total phosphorus in organic forms. The following forms of organic phosphorus can be found (in the order of decreasing abundance): monoester phosphates, such as phytate (*myo*-inositol hexakisphosphate), with the

general formula $R-O-PO_3^{2-}$ (R being an organic moiety); diester phosphates, such as nucleic acids, with the general formula $R-O-(PO_2^-)-O-R'$ (R and R′ being two organic moieties); and phosphonates with the general formula $R-C-PO_3^{2-}$ (R being an organic moiety). Finally, a small proportion of soil phosphorus can be found as pyro- and polyphosphate $((PO_3)_n)$, which are not organic molecules but can be of biological origin (Bünemann and Condron, 2007).

All the oxidation states of sulfur can be observed in soil depending on its redox status. In aerobic soils, 90% of the total sulfur is present in organic forms. The most important organic forms are in a strongly reduced state (e.g., sulfide and thiophene), in an intermediate oxidation state (sulfoxide and sulfonate), and in a strongly oxidized state (ester sulfate) (Solomon et al., 2003).

Both elements can enter the soil through rock weathering, atmospheric inputs, erosion, and inorganic and organic fertilizers (Figure 26.1). Whereas phosphorus inputs from the atmosphere are limited, sulfur inputs in gaseous form can be significant in volcanic regions and in regions with high rates of atmospheric pollution. Since air pollution control programs have been implemented in Europe and North America, the input of sulfur from the atmosphere has greatly declined.

Both elements leave the soil through the exportation of harvested products from agricultural fields, erosion, runoff, and leaching. A major difference between the phosphorus and sulfur cycles is the sometimes important loss of sulfur from soils in gaseous forms (e.g., as hydrogen sulfide or sulfur dioxide) whereas gaseous losses of phosphorus (e.g., as phosphine) are negligible.

Phosphorus in the form of orthophosphate (HPO_4^{2-} and $H_2PO_4^-$) and sulfur in the form of sulfate (SO_4^{2-}) can react with soil particles through sorption/desorption and precipitation/dissolution mechanisms. These species can also be taken up by plants and microorganisms and incorporated into stable organic compounds. Conversely, microbial activity can also trigger the release of phosphorus and sulfur compounds to the soil solution.

In most cases, abiotic and biotic reactions maintain the concentrations of sulfate and phosphate in the soil solution at a low level. This is why plants and microorganisms have developed efficient mechanisms to acquire these elements.

This short review of the phosphorus and sulfur cycles in soils covers some recent developments in analytical approaches to study pools and fluxes of these elements in soils (Section 26.1.2). Then we present the most important biotic (Section 26.1.3) reactions that these elements undergo. Finally, the factors and processes controlling phosphorus and sulfur forms and dynamics are presented for two exemplary soil systems (Section 26.1.4). The abiotic reactions that these elements undergo are not discussed here as they are presented elsewhere in this handbook. For more extensive reviews on these elements, the reader is referred to Sims and Sharpley (2005), Bünemann and Condron (2007), and Kertesz et al. (2007).
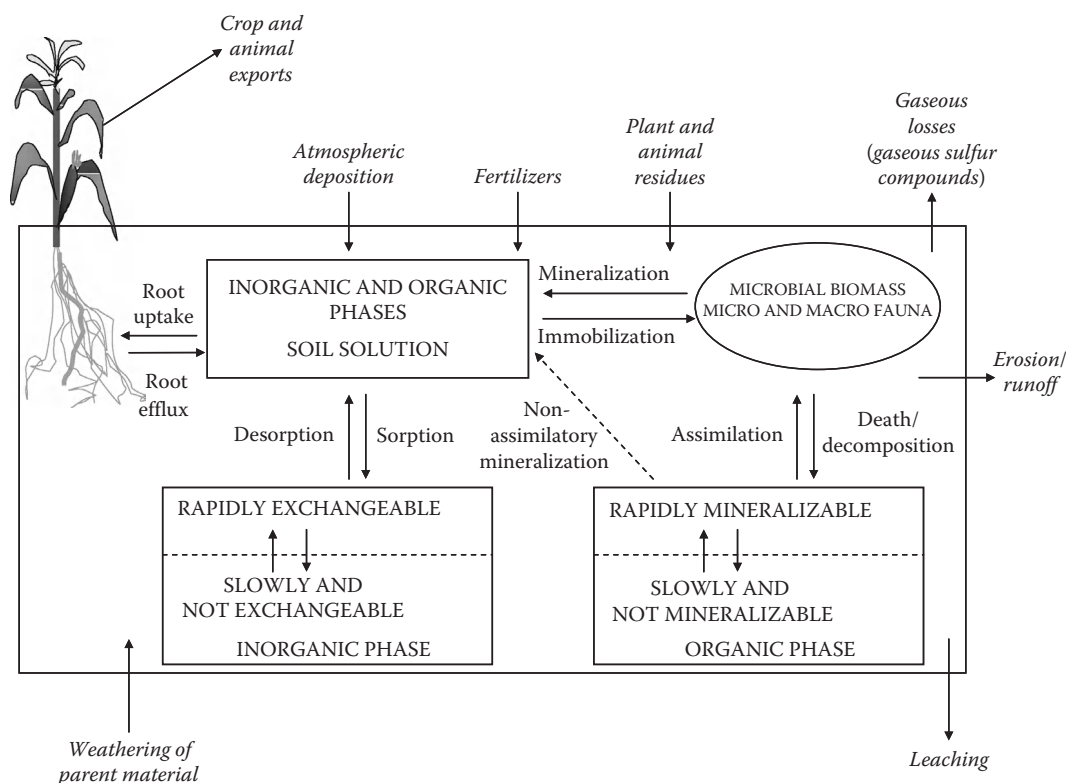
FIGURE 26.1 Schematic presentation of phosphorus and sulfur cycling in aerobic soils. Pools are written in capital letters, fluxes in the soil in small letters, and inputs to the soil and outputs from the soil in italics. (Adapted from McLaughlin, M.J., D.J. Reuter, and G.E. Rayment. 1999. Soil testing—Principles and concepts, p. 1–21. *In* K.I. Peverill, L.A. Sparrow, and D.J. Reuter (eds.) Soil analysis—An interpretation manual. The Australasian Soil and Plant Analysis Council Inc., CSIRO Publishing, Collingwood, Australia. Available online at http://www.publish.csiro.au/pid/1998.htm (accessed June 13, 2009). Published with the permission of the Australasian Soil and Plant Analysis Council Inc. and CSIRO Publishing.)

26.1.2 Methods of Investigation

26.1.2.1 Chemical/Biochemical Extractions

Chemical extractions have been extensively used to estimate the amount of phosphorus and sulfur present in different forms in soils. Dilute aqueous solutions (water or $CaCl_2$) can be used to extract the amount of plant-available phosphorus and sulfur. Extracting soils with resin strips or iron oxide paper allows an estimation of the fraction of phosphate that is sorbed and can be released to the solution. Plant extractable sulfur can be assessed by extraction with orthophosphate or with a resin.

Sequential extraction schemes have been developed for assessing phosphorus and sulfur forms in soils. For phosphorus (Hedley et al., 1982), the rapidly plant-available form is first removed from the soil with a resin. The sediment is then extracted with $NaHCO_3$ and then with NaOH. These two alkaline extractants solubilize phosphorus bound to aluminum and iron oxides and organic forms of phosphorus. The sediment is then extracted with HCl to dissolve phosphorus bound to calcium. Finally, an acid digestion can be conducted on the sediment to extract the residual phosphorus. For sulfur, a commonly used sequential extraction (Freney et al., 1975) begins with an extraction with orthophosphate, followed by

a digestion of the sediment with hydriodic acid. Ester sulfates are converted to hydrogen sulfide during this digestion while the sulfur remaining in the sediment is considered to be directly bound to carbon. The sediment can then be treated with Raney Ni to separate reducible sulfur (S in amino acids) from nonreducible sulfur forms. Sequential extractions give fair information on phosphorus and sulfur forms, but the amounts of extracted elements depend strongly on the chemical extractant that is used, and some artifacts can occur during the extraction (e.g., formation of new compounds during the extraction).

The fraction of phosphorus and sulfur stored in the microbial biomass can be estimated by comparing the amount of element extracted from a soil sample that has been fumigated with an organic solvent (e.g., chloroform, hexanol, and toluene) to that extracted from a nonfumigated sample (Saggar et al., 1981; Hedley and Stewart, 1982). The fumigation causes the lysis of microbial cells, releasing elements stored in the microbial biomass into the soil solution. This approach poses two types of problems: first, only a fraction of the microbially held element is released, and second, a fraction of the released phosphate is quickly sorbed onto soil surfaces. Whereas the correction for phosphate sorption is straightforward, it is more difficult to

correct for the incomplete extraction of the microbial phosphorus and sulfur (Oberson and Joner, 2005).

Recently, enzymatic approaches have been proposed to assess the availability and speciation of phosphorus and sulfur in organic forms in soils and solution. Hydrolases of the phosphorus or sulfur cycles are added in large amounts to soil extracts or soil suspensions, and the fraction of element released during the experiment is measured. On the one hand, the use of an enzyme cocktail able to hydrolyze a larger number of organic phosphorus compounds yields information on the total amount of phosphate that can potentially be released from organic compounds. On the other hand, the use of a specific enzyme will give precise information on the proportion of the specific form of organic phosphorus accessible to the enzyme. These enzymatic approaches have a lot of potential but need to be standardized before they can deliver results that can be correctly interpreted (Bünemann, 2008).

26.1.2.2 Spectroscopic Approaches

Two spectroscopic methods are gaining importance for the study of phosphorus and sulfur in soils: nuclear magnetic resonance (NMR) and x-ray absorption spectroscopy.

NMR is used to characterize the close chemical environment of nuclei that possess an odd number of protons and neutrons such as the ^{31}P nucleus. Two types of NMR experiments can be conducted on soils: liquid-state ^{31}P NMR and solid-state ^{31}P NMR. Liquid-state ^{31}P NMR experiments are usually conducted on NaOH–EDTA soil extracts. Earlier work had shown that this technique could allow quantitatively assessing the abundance of phosphonates, orthophosphate, monoester phosphates, diester phosphates, pyrophosphates, and polyphosphates in NaOH extracts (Condron et al., 1990); but assigning peaks to specific compounds was problematic. More recently, Turner et al. (2003a, 2003b) and Turner and Richardson (2004) proposed approaches to assign these peaks to DNA, RNA, phosphatidyl choline, phosphatidyl serine, phosphatidyl ethanolamine, polyphosphates, *scyllo*-inositol phosphates, and *myo*-inositol phosphates. Smernik and Dougherty (2007) however stress the need to carefully spike the soil extracts to obtain relevant concentrations of *myo*-inositol phosphates in soil extracts and to avoid artifacts due to changes in pH and ionic strength. This is important as *myo*-inositol phosphates, such as phytate, are considered to be the dominant pool of organic phosphorus in soils.

Soil phosphorus forms can also be studied with solid-state ^{31}P NMR. Experiments can be conducted in the cross-polarization (CP) or in the direct-polarization (DP) mode. Experiments done in CP allow the observation of ^{31}P nuclei located in the close vicinity of ^{1}H whereas experiments done in DP mode theoretically allow the observation of all ^{31}P nuclei. The interest of solid-state ^{31}P NMR is to obtain information on mineral and organic P forms present in a soil without chemical extraction. Using solid-state ^{31}P NMR, McDowell et al. (2002) report the presence of various aluminum and calcium phosphates (berlinite, wavellite, variscite, monetite, dicalcium phosphate dihydrate, hydroxyapatite,

and octocalcium phosphate) in the upper horizon of three soils containing high total P contents (>1400 mg P kg^{-1}). However, the use of this technique in soil faces a number of problems: the total phosphorus content is usually lower than 1000 mg P kg^{-1}, phosphorus is distributed in a large number of forms, and soils usually contain large amounts of iron and manganese, which hinder the observation of ^{31}P nuclei by NMR. Dougherty et al. (2005) using the technique of spin counting showed that only 9% of the total P could be observed by CP solid-state ^{31}P NMR and 22% by DP. They attributed this poor "observability" to the presence of Fe in the close vicinity of P nuclei. They conclude their paper by stating that considerable methodological work must still be carried out, considering "observability," chemical shift anisotropy, and relaxation rates, in order to obtain relevant information from solid-state NMR ^{31}P experiments.

The use of x-ray absorption spectroscopy for the study of soil phosphorus and sulfur developed during the last two decades. Most work on phosphorus and sulfur has been carried out with x-ray absorption near edge structure (XANES). Hesterberg et al. (1999) obtained phosphorus K edge XANES spectra from a number of mineral phosphorus forms, including phosphate bound to Fe, and showed the presence of calcium phosphate in a soil. Using phosphorus K edge XANES and sequential extraction, Beauchemin et al. (2003) showed the presence of phosphate adsorbed on iron and aluminum oxides and of calcium phosphates (hydroxyapatite and octocalcium P) in phosphorus-rich soils. Sato et al. (2005) observed phosphate bound to iron oxides in a forest soil and poorly crystallized calcium phosphates (dibasic calcium phosphate, amorphous calcium phosphate, and tricalcium phosphate) in soils that had received organic manures. Lombi et al. (2006) studied the distribution of phosphorus at the nanoscale in a highly calcareous soil with nano x-ray fluorescence and nano XANES, showing that phosphorus was heterogeneously distributed in this soil and that it was associated with calcium rather than iron. A big advantage of XANES compared to NMR is its ability to detect mineral phosphate forms bound to iron. A disadvantage of phosphorus K edge XANES is the lack of sensitivity of this technique for organic forms of phosphorus (Toor et al., 2006). A recent report from Kruse et al. (2009) shows however that the phosphorus $L_{2,3}$ edge XANES is richer in spectral features than the K edge XANES and therefore should allow for a more precise analysis of organic and mineral P compounds in soil. Finally, as in the case of solid-state ^{31}P NMR, XANES yields good results for soils with quite a high phosphorus content (>1000 mg P kg^{-1}).

The application of sulfur K edge XANES led to a significant leap forward in our knowledge of sulfur speciation in soils. Morra et al. (1997) showed that it was possible to differentiate sulfur forms in humic and fulvic acids according to their oxidation state. Solomon et al. (2003) could quantify with this technique the amount of sulfur present in a strongly reduced state (e.g., sulfide and thiophene), in an intermediate oxidation state (sulfoxide and sulfonate), and in a strongly oxidized state (ester sulfate) in the humic substances extracted from particle size separates of different soils. Although the sulfur K edge XANES spectra of nonextracted particle size separates could not be

quantitatively analyzed due to the high background, they were very similar to the spectra obtained from their humic extracts. The authors concluded that sulfur K edge XANES of humic substances represents adequately organic sulfur forms in total soil. Prietzel et al. (2007) on the contrary observed significant differences between the sulfur forms measured by XANES on humic and fulvic extracts, on size separates and on bulk soil samples. They suggest that sulfur speciation might have changed during the preparation of the humic extracts and that ester sulfate might be enriched in the clay fraction of soils. Prietzel et al. (2008) showed that sulfur K edge XANES could be used to distinguish adsorbed from precipitated sulfate in bulk soils and in microsites in a wide range of sulfate-rich soils.

26.1.2.3 Stable and Radioactive Isotopes

The most relevant phosphorus isotopes for the study of soil systems are ^{31}P, ^{32}P, and ^{33}P. Both ^{32}P and ^{33}P are radioactive, ^{32}P and ^{33}P have a short half-life (14.3 days for ^{32}P and 25.3 days for ^{33}P) and emit β^- radiation with a maximum of 0.25 and 1.71 MeV for ^{33}P and ^{32}P, respectively (Endt, 1990).

Labeling the soil solution with carrier-free radioactive orthophosphate ($^{32}PO_4^{3-}$ or $^{33}PO_4^{3-}$) at steady-state equilibrium for $^{31}PO_4^{3-}$ in solution, in the absence of carbon input, allows quantification of the amount of orthophosphate located on the solid phase that can be released into the solution over time, through isotopic exchange. The amount of isotopically exchangeable orthophosphate measured in a soil/solution system is called the E value. The study of the kinetics of isotopic exchange shows that most of the soil inorganic phosphate must be seen as isotopically exchangeable, but some ions (a few) appear very rapidly in the solution whereas other ions (the large majority) need a lot of time to be released into the soil solution (Frossard and Sinaj, 1997). The analysis of the specific activity (e.g., $^{33}PO_4^{3-}/^{31}PO_4^{3-}$) in the fractions of a sequential extraction gives information on the isotopic exchangeability of the phosphate present in the different fractions (Bühler et al., 2002).

The principle of isotopic exchange can be used to assess the basal mineralization of soil organic phosphorus. In biologically active soils that have been labeled with radioactive orthophosphate, the decrease of specific activity of orthophosphate in the soil solution can be related to the release of nonradioactive orthophosphate ($^{31}PO_4^{3-}$) to the solution through isotopic exchange and to the mineralization of $^{31}PO_4^{3-}$ from soil organic matter (Bünemann et al., 2007). Once the appropriate controls have been prepared to quantify the amount of $^{31}PO_4^{3-}$ released by isotopic exchange, it becomes possible to deduce the amount of phosphorus mineralized from soil organic matter. Furthermore, by measuring the concentration of radioactive phosphorus in the soil microbial biomass, it is possible to assess phosphorus immobilization (Bünemann et al., 2007).

The same principles can be applied to assess whether a growing plant can solubilize significant amounts of slowly exchangeable phosphorus to cover its needs. To do this, the fraction of soil phosphorus that is isotopically exchangeable within a given time is measured in a soil/solution system (E value) and compared to the amount of isotopically exchangeable phosphate measured over the same period with a plant growing in a labeled soil (L value). If L is higher than E, and provided that the pitfalls of both methods have been avoided (Pypers et al., 2006), then it can be concluded that the plant had access to a fraction of phosphorus that was initially not isotopically exchangeable, for example, through the root exudation of organic acids or phosphatase.

Finally, the use of phosphorus added as mineral or organic fertilizer by plants can be assessed either directly, when the fertilizer is homogeneously labeled with radioactive phosphate or, when the fertilizer cannot be homogeneously labeled by radioactive phosphate, indirectly by labeling soil-available phosphate with radioactive phosphate (Gallet et al., 2003).

The most relevant sulfur isotopes for the study of soils are ^{32}S and ^{34}S, which are both stable and ^{35}S, which is radioactive. About 95% of the total sulfur is present as ^{32}S while ^{34}S represents 4.21% of total sulfur.

The fate of sulfur in the environment can be studied using fertilizers enriched in ^{34}S or by measuring the natural abundance of ^{34}S ($\delta^{34}S$) in different sulfur forms. Knöller et al. (2005) showed by measuring $\delta^{34}S$ in sulfide and in sulfate from soils and groundwater and the natural abundance of ^{18}O ($\delta^{18}O$) in sulfate, that sulfate in water in a catchment located in eastern Germany was mostly derived from past and actual anthropogenic sulfur inputs (i.e., from the atmosphere and fertilizers) and in more restricted areas from the oxidation of sulfide. They also suggested that mineralization of C–S compounds from soil organic matter had released significant amounts of sulfur to groundwater.

Dissimilatory sulfate reduction, which is mediated by a number of enzymatic steps, strongly fractionates sulfur, with the lighter isotope being preferentially reduced, whereas sulfur isotopes are almost not fractionated during the biotic oxidation of sulfide to sulfate. Therefore, the $\delta^{34}S$ of sulfate in solution can give information on the activity of sulfate-reducing bacteria (Brunner et al., 2005). Finally, the $\delta^{18}O$ of sulfate gives information on the transfer of oxygen between water and sulfate molecules under oxidizing and reducing conditions (Brunner et al., 2005; Balci et al., 2007).

The half-life of ^{35}S is 87.5 days and this isotope emits a low-energy β^- radiation (maximum 0.17 MeV) (Endt, 1990). The fate of sulfur added to a soil/plant system in a mineral or organic compound can be directly assessed when the compound can be homogeneously labeled with ^{35}S (Vong et al., 2007). Immobilization and mineralization fluxes of sulfur in soils can be studied by labeling the soil with $^{35}SO_4^{2-}$ (Knights et al., 2001). Nziguheba et al. (2005) improved the isotope dilution method to assess sulfur immobilization by simultaneously taking into account the changes in radioactive and nonradioactive sulfate in KH_2PO_4 extracts during soil incubation. This approach gives a more precise evaluation of immobilization and therefore of net mineralization. In a further step, Nziguheba et al. (2006) extended their isotopic dilution approach to assess the mineralization and immobilization of sulfur induced by the addition of nonlabeled plant residues. In these studies, the authors obtained L values measured with Italian ryegrass that were similar to

the *E* values measured in soil/water suspensions and came to the conclusion that Italian ryegrass had very little access to nonlabeled sulfur forms.

26.1.2.4 Molecular Biology Tools for the Study of the Phosphorus and Sulfur Cycles

Molecular methods have allowed identification of changes in composition and diversity of microbial communities in response to changes in size of soil phosphorus pools and phosphate availability as achieved, for example, by different fertilization regimes (He et al., 2008; Cruz et al., 2009).

The advantage of these (mostly DNA-based) tools is mainly in their capacity to cover a broad range of both culturable and unculturable microbes, with the latter usually vastly dominating the microbial community in most soils. Sequencing of excised bands from molecular profiling gels and novel low-cost and high-throughput environmental sequencing tools (Cardenas and Tiedje) could potentially reveal the identity and abundance of microbes that specifically occur in environments with contrasting levels of available phosphate. In spite of the current technology-driven boom in metagenomics and molecular profiling tools, the DNA-based methods can provide only limited insight into the processes in phosphorus cycling, because of our limited understanding of linkages between microbial identities and functions. For example, not all members of a given taxonomic group have the same metabolic capabilities and not all microbes possessing genes for a particular metabolic pathway such as phytate degradation necessarily express the metabolic activity under all circumstances. Thus, in order to confirm the capacity of certain microbes to access certain phosphorus sources and to study the mechanisms and kinetics of the associated metabolic pathways such as release of organic acids, protons, cyanide, enzymes, and chelates (Welch et al., 2002; Chen et al., 2006; Uroz et al., 2007; Jorquera et al., 2008), it is still essential to cultivate the microbes themselves or to use transgenic organisms that express relevant genes from uncultivated organisms. Furthermore, DNA from dead cells is stable in some soils for a certain period of time (Pietramellara et al., 2009), and this may bias some of the methods based on soil DNA extractions.

These concerns have stimulated recent developments in RNA-based approaches, which offer complementary information to the analyses of phylogenetic affiliations within microbial communities and which are more directly relevant to metabolic processes than the DNA-based methods (Yergeau et al., 2009). For example, the activities of certain metabolic pathways such as cellulose degradation and N_2 fixation were shown to correlate with latitude and environmental properties along a soil transect across Antarctica, using functional gene microarray technology (Yergeau et al., 2007). Similar comprehensive studies focusing on genes related to phosphorus cycling have yet to be completed. The concepts emerging from the descriptive studies can be further scrutinized by using genetically engineered microorganisms or plants (over) expressing certain metabolic pathways (George et al., 2005).

Coupling this approach with in situ quantification of the pool sizes and reaction rates of soil enzymes, using molecular biology, proteomics, and other tools, should eventually allow mechanistic modeling of the processes involved in biological transformations of phosphorus in the soils (Nannipieri, 2006; Wallenstein and Weintraub, 2008).

Molecular methods for the analysis of microbes mediating sulfur transformations in soils have concentrated on those organisms that catalyze the inorganic sulfur cycle, with particular emphasis on sulfate/sulfite-reducing prokaryotes and on sulfur-oxidizing prokaryotes. These two groups of organisms are both taxonomically very diverse, but the members of each group share conserved pathways of sulfur-based metabolism that have allowed the development of DNA-based methods that relate functional gene diversity to taxonomic diversity. A microarray based on 16S rRNA sequences of known sulfate/sulfite-reducing prokaryotes species has been used for a number of years (Loy et al., 2002), and this has been complemented by methods based on specific oligonucleotide primers targeting the dissimilatory sulfite reductase genes (*dsrAB*) (Loy et al., 2004; Wagner et al., 2005; Schmalenberger et al., 2007), including quantification of these organisms with real-time PCR (Stubner, 2004). Polymerase chain reaction-denaturing gradient gel electrophoresis (PCR-DGGE) methods using these genes have been developed to evaluate how environmental variation affects sulfate/sulfite-reducing prokaryotes communities (Geets et al., 2006; Miletto et al., 2007) and to examine biogeography and spatial distribution of these organisms (Miletto et al., 2008). Like all DNA methods, these methods yield data on the relative abundance of the organisms concerned, but this does not necessarily reflect their activity in situ, and quantitative environmental studies based on functional RNA are still rare. Sulfur-oxidizing prokaryote diversity in various environmental habitats has also been subjected to molecular analysis, using specific primers targeting the dissimilatory adenosine-5′-phosphosulfate reductase (*aprBA*) and reverse dissimilatory sulfite reductase genes (*dsrAB*) (Meyer and Kuever, 2007a, 2007b; Loy et al., 2009). These genes are present in both sulfate-/sulfite-reducing prokaryotes and sulfur-oxidizing prokaryotes, and so an appropriate choice of primers allows simultaneous diversity analysis (Meyer and Kuever, 2007a), which promises to be a valuable tool in the future.

Molecular approaches for correlation of microbial diversity with changes in sulfur supply are less common than for the corresponding studies with phosphorus, partly because long-term fertilization experiments have seldom controlled for the sulfate levels provided (Kertesz et al., 2007). Molecular methods have been used to analyze organosulfur cycling in one such project, evaluating changes in microbial sulfonatase diversity in sulfate-rich and sulfate-limited rhizospheres using a sulfonatase-related gene (*asfA*) (Schmalenberger and Kertesz, 2007; Schmalenberger et al., 2009). However, the most widely used marker of organosulfur cycling in soils is the arylsulfatase enzyme, and molecular studies of the diversity and expression of sulfatase genes in the soil have not yet been reported.

26.1.3 Biotic Processes

26.1.3.1 Uptake of Phosphorus and Sulfur by Plants

Plants take up orthophosphate from the soil solution through root hairs and mycorrhizal symbionts. The amount of phosphate present in the soil solution at a given time is far lower than the total plant need. When a plant takes up phosphate, a phosphate-depleted zone appears along the root creating a concentration gradient from the soil particles to the root surface triggering the diffusion of phosphate from soil particles to the surface of the root.

Since phosphate ions cannot diffuse over long distances in the soil, root morphology, root hairs, and the presence of mycorrhizal symbionts are determinants for phosphate uptake. Some plant species can increase the rate of phosphate solubilization in their rhizosphere through the exudation of protons, organic acids, and phosphatases (Hinsinger, 2001; Turner, 2008). Protons dissolve calcium-bound phosphate, organic acids chelate cations bound to phosphate and phosphatases release phosphate from mono- and diester-P. White lupin is known for being extremely efficient in solubilizing scarcely soluble phosphate forms and for efficient phosphate uptake through its cluster roots (Neumann et al., 1999). *Brassicaceae* are also efficient in using scarcely soluble mineral phosphate compounds (Marschner et al., 2007).

Most plant species, however, do not exhibit such adaptations and rely on the symbiosis they establish with mycorrhizal fungi for phosphate uptake. Thanks to their extra-radical hyphae these fungi are able to take up phosphate from a volume of soil that is much larger than that explored by the roots. Phosphate is then transported to the plant and in exchange the plant provides reduced carbon to the fungus. Arbuscular mycorrhizal fungi are the most widespread; they form a symbiosis with 80% of higher plant species and are particularly important for crop plants (Jansa et al., 2006). They have a high level of functional diversity and are influenced by agricultural management practices (Jansa et al., 2006).

A number of phosphate transporters have been identified in AMF hyphae, in the plant for uploading phosphate from AMF arbuscule, in root hairs for uploading phosphate from the soil solution, for up- and downloading phosphate in the xylem and phloem, and for phosphate transfer within plant cells (Bucher, 2007). The expression of genes coding for some of these transporters (e.g., on root hairs) is upregulated when the phosphate concentration in the soil solution decreases.

Most of the sulfur in plants is taken up by roots as sulfate from the soil solution. Plants can also take up sulfur dioxide through their stomata when sulfur inputs from the atmosphere are high (Kertesz et al., 2007). Since plants can release and take up amino acids through roots, they are probably also able to take up sulfur-containing amino acids (Kertesz et al., 2007). A series of sulfate transporters and amino acid transporters have been identified for their uptake in the roots, for the uploading and downloading from the xylem and phloem, and for sulfur distribution within cells.

The role of mycorrhizal fungi on sulfur uptake by plants is not well understood. Mycorrhized plants appear to take up more sulfur than non-mycorrhized plants. Kertesz et al. (2007) hypothesize that this higher uptake could be due to the association of soil bacteria that would release sulfur to the solution, which would then be taken up by mycorrhizal hyphae. Once in the plant, sulfate is reduced and incorporated in methionine and cysteine, which are then incorporated in proteins. *Brassicaceae* and *Alliaceae* also use a large fraction of their sulfur for the production of secondary metabolites.

26.1.3.2 Role of Soil Microorganisms in the Phosphorus Cycle

Soil microorganisms can affect nutrient speciation and at the same time act as a source or sink for nutrients (Oberson and Joner, 2005). Many soil microorganisms can solubilize calcium, iron, and aluminum phosphate minerals under laboratory conditions. Under these conditions, acidification of the medium and release of various organic acids were shown to be the main mechanisms through which microorganisms could solubilize calcium phosphate compounds (Gyaneshwar et al., 2002).

Gyaneshwar et al. (2002) report however the limited effects on crop phosphate nutrition of these solubilizing microorganisms when applied in the field. These limited effects can be explained by the high buffer capacity of calcium phosphate-rich soils (generally calcareous soils) leading to a rapid disappearance of the protons released by the microorganisms (Gyaneshwar et al., 2002), by the rapid degradation of organic acids by soil microorganisms (Jones and Darrah, 1994), and/or by the growing conditions of the microorganisms in the rhizosphere that are totally different from those used in the laboratory to assess phosphate solubilization rates.

Microorganisms can develop other strategies to release phosphate from scarcely soluble minerals. Hamdali et al. (2008) show that actinomycetes isolated from Moroccan rock phosphates solubilize calcium phosphate through the production of siderophores and not organic acids. Hoberg et al. (2005) show that an actinomycete (*Gordonia* sp.) and a bacterium (*Pseudomonas fluorescens*) access phosphate sorbed on the surface of a goethite through different mechanisms. *Gordonia* sp. increases the pH of the solution, which increases the density of negative charges of the oxide surface, resulting in phosphate desorption, while *P. fluorescens* exudes citrate and decreases the pH of the solution leading to a probable dissolution of the goethite surface and to a release of phosphate.

Soil microorganisms produce extracellular phosphatases to hydrolyze esters P to phosphate. Fungal and bacterial phosphatases (phytases, acid phosphatases, and alkaline phosphatases) have been characterized, and their genes cloned and expressed in various microorganisms (Wyss et al., 1999; Moura et al., 2001). A gene encoding a protein that regulates phosphorus metabolism (phosphorus starvation response *psr1*) has been described in the green alga *Chlamydomonas reinhardtii* (Wykoff et al., 1999). It is hypothesized that during a period of phosphate starvation, the protein PSR1 is activated

and temporally regulates the transcription of specific genes encoding, for example, the production and excretion of alkaline phosphatases in the extracellular medium (Wykoff et al., 1999). Tarafdar et al. (2002) showed a strong linear relationship between the intra- versus extracellular alkaline and acid phosphatases and phytase activity in six soil fungi. About 25% of the acid and alkaline phosphatases activity was observed outside of the cell, while more than 90% of the phytase activity was observed in the extracellular medium.

Finally, some microorganisms (*Pseudomonas* sp., *Enterobacter* sp., and *Pantoea* sp.) have the capacity both to solubilize calcium phosphates and to hydrolyze phytate and other organic P (Jorquera et al., 2008). Gyaneshwar et al. (2002) consider however that more work is needed on the mechanisms and the genetics of phosphate solubilization by soil microorganisms before these can be used to efficiently improve crop nutrition.

Once released in the solution, phosphate can be taken up by microorganisms to cover their needs and either be remineralized or stored in organic forms for longer periods. The amount of phosphate stored in the microbial biomass as measured by fumigation–extraction techniques ranges between 1% and 10% of total P in the upper horizon of cropped and grassland soils, and reaches higher values in forest soils.

Microbial phosphate content is determined by the total microbial biomass and its activity, which in turn is controlled by the soil organic matter content and organic matter inputs (Oberson and Joner, 2005). Bünemann et al. (2008) used liquid-state ^{31}P NMR to characterize the forms of phosphorus stored in bacteria and fungi isolated from two soils. They observed mono- and diester phosphates in bacteria and fungi, but the concentrations of these phosphate species were higher in bacteria. They also observed higher concentrations of pyro- and polyphosphate in fungi. Finally, they showed that the forms of phosphorus observed in fungi and bacteria were similar to those observed in NaOH–EDTA soil extracts.

When microbial growth is favored, for example, following the addition of organic substances, or when phosphate availability is very limited, microorganisms tend to accumulate phosphorus without releasing it to plants (Bünemann et al., 2004). Phosphate release then occurs following strong changes in the microenvironment, such as drying/rewetting or freezing/thawing cycles (Turner and Haygarth, 2001), or following predation, for example, by protozoa, nematodes or amoebae. Microbial phosphorus and soil organic phosphorus can be remineralized either following the death of microorganisms or through the action of extracellular enzymes as described above.

Soil organic phosphorus mineralization has not been studied extensively because of methodological difficulties. The rate of organic phosphorus mineralization is linked to soil biological activity (Oehl et al., 2004) and most probably also to the microbial community structure and to the organic phosphorus content and forms. The application of exogenous organic matter can increase or decrease phosphate release to the solution depending on its quality. Ha et al. (2007) observed that residues of flowering pea with a low C/P led to a strong release of phosphate while residues of mature pea with a very high C/P ratio led to phosphate immobilization by microorganisms.

26.1.3.3 Role of Soil Microorganisms in the Sulfur Cycle

Soil microorganisms can affect sulfur forms and availability in four ways: (1) through oxidation, when reduced sulfur is used as an electron donor for respiration; (2) through reduction when oxidized sulfur is used as a terminal electron acceptor; (3) through immobilization when inorganic sulfur is taken up by microorganisms to build proteins and structural cell components; and (4) through mineralization when microbial sulfur is released as inorganic sulfur to the soil solution (Siciliano and Germida, 2005).

Many soil bacteria are able to use sulfur as a source of energy by oxidizing elemental sulfur (S°) to sulfite and sulfate using O_2 or nitrate as final electron acceptors. It is assumed that *Thiobacillus* sp. dominate when energy and organic matter are limited and when dissolved reduced sulfur compounds are present in high concentrations. In soils rich in organic matter, when levels of reduced sulfur compounds are low, sulfur oxidation is carried out by heterotrophs (Siciliano and Germida, 2005).

Sulfate-reducing bacteria are found in highly anaerobic environments. They oxidize organic compounds (lactate, pyruvate, and acetate) or hydrogen as electron donors, and use the enzymes adenosine-5′-phosphosulfate reductase and dissimilatory sulfite reductase to transfer electrons to adenosine-5′-phosphosulfate and sulfite leading to the production of hydrogen sulfide that precipitates in the presence of metals.

Finally, some bacteria with phylogenetic affiliation to the sulfate-reducing bacteria can also disproportionate sulfur (Finster, 2008). Such bacteria can for instance oxidize one sulfur atom of thiosulfate to sulfate while simultaneously reducing the other sulfur atom to hydrogen sulfide (Equation 26.1):

$$S_2O_3^{2-} + H_2O \rightarrow SO_4^{2-} + HS^- + H^+ \qquad (26.1)$$
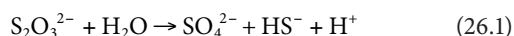
Like plants, soil microorganisms take up sulfate from the solution through transporters, reduce it to sulfite, and incorporate it into cysteine and methionine and then into proteins. When sulfate has been exhausted, some microorganisms take up sulfonates and transform them into sulfite to cover their sulfur needs. The genetic control and the transformation pathways of sulfonate assimilation by bacteria have been studied extensively. Three gene clusters seem important: the *ats* gene cluster, which is responsible for the binding of aromatic sulfonate, the *ssu* gene cluster, which controls the cleavage of the C–S bond, and the *asf* gene cluster, which provides reducing equivalents (Siciliano and Germida, 2005; Kertesz et al., 2007).

The use of sulfonate by bacteria is a very relevant phenomenon as (1) during plant growth, sulfate gets depleted along plant roots, (2) sulfonates make up an important fraction of organic S in aerobic soils, and (3) plants are not able to use sulfonates (Kertesz et al., 2007). The principle rhizosphere genera able to desulfurize sulfonates belong to the β-proteobacteria (e.g., *Variovorax*, *Acidovorax*, and *Polaromonas* strains) and actinobacteria

(mostly *Rhodococcus* strains) (Kertesz et al., 2007). However, given the complex environment of sulfonate molecules in the soil, these bacteria or other rhizosphere organisms must be able to extract sulfonate from the soil matrix and make it available to the bacteria. These mechanisms remain to be discovered. Once the needs of the microorganisms for sulfur are covered, they downregulate the genes controlling the use of sulfonate. Kertesz et al. (2007) suggest that through predation of these bacteria by, for example, protozoa, sulfur contained in these bacteria could be released into the soil solution and become available to the plant.

In response to sulfate deficiency, soil microorganisms can also produce arylsulfatase enzymes that release sulfate from sulfate ester, but little is known about the microorganisms producing these enzymes. Work done by Vong et al. (2007) suggests that the release of carbon and nitrogen from roots increases microbial immobilization of sulfur and sulfate mineralization, and that sulfate net mineralization is much more important in the rhizosphere of oilseed rape than in the rhizosphere of barley. They showed furthermore that arylsulfatase activity was higher in the rhizosphere soil of oilseed rape. These results complement those obtained by Dedourge et al. (2004) showing that the application of glucose and of model "rhizodeposits" to soils result in sulfur immobilization and that the increase in intra- and extracellular microbial arylsulfatase activity is positively related to the amount of immobilized sulfur. Altogether these results suggest that the dynamics of sulfur in the rhizosphere is mediated by microorganisms using rhizodeposits as carbon source.

Once sulfate has been released into the soil, it can be taken up again by microorganisms and stored in different organic forms or be remineralized. Soil microbial sulfur as measured by the fumigation technique ranges between 1% and 3% of total soil S (Zhao et al., 1996). A case study showing the relationships between microbial activity and sulfur dynamics as affected by land use and fertilization is presented in Section 26.1.4.2.

26.1.3.4 Role of Soil Macrofauna in the Phosphorus and Sulfur Cycles

The soil macrofauna (earthworms, termites, ants, etc.) affects nutrient cycling through decomposition of organic residues and through effects on the spatial distribution and accessibility of nutrients in the soil. Since most work has concentrated on the role of earthworms in phosphorus dynamics, the following section will focus on that aspect.

Earthworms create structures that represent habitats with characteristics different from the surrounding soil (Decaens et al., 2002). Plant-available phosphate concentrations are usually higher in earthworm casts (Jiménez et al., 2003; Chapuis-Lardy et al., 2009) and in burrow walls (Tiunov et al., 2001). Under field conditions where earthworms ingest soil plus organic residues, total phosphorus concentrations are higher in casts than in surrounding soil (Jiménez et al., 2003). This increase is distributed over all inorganic and organic phosphorus fractions, including microbial phosphorus. Thus, particularly under phosphate-deficient conditions, earthworm casts present spots with clearly increased phosphate availability for plants. Ouedraogo et al.

(2005) report that earthworms contribute to the solubilization of rock phosphate.

However, negative impacts of earthworms on P cycling have also been reported. The invasion of north temperate forests by exotic species of earthworms resulted in the loss of the forest litter, which altered the location and nature of nutrient cycling in the soil profile (Suaréz et al., 2004). On cropped land, earthworm surface-casts can increase soil erosion and particulate transfer if the casts become detached from the soil matrix (Le Bayon et al., 2002). Thus, earthworm surface-casts can contribute to the loss of organic matter and phosphorus from soils because of the selective ingestion of the finest particles and organic material by earthworms and the resulting higher cast phosphorus concentration.

Complex trophic interactions exist between soil macrofauna (e.g., earthworms, termites, and diplopods), mesofauna (mainly microarthropods such as collembola, acarids, and enchytraeids), microfauna (nematodes and protozoa), and microflora (bacteria and fungi) (Cortet et al., 2003; Thies and Grossman, 2006). Microbial biomass, basal respiration, microbial volume, and abundance of protozoa are greater in *Lumbricus terrestris* burrow walls than in surrounding control soil (Tiunov et al., 2001), showing that several groups of organisms affect phosphate availability in the microhabitats created by earthworms.

Land use has a strong impact on macrofaunal communities (Decaens et al., 2004). The conversion of tropical forests or native savanna to agropastoral systems resulted in a sharp decrease in the volume of biostructures created by earthworms and ants when an annual crop was introduced (Decaens et al., 2002) or after slashing and burning (Decaens et al., 2004). As soils of these systems are often phosphate limited, this decrease represents a loss of niches with higher phosphate availability for plants. In contrast, the density of macrofauna and facultative phytophagous, bacterial feeding and predatory nematodes in soils under maize–legume rotations are higher than in maize-monocropped soils, and this promotes soil structure and nutrient availability (Blanchart et al., 2006). In agroforestry systems established on savanna soils, organic fertilization increased earthworm abundance (Lopez-Hernandez et al., 2004). Macroorganisms play a dominant role in the initial phase of manure decomposition (Esse et al., 2001).

While the role of the soil fauna in the cycling of phosphorus is acknowledged, the role of the fauna in the cycling of sulfur remains largely unknown and needs to be investigated.

26.1.4 Case Studies: Phosphorus and Sulfur Cycling in Two Exemplary Soil Systems

26.1.4.1 Phosphorus Forms and Dynamics in the Soil Chronosequence of the Franz Josef Glacier (New Zealand)

The relationships between soil phosphorus and soil development have been studied intensively during the last decades. These studies show that as the soil gets older, total phosphorus stock (expressed in mass of phosphorus per unit of surface) decreases,

the proportion of phosphate bound to calcium decreases, and the proportion of phosphorus bound to organic matter and oxides first increases and then decreases and stabilizes (Walker and Syers, 1976). These studies however relied mostly on sequential extraction schemes for the characterization of soil phosphorus forms and did not consider soil biological activity as a driver of these changes. Work recently conducted in the soil chronosequence of the Franz Josef Glacier located on the West Coast of the South Island of New Zealand has helped to fill this gap (Richardson et al., 2004; Allison et al., 2007; Turner et al., 2007; Doblas-Miranda et al., 2008).

Sites with soil ages varying between 5 and 120,000 years developed on the same parent material (graywacke and schist) were identified along this chronosequence. Strong changes in plant, soil fauna and soil microorganism communities were observed along these sites. Evergreen angiosperms dominated in the younger sites and then decreased while conifers (*Podocarpaceae*), which were absent from the younger sites, appeared in the 500 year old site and became dominant in the 120,000 year old site. Foliar analyses suggested that vegetation became phosphate limited in sites older than 500 years (Richardson et al., 2004). The nitrogen to phosphorus ratio of the leaves of the *Podocarpaceae* found in the older sites indicated that it was less phosphate limited than the evergreen angiosperms. Doblas-Miranda et al. (2008) observed an increase in microbial-feeding nematodes and omnivorous nematodes from the younger site to the 5,000 year old site and then a general decrease of these nematodes till the 120,000 year old site. Interestingly, they observed a dominance of fungal-feeding nematodes in the older sites. Finally, Allison et al. (2007), using phospholipid fatty acid analyses, observed a shift from bacterial to fungal microbial communities from the youngest to the oldest sites. The same authors observed an increase in phosphatase (phosphomonoesterase and phosphodiesterase) activities with increasing soil age and even an increase in phosphatase efficiency (phosphatase activity per unit microbial biomass). Since these increases were correlated with a decrease in soil phosphorus content, these changes in phosphatase activity were interpreted as strategies developed by organisms to scavenge phosphorus in an environment that becomes more and more limiting.

Turner et al. (2007) confirmed the results presented by Walker and Syers (1976) on changes in total P and in P forms, and also conducted a detailed analysis of soil organic phosphorus forms in NaOH–EDTA soil extracts using liquid-state ^{31}P NMR (Figure 26.2). They observed an increase with soil age till 500 years of two isomers of inositol hexaphosphate (phytate and *scyllo*-inositol hexakisphosphate), other monoesters, DNA, phospholipids, phosphonates, pyrophosphate, and orthophosphate. The amount of phosphorus present in these forms then decreased in the older sites. The proportion of extractable organic phosphorus in the form of *myo*-inositol hexakisphosphate decreased after 500 years whereas the proportion of organic phosphorus present as DNA and the proportion of extractable inorganic phosphorus present as pyrophosphate continuously increased with soil age. These results are astonishing as *myo*-inositol hexakisphosphate is thought to be very stable in soils whereas DNA and pyrophosphate are thought to degrade rapidly. The increase in pyrophosphate in the older sites can be linked to the increased presence of fungi as noted by Bünemann et al. (2008) or to the stabilization of pyrophosphate by organic matter (Turner et al., 2007). The increase in DNA could be linked to its stabilization by organic matter (Turner et al., 2007) or on clay minerals. The decrease in *myo*-inositol hexakisphosphate could be related to the decrease in iron and aluminum oxides content. Much remains to be done to understand by which mechanisms phosphorus can be recycled in the older soils where organic phosphorus forms appear to be very stable and available phosphate is very low (Turner et al., 2007). Rhizosphere processes occurring in the root zone of the *Podocarpaceae*, fungal activity, and nutrient recycling through soil fauna might play a determining role (Turner et al., 2007). Although Turner et al. (2007) do not mention them, the mechanisms controlling phosphate release during the very initial steps of soil development would also be interesting to study as this release happens very rapidly.
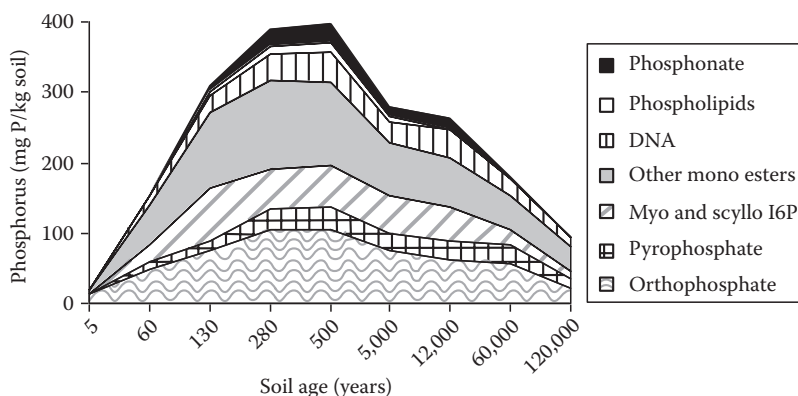
FIGURE 26.2 Forms and quantities of phosphorus extracted by NaOH–EDTA from the upper horizon of soils of the 120,000 year Franz Josef postglacial chronosequence, New Zealand. (Adapted with kind permission from Springer Science+Business Media: Turner, B.L., L.M. Condron, S.J. Richardson, D.A. Peltzer, and V.J. Allison. 2007. Soil organic phosphorus transformations during pedogenesis. Ecosystems 10:1166–1181.)

TABLE 26.1 Percentage Distribution of Organic Sulfur Determined by Sulfur K Edge XANES in Humic Acids Extracts and Sulfur Immobilization and Gross Mineralization Measured in a 53 Day Incubation in the Presence of ^{35}S in Topsoils from the Long-Term Broadbalk Experiment

Soil	Reduced S[a] (%)[b]	Intermediate S[c] (%)[b]	Oxidized S[d] (%)[b]	S Immobilization in Soil (mg S kg[−1])	Gross S Mineralization in Soil (mg S kg[−1])
Woodland since 1882	24.8	46.0	29.1	nd[e]	nd
Grassland since 1882	18.4	37.4	44.3	3.9 (0.20)	6.2 (1.08)
Wheat not fertilized since 1843	14.7	32.8	52.5	3.5 (0.32)	0.9 (0.05)
Wheat fertilized with farmyard manure since 1843	28.7	49.7	21.5	4.5 (0.96)	6.8 (1.26)

Source: Adapted from Zhao, F.J., J. Lehmann, D. Solomon, M.A. Fox, and S.P. McGrath. 2006. Sulphur speciation and turnover in soils: Evidence from sulphur K-edge XANES spectroscopy and isotope dilution studies. Soil Biol. Biochem. 38:1000–1007.

[a] S in a strongly reduced state (e.g., sulfide and thiophene).

[b] Percent of organic S measured using XANES (reduced S + intermediate S + oxidized S = 100%).

[c] S in an intermediate oxidation state (sulfoxide and sulfonate).

[d] S in a strongly oxidized state (ester sulfate).

[e] Not determined.

26.1.4.2 Sulfur Forms and Dynamics in the Broadbalk Long-Term Experiment

Long-term field experiments are essential to assess the effects of agricultural practices and environmental factors on crop yield and soil fertility. The Broadbalk experiment is the world's oldest agricultural field experiment and is extensively described in Poulton et al. (2006). The experiment was set up in 1843 on an Aquic Paleudalf in Rothamsted to the north of London. Its aim was to evaluate the effect of different fertilizations (e.g., no fertilization, full mineral fertilization [NPKS], and farmyard manure) on winter wheat. In 1882, a fraction of the surface under wheat was set aside, and a woodland and a grassland were allowed to develop. Research conducted in this long-term field experiment allowed assessing the effect of sulfur inputs from the atmosphere on the sulfur budget of the soil-crop system, to understand the relationships between land use, fertilization, and sulfur forms in soils, and more recently assessing the effect of fertilization on microbial communities that are important for the mineralization of sulfonates.

Knights et al. (2000) and Zhao et al. (2003) showed that emissions of sulfur to the atmosphere that increased between 1840 and 1970 and then decreased sharply resulted in a significant decrease in δ^{34}S in soil and wheat samples. These results suggest that at the peak of sulfur emissions, 62%–78% of S in the wheat and 28%–37% of S in the top soil of the nonfertilized treatment was derived from sulfur dioxide inputs from the atmosphere. The incorporation of sulfur from atmospheric inputs was higher in the soil that had received manure and on the soils from the woodland and the grassland, leading to an increase in total sulfur content. In treatments that had not received organic matter input in contrast large amounts of sulfur were lost by leaching, resulting in constant total soil sulfur content (Knights et al., 2000).

Zhao et al. (2006) analyzed sulfur speciation by sulfur K edge XANES and the rate of sulfur mineralization and immobilization using a ^{35}S dilution technique approach in the nonfertilized, the NPKS-fertilized, the manure fertilized, the woodland, and the grassland soils. They showed that soils with high organic matter content (woodland, grassland, and manured soils) contained more reduced and intermediate sulfur and less oxidized sulfur than arable soils that did not receive organic matter. Furthermore, the fraction of sulfur released by mineralization was correlated to the fraction of sulfur present in reduced and intermediate forms (Table 26.1).

Finally, Schmalenberger et al. (2008, 2009) identified *Variovorax*, *Polaromonas* (both β-proteobacteria), and *Rhodococcus* (actinobacteria) strains from the rhizosphere of wheat from the Broadbalk experiment able to desulfurize sulfonate. Based on the analysis of DGGE profiles, they showed that the communities of β-proteobacteria and actinobacteria sampled in rhizosphere of wheat were affected by sulfate and manure applications. Analysis of the functional gene *asfA* (oxidoreductase) showed that 40% of the clones belonged to the genus *Variovorax*, that this gene was present in higher number in the rhizosphere soil (where sulfate is depleted) than in the bulk soil, and that higher copy numbers were present in the sulfate-limited soils.

These results suggest that rhizobacteria play an important role in the cycle of sulfur in the rhizosphere and therefore on wheat mineral nutrition, and at a higher level, that land use and organic matter inputs drive sulfur speciation and sulfur immobilization/mineralization in soil.

References

Allison, V.J., L.M. Condron, D.A. Peltzer, S.J. Richardson, and B.L. Turner. 2007. Changes in enzyme activities and soil microbial community composition along carbon and nutrient gradients at the Franz Josef chronosequence. N.Z. Soil Biol. Biochem. 39:1770–1781.

Balci, N., W.C. Shanks, B. Mayer, and K.W. Mandernack. 2007. Oxygen and sulfur isotope systematics of sulfate produced by bacterial and abiotic oxidation of pyrite. Geochim. Cosmochim. Acta 71:3796–3811.

Beauchemin, S., D. Hesterberg, J. Chou, M. Beauchemin, R.R. Simard, and D.E. Sayers. 2003. Speciation of phosphorus in phosphorus-enriched agricultural soils using X-ray absorption near-edge structure spectroscopy and chemical fractionation. J. Environ. Qual. 32:1809–1819.

Blanchart, E., C. Villenave, A. Viallatoux, B. Barthes, C. Girardin, A. Azontonde, and C. Feller. 2006. Long-term effect of a legume cover crop (*Mucuna pruriens* var. *utilis*) on the communities of soil macrofauna and nematofauna, under maize cultivation, in southern Benin. Eur. J. Soil Biol. 42:S136–S144.

Brunner, B., S.M. Bernasconi, J. Kleikemper, and M.H. Schroth. 2005. A model for oxygen and sulfur isotope fractionation in sulfate during bacterial sulfate reduction processes. Geochim. Cosmochim. Acta 69:4773–4785.

Bucher, M. 2007. Functional biology of plant phosphate uptake at root and mycorrhiza interfaces. New Phytol. 173:11–26.

Bühler, S., A. Oberson, I.M. Rao, E. Frossard, and D.K. Friesen. 2002. Sequential phosphorus extraction of a ^{33}P-labeled oxisol under contrasting agricultural systems. Soil Sci. Soc. Am. J. 66:868–877.

Bünemann, E.K. 2008. Enzyme additions as a tool to assess the potential bioavailability of organically bound nutrients. Soil Biol. Biochem. 40:2116–2129.

Bünemann, E.K., and L.M. Condron. 2007. Phosphorus and sulphur cycling in terrestrial ecosystems, p. 65–92. *In* P. Marschner and Z. Rengel (eds.) Soil biology. Vol. 10. Nutrient cycling in terrestrial ecosystems. Springer-Verlag, Heidelberg, Germany.

Bünemann, E.K., P. Marschner, A.M. McNeill, and M.J. McLaughlin. 2007. Measuring rates of gross and net mineralisation of organic phosphorus in soils. Soil Biol. Biochem. 39:900–913.

Bünemann, E.K., R.J. Smernik, A.L. Doolette, P. Marschner, R. Stonor, S.A. Wakelin, and A.M. McNeill. 2008. Forms of phosphorus in bacteria and fungi isolated from two Australian soils. Soil Biol. Biochem. 40:1908–1915.

Bünemann, E.K., F. Steinebrunner, P.C. Smithson, E. Frossard, and A. Oberson. 2004. Phosphorus dynamics in a highly weathered soil as revealed by isotopic labeling techniques. Soil Sci. Soc. Am. J. 68:1645–1655.

Cardenas, E., and J.M. Tiedje. 2008. New tools for discovering and characterizing microbial diversity. Curr. Opin. Biotechnol. 19:544–549.

Chapuis-Lardy, L., R.S. Ramiandrisoa, L. Randriamanantsoa, C. Morel, L. Rabeharisoa, and E. Blanchart. 2009. Modification of P availability by endogeic earthworms (Glossoscolecidae) in ferralsols of the Malagasy Highlands. Biol. Fertil. Soils 45:415–422.

Chen, Y.P., P.D. Rekha, A.B. Arun, F.T. Shen, W.A. Lai, and C.C. Young. 2006. Phosphate solubilizing bacteria from subtropical soil and their tricalcium phosphate solubilizing abilities. Appl. Soil Ecol. 34:33–41.

Condron, L.M., E. Frossard, H. Tiessen, R.H. Newman, and J.W.B. Stewart. 1990. Chemical nature of organic phosphorus in cultivated and uncultivated soils under different environmental conditions. J. Soil Sci. 41:41–50.

Cordell, D., J.O. Drangert, and S. White. 2009. The story of phosphorus: Global food security and food for thought. Global Environ. Change-Hum. Policy Dimens. 19:292–305.

Cortet, J., R. Joffre, S. Elmholt, and P.H. Krogh. 2003. Increasing species and trophic diversity of mesofauna affects fungal biomass, mesofauna community structure and organic matter decomposition processes. Biol. Fertil. Soils 37:302–312.

Cruz, A.F., C. Hamel, K. Hanson, F. Selles, and R.P. Zentner. 2009. Thirty-seven years of soil nitrogen and phosphorus fertility management shapes the structure and function of the soil microbial community in a Brown Chernozem. Plant Soil 315:173–184.

Decaens, T., N. Asakawa, J.H. Galvis, R.J. Thomas, and E. Amezquita. 2002. Surface activity of soil ecosystem engineers and soil structure in contrasted land use systems of Colombia. Eur. J. Soil Biol. 38:267–271.

Decaens, T., J.J. Jiménez, E. Barrios, A. Chauvel, E. Blanchart, C. Fragoso, and P. Lavelle. 2004. Soil macrofaunal communities in permanent pastures derived from tropical forest or savanna. Agric. Ecosyst. Environ. 103:301–312.

Dedourge, O., P.C. Vong, F. Lasserre-Joulin, E. Benizri, and A. Guckert. 2004. Effects of glucose and rhizodeposits (with or without cysteine-S) on immobilized-^{35}S, microbial biomass-^{35}S and arylsulphatase activity in a calcareous and an acid brown soil. Eur. J. Soil Sci. 55:649–656.

Doblas-Miranda, E., D.A. Wardle, D.A. Peltzer, and G.W. Yeates. 2008. Changes in the community structure and diversity of soil invertebrates across the Franz Josef Glacier chronosequence. Soil Biol. Biochem. 40:1069–1081.

Dougherty, W.J., R.J. Smernik, and D.J. Chittleborough. 2005. Application of spin counting to the solid-state ^{31}P NMR analysis of pasture soils with varying phosphorus content. Soil Sci. Soc. Am. J. 69:2058–2070.

Endt, P.M. 1990. Energy levels of A = 21–44 nuclei (VII). Nucl. Phys. A 521:1–830.

Esse, P.C., A. Buerkert, P. Hiernaux, and A. Assa. 2001. Decomposition of and nutrient release from ruminant manure on acid sandy soils in the Sahelian zone of Niger, West Africa. Agric. Ecosyst. Environ. 83:55–63.

Finster, K. 2008. Microbiological disproportionation of inorganic sulfur compounds. J. Sulfur Chem. 29:281–292.

Freney, J.R., G.E. Melville, and C.H. Williams. 1975. Soil organic matter fractions as sources of plant available sulfur. Soil Biol. Biochem. 7:217–221.

Frossard, E., and S. Sinaj. 1997. The isotope exchange kinetic technique: A method to describe the availability of inorganic nutrients. Applications to K, P, S and Zn. Isot. Environ. Health Stud. 33:61–77.

Gallet, A., R. Flisch, J.-P. Ryser, J. Nösberger, E. Frossard, and S. Sinaj. 2003. Uptake of residual phosphate and freshly applied diammonium phosphate by *Lolium perenne* and *Trifolium repens*. J. Plant Nutr. Soil Sci. 166:557–567.

Geets, J., B. Borrernans, L. Diels, D. Springael, J. Vangronsveld, D. van der Lelie, and K. Vanbroekhoven. 2006. *DsrB* gene-based DGGE for community and diversity surveys of sulfate-reducing bacteria. J. Microbiol. Methods 66:194–205.

George, T.S., R.J. Simpson, P.A. Hadobas, and A.E. Richardson. 2005. Expression of a fungal phytase gene in *Nicotiana tabacum* improves phosphorus nutrition of plants grown in amended soils. Plant Biotechnol. J. 3:129–140.

Gyaneshwar, P., G.N. Kumar, L.J. Parekh, and P.S. Poole. 2002. Role of soil microorganisms in improving P nutrition of plants. Plant Soil 245:83–93.

Ha, K.V., P. Marschner, E.K. Bünemann, and R. Smernik. 2007. Chemical changes and phosphorus release during decomposition of pea residues in soil. Soil Biol. Biochem. 39:2696–2699.

Hamdali, H., B. Bouizgarne, M. Hafidi, A. Lebrihi, M.J. Virolle, and Y. Ouhdouch. 2008. Screening for rock phosphate solubilizing actinomycetes from Moroccan phosphate mines. Appl. Soil Ecol. 38:12–19.

He, J.Z., Y. Zheng, C.R. Chen, Y.Q. He, and L.M. Zhang. 2008. Microbial composition and diversity of an upland red soil under long-term fertilization treatments as revealed by culture-dependent and culture-independent approaches. J. Soils Sediments 8:349–358.

Hedley, M.J., and J.W.B. Stewart. 1982. Method to measure microbial phosphate in soils. Soil Biol. Biochem. 14:377–385.

Hedley, M.J., J.W.B. Stewart, and B.S. Chauhan. 1982. Changes in inorganic and organic soil phosphorus fractions induced by cultivation practices and by laboratory incubations. Soil Sci. Soc. Am. J. 46:970–976.

Hesterberg, D., W.Q. Zhou, K.J. Hutchison, S. Beauchemin, and D.E. Sayers. 1999. XAFS study of adsorbed and mineral forms of phosphate. J. Synchrotron Radiat. 6:636–638.

Hinsinger, P. 2001. Bioavailability of soil inorganic P in the rhizosphere as affected by root-induced chemical changes: A review. Plant Soil 237:173–195.

Hoberg, E., P. Marschner, and R. Lieberei. 2005. Organic acid exudation and pH changes by *Gordonia* sp. and *Pseudomonas fluorescens* grown with P adsorbed to goethite. Microbiol. Res. 160:177–187.

Jansa, J., A. Wiemken, and E. Frossard. 2006. The effects of agricultural practices on arbuscular mycorrhizal fungi, p. 89–116. *In* E. Frossard, W.E.H. Blum, and B.P. Warkentin (eds.) Function of soils for human societies and the environment. Special Publication No. 266. Geological Society, London, U.K.

Jiménez, J.J., A. Cepeda, T. Decaëns, A. Oberson, and D.K. Friesen. 2003. Phosphorus fractions and dynamics in surface earthworm casts under native and improved grasslands in a Colombian savanna oxisol. Soil Biol. Biochem. 35:715–727.

Jones, D.L., and P.R. Darrah. 1994. Role of root derived organic acids in the mobilization of nutrients from the rhizosphere. Plant Soil 166:247–257.

Jorquera, M.A., M.T. Hernandez, Z. Rengel, P. Marscher, and S.D. Mora. 2008. Isolation of culturable phosphobacteria with both phytate-mineralization and phosphate-solubilization activity from the rhizosphere of plants growing in a volcanic soil. Biol. Fertil. Soils 44:1025–1034.

Kertesz, M.A., E. Fellows, and A. Schmalenberger, A. 2007. Rhizobacteria and plant sulfur supply. Adv. Appl. Microbiol. 62:235–268.

Knights, J.S., F.J. Zhao, S.P. McGrath, and N. Magan. 2001. Long-term effects of land use and fertiliser treatments on sulphur transformations in soils from the Broadbalk experiment. Soil Biol. Biochem. 33:1797–1804.

Knights, J.S., F.J. Zhao, B. Spiro, and S.P. McGrath. 2000. Long-term effects of land use and fertilizer treatments on sulfur cycling. J. Environ. Qual. 29:1867–1874.

Knöller, K., R. Trettin, and G. Strauch. 2005. Sulphur cycling in the drinking water catchment area of Torgau-Mockritz (Germany): Insights from hydrochemical and stable isotope investigations. Hydrol. Processes 19:3445–3465.

Kruse, J., P. Leinweber, K.-U. Eckhardt, F. Godlinski, Y. Hu, and L. Zuin. 2009. Phosphorus $L_{2,3}$-edge XANES: Overview of reference compounds. J. Synchrotron Radiat. 16:247–259.

Le Bayon, R.C., S. Moreau, C. Gascuel-Odoux, and F. Binet. 2002. Annual variations in earthworm surface-casting activity and soil transport by water runoff under a temperate maize agroecosystem. Geoderma 106:121–135.

Lombi, E., K.G. Scheckel, R.D. Armstrong, S. Forrester, J.N. Cutler, and D. Paterson. 2006. Speciation and distribution of phosphorus in a fertilized soil: A synchrotron-based investigation. Soil Sci. Soc. Am. J. 70:2038–2048.

Lopez-Hernandez, D., Y. Araujo, A. Lopez, I. Hernandez-Valencia, and C. Hernandez. 2004. Changes in soil properties and earthworm populations induced by long-term organic fertilization of a sandy soil in the Venezuelan Amazonia. Soil Sci. 169:188–194.

Loy, A., S. Duller, C. Baranyi, M. Mussmann, J. Ott, I. Sharon et al. 2009. Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. Environ. Microbiol. 11:289–299.

Loy, A., K. Kusel, A. Lehner, H.L. Drake, and M. Wagner. 2004. Microarray and functional gene analyses of sulfate-reducing prokaryotes in low-sulfate, acidic fens reveal co-occurrence of recognized genera and novel lineages. Appl. Environ. Microbiol. 70:6998–7009.

Loy, A., A. Lehner, N. Lee, J. Adamczyk, H. Meier, J. Ernst et al. 2002. Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. Appl. Environ. Microbiol. 68:5064–5081.

Marschner, P., Z. Solaiman, and Z. Rengel. 2007. *Brassica* genotypes differ in growth, phosphorus uptake and rhizosphere properties under P-limiting conditions. Soil Biol. Biochem. 39:87–98.

McDowell, R.W., L.M. Condron, N. Mahieu, P.C. Brookes, P.R. Poulton, and A.N. Sharpley. 2002. Analysis of potentially mobile phosphorus in arable soils using solid state nuclear magnetic resonance. J. Environ. Qual. 31:450–456.

McLaughlin, M.J., D.J. Reuter, and G.E. Rayment. 1999. Soil testing—Principles and concepts, p. 1–21. *In* K.I. Peverill, L.A. Sparrow, and D.J. Reuter (eds.) Soil analysis—An interpretation manual. The Australasian Soil and Plant Analysis Council Inc., CSIRO Publishing, Collingwood, Australia. Available online at http://www.publish.csiro.au/pid/1998.htm (accessed June 13, 2009).

Meyer, B., and J. Kuever. 2007a. Molecular analysis of the diversity of sulfate-reducing and sulfur-oxidizing prokaryotes in the environment, using *aprA* as functional marker gene. Appl. Environ. Microbiol. 73:7664–7679.

Meyer, B., and J. Kuever. 2007b. Molecular analysis of the distribution and phylogeny of dissimilatory adenosine-5′-phosphosulfate reductase-encoding genes (*aprBA*) among sulfur-oxidizing prokaryotes. Microbiology 153:3478–3498.

Miletto, M., P.L.E. Bodelier, and H.J. Laanbroek. 2007. Improved PCR-DGGE for high resolution diversity screening of complex sulfate-reducing prokaryotic communities in soils and sediments. J. Microbiol. Methods 70:103–111.

Miletto, M., A. Loy, A.M. Antheunisse, R. Loeb, P.L.E. Bodelier, and H.J. Laanbroek. 2008. Biogeography of sulfate-reducing prokaryotes in river floodplains. FEMS Microbiol. Ecol. 64:395–406.

Morra, M.J., S.E. Fendorf, and P.D. Brown. 1997. Speciation of sulfur in humic and fulvic acids using X-ray absorption near-edge structure (XANES) spectroscopy. Geochim. Cosmochim. Acta 61:683–688.

Moura, R.S., J.F. Martin, A. Martin, and P. Liras. 2001. Substrate analysis and molecular cloning of the extracellular alkaline phosphatase of *Streptomyces griseus*. Microbiol. SGM 147:1525–1533.

Nannipieri, P. 2006. Role of stabilised enzymes in microbial ecology and enzyme extraction from soil with potential applications in soil proteomics, p. 75–94. *In* P. Nannipieri and K. Smalla (eds.) Nucleic acids and proteins in soil. Springer-Verlag, Heidelberg, Germany.

Neumann, G., A. Massonneau, E. Martinoia, and V. Romheld. 1999. Physiological adaptations to phosphorus deficiency during proteoid root development in white lupin. Planta 208:373–382.

Nziguheba, G., E. Smolders, and R. Merckx. 2005. Sulphur immobilization and availability in soils assessed using isotope dilution. Soil Biol. Biochem. 37:635–644.

Nziguheba, G., E. Smolders, and R. Merckx. 2006. Mineralization of sulfur from organic residues assessed by inverse isotope dilution. Soil Biol. Biochem. 38:2278–2284.

Oberson, A., and E. Joner. 2005. Microbial turnover of phosphorus in soil, p. 133–164. *In* B.L. Turner, E. Frossard, and D.S.P. Baldwin (eds.) Organic phosphorus in the environment. CABI, Wallingford, U.K.

Oehl, F., E. Frossard, A. Fliessbach, D. Dubois, and A. Oberson. 2004. Basal phosphorus mineralisation in soils under different farming systems. Soil Biol. Biochem. 36:667–675.

Ouedraogo, E., L. Brussaard, A. Mando, and L. Stroosnijder. 2005. Organic resources and earthworms affect phosphorus availability to sorghum after phosphate rock addition in semi-arid West Africa. Biol. Fertil. Soils 41:458–465.

Pietramellara, G., J. Ascher, F. Borgogni, M.T. Ceccherini, G. Guerri, and P. Nannipieri. 2009. Extracellular DNA in soil and sediment: Fate and ecological relevance. Biol. Fertil. Soils 45:219–235.

Poulton et al. 2006. Guide to the classical and other long-term experiments, datasets and sample archive. Rothamsted Research. Harpenden, Herts, U.K. Available online at: http://www.rothamsted.bbsrc.ac.uk/resources/Long TermExperiments.pdf (accessed June 14, 2009).

Prietzel, J., J. Thieme, A. Herre, M. Salome, and D. Eichert. 2008. Differentiation between adsorbed and precipitated sulphate in soils and at micro-sites of soil aggregates by sulphur K-edge XANES. Eur. J. Soil Sci. 59:730–743.

Prietzel, J., J. Thieme, M. Salome, and H. Knicker. 2007. Sulfur K-edge XANES spectroscopy reveals differences in sulfur speciation of bulk soils, humic acid, fulvic acid, and particle size separates. Soil Biol. Biochem. 39:877–890.

Pypers, P., L. Van Loon, J. Diels, R. Abaidoo, E. Smolders, and R. Merckx. 2006. Plant-available P for maize and cowpea in P-deficient soils from the Nigerian Northern Guinea Savanna—Comparison of E- and L-values. Plant Soil 283:251–264.

Richardson, S.J., D.A. Peltzer, R.B. Allen, M.S. McGlone, and R.L. Parfitt. 2004. Rapid development of phosphorus limitation in temperate rainforest along the Franz Josef soil chronosequence. Oecologia 139:267–276.

Saggar, S., J.R. Bettany, and J.W.B. Stewart. 1981. Measurement of microbial sulfur in soil. Soil Biol. Biochem. 13:493–498.

Sato, S., D. Solomon, C. Hyland, Q.M. Ketterings, and J. Lehmann. 2005. Phosphorus speciation in manure and manure-amended soils using XANES spectroscopy. Environ. Sci. Technol. 39:7485–7491.

Schmalenberger, A., H.L. Drake, and K. Kusel. 2007. High unique diversity of sulfate-reducing prokaryotes characterized in a depth gradient in an acidic fen. Environ. Microbiol. 9:1317–1328.

Schmalenberger, A., S. Hodge, A. Bryant, M.J. Hawkesford, B.K. Singh, and M.A. Kertesz. 2008. The role of *Variovorax* and other *Comamonadaceae* in sulfur transformations by microbial wheat rhizosphere communities exposed to different sulfur fertilization regimes. Environ. Microbiol. 10:1486–1500.

Schmalenberger, A., S. Hodge, M.J. Hawkesford, and M.A. Kertesz. 2009. Sulfonate desulfurization in *Rhodococcus* from wheat rhizosphere communities. FEMS Microbiol. Ecol. 67:140–150.

Schmalenberger, A., and M.A. Kertesz. 2007. Desulfurization of aromatic sulfonates by rhizosphere bacteria—High diversity of the *asfA* gene. Environ. Microbiol. 9:535–545.

Siciliano, S.D., and J.J. Germida. 2005. Sulfur in soils/biological transformations, p. 85–90. *In* D. Hillel (ed.) Encyclopedia of soils in the environment. Elsevier, Amsterdam, the Netherlands.

Sims, J.T., and A.N. Sharpley. 2005. Phosphorus: Agriculture and the environment. Number 46 in the series Agronomy. ASA, CSSA and SSSA, Madison, WI.

Smernik, R.J., and W.J. Dougherty. 2007. Identification of phytate in phosphorus-31 nuclear magnetic resonance spectra: The need for spiking. Soil Sci. Soc. Am. J. 71:1045–1050.

Solomon, D., J. Lehmann, and C.E. Martinez. 2003. Sulfur K-edge XANES spectroscopy as a tool for understanding sulfur dynamics in soil organic matter. Soil Sci. Soc. Am. J. 67:1721–1731.

Stubner, S. 2004. Quantification of Gram-negative sulphate-reducing bacteria in rice field soil by 16S rRNA gene-targeted real-time PCR. J. Microbiol. Methods 57:219–230.

Suaréz, E.R., D.M. Pelletier, T.J. Fahey, P.M. Groffman, P.J. Bohlen, and M.C. Fisk. 2004. Ecosystem consequences of exotic earthworm invasion of north temperate forests. Ecosystems 7:28–44.

Tarafdar, J.C., R.S. Yadav, and R. Niwas. 2002. Relative efficiency of fungal intra- and extracellular phosphatases and phytase. J. Plant Nutr. Soil Sci. 165:17–19.

Thies, J.E., and J.M. Grossman. 2006. The soil habitat and soil ecology, p. 59–78. *In* N. Uphoff, A.S. Ball, E. Fernandes, H. Herren, O. Husson, M. Laing, C. Palm, J. Pretty, P. Sanchez, N. Sanginga, and J. Thies (eds.) Biological approaches to sustainable soil systems. CRC Press, Boca Raton, FL.

Tiunov, A.V., M. Bonkowski, J. Alphei, and S. Scheu. 2001. Microflora, protozoa and nematoda in *Lumbricus terrestris* burrow walls: A laboratory experiment. Pedobiologia 45:46–60.

Toor, G.S., S. Hunger, J.D. Peak, J.T. Sims, and D.L. Sparks. 2006. Advances in the characterization of phosphorus in organic wastes: Environmental and agronomic applications. Adv. Agron. 89:1–72.

Turner, B.L. 2008. Resource partitioning for soil phosphorus: A hypothesis. J. Ecol. 96:4:698–702.

Turner, B.L., L.M. Condron, S.J. Richardson, D.A. Peltzer, and V.J. Allison. 2007. Soil organic phosphorus transformations during pedogenesis. Ecosystems 10:1166–1181.

Turner, B.L., and P.M. Haygarth. 2001. Phosphorus solubilization in rewetted soils. Nature 411:258–258.

Turner, B.L., N. Mahieu, and L.M. Condron. 2003a. Quantification of *myo*-inositol hexakisphosphate in alkaline soil extracts by solution ^{31}P NMR spectroscopy and spectral deconvolution. Soil Sci. 168:469–478.

Turner, B.L., N. Mahieu, and L.M. Condron. 2003b. Phosphorus-31 nuclear magnetic resonance spectral assignments of phosphorus compounds in soil NaOH–EDTA extracts. Soil Sci. Soc. Am. J. 67:497–510.

Turner, B.L., and A.E. Richardson. 2004. Identification of *scyllo*-inositol phosphates in soil by solution phosphorus-31 nuclear magnetic resonance spectroscopy. Soil Sci. Soc. Am. J. 68:802–808.

Uroz, S., C. Calvaruso, M.P. Turpault, J.C. Pierrat, C. Mustin, and P. Frey-Klett. 2007. Effect of the mycorrhizosphere on the genotypic and metabolic diversity of the bacterial communities involved in mineral weathering in a forest soil. Appl. Environ. Microbiol. 73:3019–3027.

Vong, P.C., S. Piutti, E. Benizri, S. Slezack-Deschaumes, C. Robin, and A. Guckert. 2007. Water-soluble carbon in roots of rape and barley: Impacts on labile soil organic carbon, arylsulphatase activity and sulphur mineralization. Plant Soil 294:19–29.

Wagner, M., A. Loy, M. Klein, N. Lee, N.B. Ramsing, D.A. Stahl, and M.W. Friedrich. 2005. Functional marker genes for identification of sulfate-reducing prokaryotes. Methods Enzymol. 397:469–489.

Walker, T.W., and J.K. Syers. 1976. Fate of phosphorus during pedogenesis. Geoderma 15:1–19.

Wallenstein, M.D., and M.N. Weintraub. 2008. Emerging tools for measuring and modeling the *in situ* activity of soil extracellular enzymes. Soil Biol. Biochem. 40:2098–2106.

Welch, S.A., A.E. Taunton, and J.F. Banfield. 2002. Effect of microorganisms and microbial metabolites on apatite dissolution. Geomicrobiol. J. 19:343–367.

Wykoff, D.D., A.R. Grossman, D.P. Weeks, H. Usuda, and K. Shimogawara. 1999. Psr1, a nuclear localized protein that regulates phosphorus metabolism in *Chlamydomonas*. Proc. Natl. Acad. Sci. USA 96:15336–15341.

Wyss, M., R. Brugger, A. Kronenberger, R. Remy, R. Fimbel, G. Oesterhelt, M. Lehmann, and A.P.G.M. van Loon. 1999. Biochemical characterization of fungal phytases (*myo*-inositol hexakisphosphate phosphohydrolases): Catalytic properties. Appl. Environ. Microbiol. 65:367–373.

Yergeau, E., S. Kang, Z. He, J. Zhou, and G.A. Kowalchuk. 2007. Functional microarray analysis of nitrose and carbon cycling genes across an Antarctic latitudinal gradient. The ISME J. 1:163–179.

Yergeau, E., S.A. Schoondermark-Stolk, E.L. Brodie, S. Déjean, T.Z. DeSantis, O. Gonçalves, Y.M. Piceno, G.L. Andersen, and G.A. Kowalchuk. 2009. Environmental microarray analyses of Antarctic soil microbial communities. ISME J. 3:340–351.

Zhao, F.J., J.S. Knights, Z.Y. Hu, and S.P. McGrath. 2003. Stable sulfur isotope ratio indicates long-term changes in sulfur deposition in the Broadbalk experiment since 1845. J. Environ. Qual. 32:33–39.

Zhao, F.J., J. Lehmann, D. Solomon, M.A. Fox, and S.P. McGrath. 2006. Sulphur speciation and turnover in soils: Evidence from sulphur K-edge XANES spectroscopy and isotope dilution studies. Soil Biol. Biochem. 38:1000–1007.

Zhao, F.J., J. Wu, and S.P. McGrath. 1996. Soil organic sulphur and its turnover, p. 467–506. *In* A. Piccolo (ed.) Humic substances in terrestrial ecosystems. Elsevier Science, Amsterdam, the Netherlands.

26.2 Bacterial Transformations of Metals in Soil

Terence L. Marsh

David Long

Thomas Voice

26.2.1 Introduction

Soil ecosystems contain one of the most complex microbial communities on earth. The number of species per gram of soil is controversial (Torsvik et al., 1990, 2002; Dunbar et al., 2002; Torsvik and Ovreas, 2002) but it is certainly large. Each species may interact with metals in different ways. The number of different soil types and the resulting number of possible metal–bacteria complexes result in an interactive matrix of a size well beyond the scope of this chapter. Our objective therefore is to provide an overview of the microbial–metal interactions in soil rather than an exhaustive accounting of all known interactions. Given the physiological diversity of microbes and their ubiquity, they can influence nearly every abiotic process of the metal–soil repertoire through their own habitat-altering metabolic capabilities. In addition, they can reduce or oxidize metals directly through metabolism that can result in habitat alteration at landscape scales. Below we provide a brief description of the fate of metals in soils followed by the microbial processes that alter and transform the state of metals.

26.2.2 Historical Perspective

The work of Winogradsky (e.g., Waksman, 1953) established the concept of chemoautotrophy whereby microorganisms fix CO_2 and obtain energy from inorganic compounds, including metals. During the ensuing years, European microbiologists led the investigations on the "iron bacteria" (e.g., Molisch, 1910), those bacteria capable of oxidizing ferrous. Sixty years of largely anecdotal reports of bacteria capable of precipitating iron followed (see reviews by Pringsheim, 1949; Stokes, 1953) before bacterial iron oxidation was unequivocally linked to growth (Colmer et al., 1950). At around the same time, metal reduction was directly linked to bacterial growth in *Bacillus* (Roberts, 1947; Broomfield, 1954). These observations relating bacterial growth to the valence state and solubility of metals led to new views regarding the deposition of metals in, for example, banded iron formations and bog iron ore. An early monograph by Harder (1919) titled "Iron-depositing bacteria and their geologic relations" suggested the possibility that bacteria played a key role in the formation of Precambrian banded iron formations. A review by Lundgren and Dean (1979) provides a good discussion about early views of this still debated topic. Several viable scenarios have since been put forward that are based on biological activity including oxidation of ferrous by cyanobacteria-produced oxygen (Pierson and Parenteau, 2000) or more recently through direct oxidation by anoxygenic phototrophic bacteria (Klapper

et al., 2005). The origins of bog iron ores are similarly controversial. Early investigators separated marine ferromanganese nodules from freshwater deposits and bog iron ores (Ehrlich et al., 1972). Yet the distinction between lake deposits and bog iron ore was blurred by Arrhenius (1967), who pointed out that landscape succession of shallow lakes developing into vegetation covered bogs linked the deposits found in lakes with bog iron. The origins of bog iron remain a point of discussion with both direct (Gillette, 1961; Chapnick et al., 1982) and indirect (Dean and Ghosh, 1978) microbial activities invoked as well as abiotic scenarios (Varentsov, 1972). These paths of investigation led to a greater appreciation of the metabolic versatility of the bacteria and an increased understanding of their contribution to terrestrial geology.

26.2.3 The State and Fate of Metals in Soils

Metals are natural components of soil and rock minerals, but their presence in soil may also result from anthropogenic activities. Physical, chemical, and biological processes serve to transform and transport metals such that they may be found as precipitates, bound to organic or inorganic surfaces of the soil or colloids in solution, dissolved in the soil solution or in the soil atmosphere (McLean and Bledsoe, 1992). In solution, metals can exist as free cations, but many also form complexes with other organic and inorganic ligands such that the resulting molecule can be cationic, anionic, or neutral (Stumm and Morgan, 1996). The metals themselves may have multiple oxidation states, each of which differ in their ability to form complexes, and the complexes may participate in acid–base reactions (Evangelou, 1998). Both free metals and soluble complexes can form precipitates, either as pure solids, or as mixed solids when multiple metals coprecipitate (McBride, 1994). Soluble metals can also associate with surfaces, or become incorporated into soil components, through a variety of sorption, exchange, and diffusion mechanisms (Sposito, 1989). Precipitation and surface association are generally reversible reactions, with the result that metals can be released back into solution through dissolution, desorption, and exchange, if solution chemistry changes (Sparks, 1995). A few metals, metalloids, and their complexes, notably, Hg, As, Se, and Ra, are sufficiently volatile to be found in the soil atmosphere (Pecheyran et al., 1998). These processes are diagrammed in Figure 26.3.

Understanding the chemical form and soil phase of a metal in the soil environment can be extremely challenging due to the large number of possible reactions. It is theoretically possible to determine the predominate forms by solving a set of equations for the various acid–base, oxidation–reduction, precipitation–dissolution, sorption/ion-exchange and complexation reactions for each of the species in the system using analytical, graphic, or geochemical modeling techniques (Butler, 1998; Langmuir, 1998). These techniques may be limited, however, by the incomplete knowledge of the reacting species and phases, and of equilibrium and kinetic constants associated with the reactions of interest. Particularly problematic in this regard are reactions with organic matter, which are not well characterized chemically
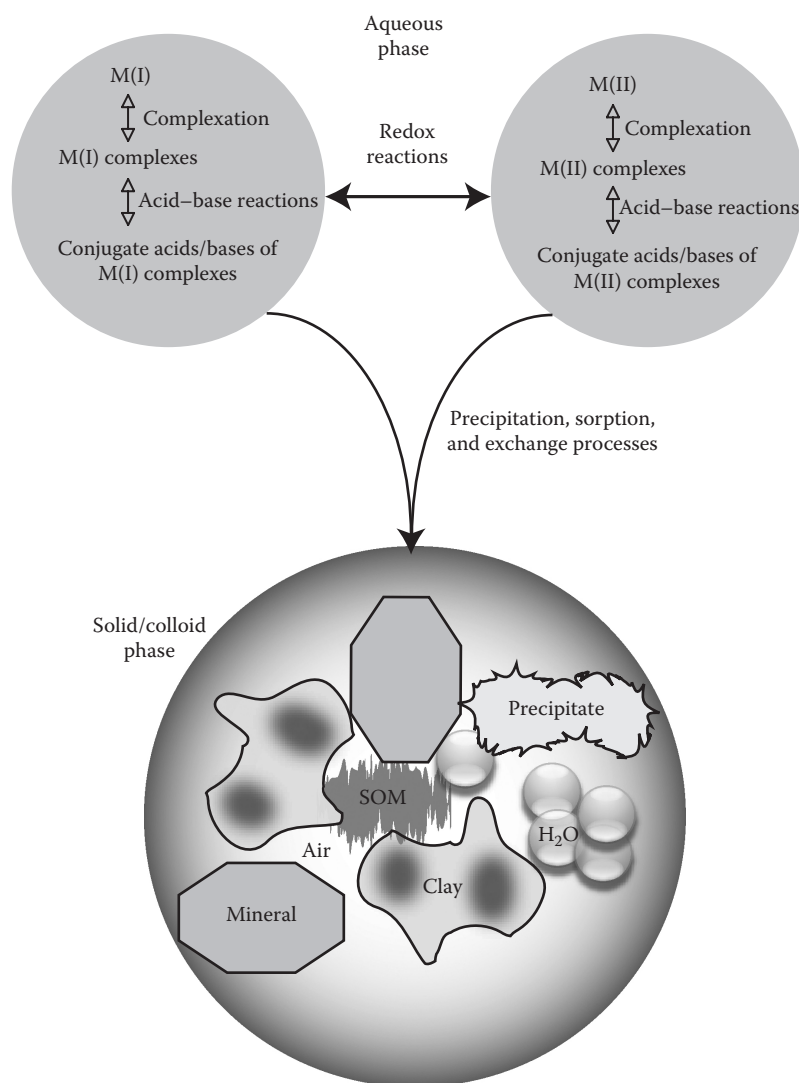
FIGURE 26.3 Processes affecting metal speciation in the soil environment. Uncomplexed metal ions are probably a rare commodity in the aqueous phase with metal complexes predominating and different complexes in equilibrium. Precipitation, sorption, and exchange processes remove metals from solution depositing them in the solid/colloidal phase of soil. Each process can be influenced by microbes through their abilities in (1) providing extracellular organic molecules and specific chelating agents for complexation, (2) binding metal ions to the cell surface and transporting them into the cytoplasm, (3) altering the pH of the local environment through metabolic reactions, (4) reducing or oxidizing metals directly through cell-to-metal contact, and (5) causing precipitation of metallic colloids.

and highly variable such that accurate constants are not generally available (Dudal and Gerard, 2004).

It is within this "dynamic equilibrium in a polyphasic system" (Hopkins, 1913) of metal–soil–solvent that microbes impose their nutritional requirements for metals. Success at this endeavor means that they will grow, metabolize, and evolve. Hence, these are not casual interactions but robust microbial activities aimed at acquiring the necessary metals for growth under conditions of low metal concentrations or detoxifying high metal concentrations, so that growth is possible. Within these two extreme conditions, the diversity of the microbial world has evolved sophisticated approaches to take advantage of metals in soils.

26.2.4 Metals and Microbes

The number of ways that microbes interact with metals is finite and provides an organizational scheme. These interactions are presented graphically in Figure 26.4 and summarized as follows:

- Direct association: Microbes frequently form biofilms and establish microcolonies on surfaces and crevasses of rocks, minerals, and metals. In so doing, these communities actually contribute to the formation of soils by accelerating degradation of parent material. In addition, microbial communities are known to form biofilms and microcolonies on and within soil aggregates as well as on the roots of plants. Thus, the microbial community
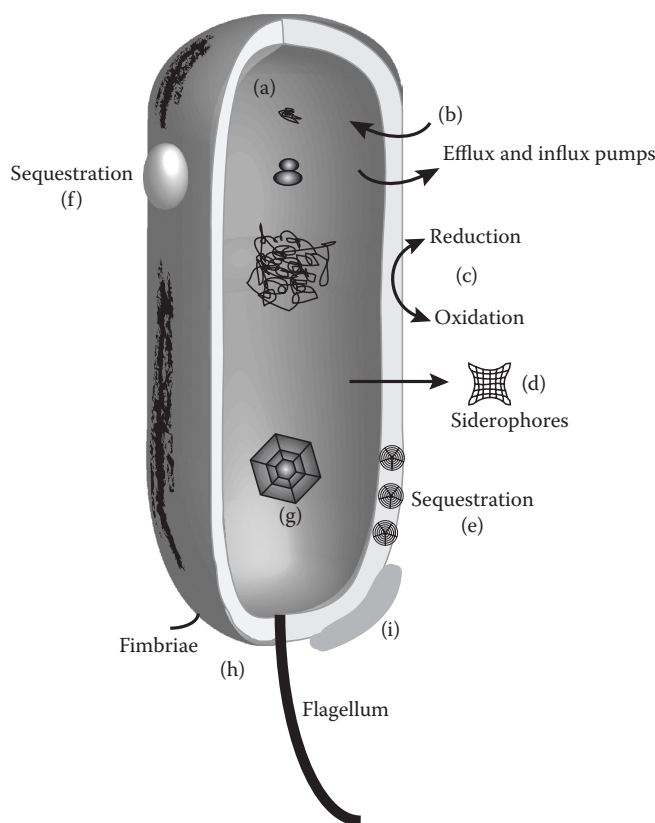
FIGURE 26.4 The interactions of microbes with metals. Metals are essential for the structure and function of macromolecules including DNA, RNA, and protein (a). The requirement for metals has led to elaborate systems of homeostasis based on influx and efflux pumps (b). Metals are used as a source of electrons by some populations and as an electron sink by others (c). Microbial cells will actively strip metals from their abiotic environment using excreted siderophores to chelate and transport metals into cells (d). Microbes have also developed techniques to sequester metals. Extracellular sequestration (e and f) locally reduces toxic levels while intracellular sequestration within bacterioferritin (g) allows cells to bank an essential nutrient. Both the expression of fimbriae and flagella (h) can be regulated by metal concentration in the environment, and both may be involved with the formation of biofilm on metal surfaces. Finally, extracellular polysaccharide (i) can bind metals to the surface of cells.

partitions into populations that are either attached to soil particulates and plant surfaces or pelagic within pore water.

- Metals as required nutrients: Metals are required by all known microbes as cofactors for a diverse collection of enzymatically driven reactions as well as an essential component in the folding and conformation of biological macromolecules. Because of these requirements, microbes have developed mechanisms, sometimes elaborate, to recruit metals from the environment. Microbiologically, these processes are described as assimilatory, referring to the assimilation of the substance, in this case metals, into cellular material. Because of the great biological importance of metals, microbial populations have evolved

sensory systems linked to gene expression to detect metals and respond accordingly.

- Metals as a source or sink for electrons: Metals can be used by microbes in a dissimilatory manner as well. In this case, metals serve as electron acceptors or donors in cell-mediated reactions that contribute to the bioenergetics of the cell.
- Sequestration and resistance: In environments with high concentrations of metals, microbes have developed survival mechanisms that diminish the intracellular toxicity of metals.

26.2.4.1 Direct Association

Since the codification of microbial biofilms as a mode of life (Costerton et al., 1995; Ghannoum and O'Toole, 2001; Stoodley et al., 2002), many communities with populations bound to a substratum have been described as a biofilm. While the existence of biofilm in soils under natural conditions is debated, populations that adhere to soil particles have been identified (Mills and Powelson, 1996) and bacterial isolates of soil have been experimentally determined to form biofilms in the laboratory (Burmølle et al., 2007). Microbes can form biofilms on soil particulates containing metallic compounds (Lower et al., 2001; Luttge et al., 2005) as well as on pure metal (Mansfeld, 2007; Sheng et al., 2008; Waters et al., 2008). Moreover, biofilm matrices and extracellular polysaccharides bind metal ions (Ghannoum and O'Toole, 2001; Teitzel and Parsek, 2003; Raize et al., 2004; Quintelas et al., 2008; Ueshima et al., 2008). It has been noted as well that cells within a biofilm matrix are more resistant to high concentrations of metals (Harrison et al., 2007). Hence, life as biofilm attached to surfaces, including minerals and metals, may provide distinct fitness advantages.

Gorbushina (2007) describes the "subaerial biofilm" that covers the surface of rocks as a relatively simple community composed of fungi, algae, cyanobacteria, and heterotrophic Bacteria and Archaea. Walker and Pace (2007) provides an excellent review that summarizes what we know of community composition of endolithic communities that are typically composed of *Cyanobacteria*, *Actinobacteria*, *Proteobacteria*, and *Bacteriodetes* with other phylogenetic groups found occasionally. It is of course the metabolism of these communities that ultimately alters the rock and contributes to the genesis of soil. Omelon et al. (2007) have correlated endolithic community structure with metal composition of the habitat and concluded that *Cyanobacteria* dominated under higher pH with elevated calcium and magnesium whereas fungi and algae dominated lower pH habitats with elevated iron and aluminum. This suggests that metals were a part of the selective pressures on developing communities. Ultimately, colonization of this type leads to biogeochemical weathering (Barker et al., 1997).

26.2.4.2 Metals as Required Nutrients

Because metals are essential for enzymatic processes and macromolecular structure within microbial cells, elaborate systems for

metal capture and sequestration have evolved. For example, membrane-located influx pumps have been identified for essential metals including iron (Andrews et al., 2003), magnesium (Maguire, 1992), and copper (Lu et al., 1999; Lu and Solioz, 2002). These pumps require contact of the cell with a free or complexed ion in solution. The metal is transported across the outer membrane (Gram-negative cell) by specific energy-requiring receptors. A periplasmic binding protein guides the metal complex to the cytoplasmic membrane transport system and transport to the cytoplasm follows. Many bacteria release a siderophore or hemophore (chelating agent) into the environment to assist in the acquisition of metals (Wandersman and Delepelaire, 2004). Siderophores are low-molecular-weight molecules while hemophores are proteins. Both have relatively high-affinity constants for metals and siderophores can aggressively strip metals from the immediate habitat, a biomediated dissolution of metals. Siderophores complexed with the target metal ion are reassimilated by the cell using specific receptors on the surface. It has been argued that free iron is at low concentrations in every environment. Hence, most bacteria have evolved tactics, for example, siderophores or hemophores, to find and sequester iron. Approximately 500 different siderophores have been described (Wandersman and Delepelaire, 2004), underscoring bacterial diversity and the potential complexity of the soil environment. Because of the importance of iron in cell metabolism, some species have evolved intracellular storage systems that are constructed when iron is abundant. Intracellular storage of iron is mediated by caged structures composed of protein. Three versions have been described, ferritin, bacterioferritin, and Dps proteins (Andrews et al., 2003). These proteins form intracellular multisubunit structures capable of storage of up to 3000 iron atoms.

26.2.4.3 Metals as a Source or Sink for Electrons

Metals are of course essential to the redox chemistry of cells through the coordinated metals within enzymes. What has become recently apparent is the importance of reduction and oxidation of metals in a dissimilatory manner, that is, as donors or acceptors of electrons in reactions that are pivotal to cell bioenergetics (Lovley, 1993; Nealson and Little, 1997; Lloyd, 2003; Lovley et al., 2004). A relatively recent text (Lovley, 2000) provides a good overview of the use of metals as electron donors or acceptors. Perhaps the best-studied organisms capable of dissimilatory metal reduction are *Geobacter* and *Shewanella*. *Geobacter* requires direct contact with insoluble Fe(III) before it can pass metabolic electrons to the metal, reducing it to Fe(II). In contrast, *Shewanella* constructs and excretes soluble quinones or flavin and riboflavins that carry the terminal metabolic electron to Fe(III), perhaps some distance from the cell (Newman and Kolter, 2000; von Canstein et al., 2008). During this process, irregularities on the mineral surface are thought to play a large role in stimulating the process (Brown et al., 1999). These irregularities are the results of imperfect crystallization resulting in series of steps or terraces as one surface gives way to another with each step, providing point defects in the form of kinks in the unit structure. These defects are relevant to the reactivity

of the metal oxide surfaces because they present regions where the energy is perturbed. Rosso et al. (2003) have shown that the dissolution of hematite by *Shewanella*, when using a nonlocal electron transfer system like quinones, can take advantage of structural defects distant from the attached cell. Investigations such as these provide insight into the interactions of microbe and metal at the molecular level.

Finally, microbially synthesized nanowires have been implicated in direct electron transfer, opening a new field of study (Reguera et al., 2005, 2006; Gorby et al., 2006). The specific molecular events leading to the direct transfer of electrons from the cell surface to insoluble metal oxides are not well elucidated but outer membrane cytochromes have been invoked. However, the problem electron transfer has risen to the level of a grand challenge (Fredrickson and Zachara, 2008) and so progress is anticipated. Clearly the microbe must come into direct contact with the mineral surface through an attachment between the outer envelop of the microbe with the mineral surface. The oxidation of metals by microbes is not as well understood. Emerson (2000) provides a good overview of Fe and Mn oxidizers and the geochemical problems they face.

The presence of metals in an environment can dictate a particular life style for microbes. Not only can microbes excrete metal chelating compounds, but also upon sensing the presence of metals (iron in this case) behavioral traits can be stimulated. In the case of *Geobacter* (Childers et al., 2002), insoluble Fe(III) oxide is used as an electron acceptor and soluble Fe(II) is an inducer of flagella and pili synthesis. Hence, it is argued from an ecological perspective that elevated concentrations of Fe(II) are indicative of depleted Fe(III); hence, motility, induced by increasing Fe(II) concentrations, provides the means to relocate to more Fe(III)-rich environs.

26.2.4.4 Resistance and Sequestration

Development of resistance to metals by microbes (Nies, 1992, 1999; Cervantes and Gutierrezcorona, 1994; Silver and Phung, 1996; Barkay et al., 2003) has resulted from long-term exposure to elevated concentrations of metals. Natural environments, for example, serpentine soils (Haferburg and Kothe, 2007), provide habitats for selection under these conditions. Microbes have evolved numerous mechanisms to deal with elevated metal concentrations including efficient and numerous efflux pumps (Nies, 2003). The reference organism for metal resistance is *Cupriavidus metallidurans* (aka *Ralstonia metallidurans*). Originally isolated from metal-contaminated soils, this bacterium has an enormous repertoire of resistance determinants (Mergeay et al., 2003; Noel-Georis et al., 2004). So accomplished is this organism in its interactions with metals that a special issue of *Antonie Van Leeuwenhoek* devoted to *C. metallidurans* is planned (Silver and Mergeay, 2009).

Maintaining nontoxic intracellular concentrations of metals can be accomplished by sequestration as well as pumping. Several examples of microbes capable of binding metals to their extracellular matrix have been described (Sar and D'Souza, 2001; Raize et al., 2004; Lu et al., 2006; Quintelas et al., 2008; Kazy et al., 2009).

In some cases, the binding of metals to dead cells is as effective as it is to live cells (Raize et al., 2004). Not all bacteria have high metal-binding capabilities but clearly those that do can bind significant quantities on their exterior surfaces (Yang et al., 2010). Acting as a large chelator, bacteria can directly influence the transport of metals in the subsurface. Another unusual example of sequestration is the formation of surface blebs or vesicles. Several examples of bacteria capable of producing extracellular vesicles that are enriched in metal have been described (Oremland et al., 2004). This sequestration strategy can accumulate significant quantities of metals, thus altering the distribution of metals locally. In some cases, not only are the metals sequestered but they are also reduced within the vesicle by the bacterium and shifted from the soluble to the insoluble phase.

26.2.5 Discussion

Soils are essential to our survival. However, as our population grows, we are placing soils under increased stress that frequently diminishes their fertility and longevity. The transformation of metals by microbes within the soil environment is a critical process, which, if understood more completely, would assist in the design of sustainable practices. The dynamic equilibrium of metals in soils can be influenced by microbes at every step depicted in Figure 26.3, in both aqueous and solid/colloidal phases. Simply considering the enormous number of siderophores already described makes the combined equilibria of all metals and siderophores, in addition to inorganic binding ligands, a fairly complex equation to solve. Recall that there are approximately 10^9 microbial cells per gram of soil. Haferburg and Kothe (2007) correctly point out that when due consideration is placed on the uncultured majority of the microbes residing in soil, the diversity of known siderophores will increase and the complexity of the equilibria calculations will increase concomitantly. If this is insufficient complexity to challenge the intellect, we might add yet another layer. We have considered the bacteria in our discussions, leaving the viral world unmentioned. Estimates place the concentration of virus at approximately the same order of magnitude as the bacteria in soils (Srinivasiah et al., 2008; Williamson et al., 2010; Section 24.1). Wayne et al. (1975) pointed out three decades ago that bacteriophage Φ80 and an iron siderophore share the same binding site on *Escherichia coli*. Hence, in this one case, producing a siderophore makes one more susceptible to phage. Moreover, the mature virion of T4 has been found to bind metals (Thomassen et al., 2003), adding perhaps many more "chelators" to the mix. One might be led to conclude that the greatest influence on the state of metals in soils is biological.

References

Andrews, S.C., A.K. Robinson, and F. Rodriguez-Quinones. 2003. Bacterial iron homeostasis. FEMS Microbiol. Rev. 27:215–237.

Arrhenius, O. 1967. Ore, iron, artifacts and corrosion. Sveriges geologiska undersökning Series C 626:39.

Barkay, T., S.M. Miller, and A.O. Summers. 2003. Bacterial mercury resistance from atoms to ecosystems. FEMS Microbiol. Rev. 27:355–384.

Barker, W.W., S.A. Welch, and J.F. Banfield. 1997. Biogeochemical weathering of silicate minerals, p. 391–428. *In* J.F. Banfield and K.H. Nealson (eds.) Reviews in mineralogy. Vol. 35. Mineralogical Society of America, Washington, DC.

Broomfield, S.M. 1954. The reduction of iron oxide by bacteria. J. Soil Sci. 5:129–139.

Brown, G.E., V.E. Henrich, W.H. Casey, D.L. Clark, C. Eggleston, A. Felmy, D.W. Goodman et al. 1999. Metal oxide surfaces and their interactions with aqueous solutions and microbial organisms. Chem. Rev. 99:77–174.

Burmølle, M., L.H. Hansen, and S.J. Sørensen. 2007. Establishment and early succession of a multispecies biofilm composed of soil bacteria. Microb. Ecol. 54:352–362.

Butler, J.N. 1998. Ionic equilibrium: Solubility and pH calculations. John Wiley & Sons, New York.

Cervantes, C., and F. Gutierrezcorona. 1994. Copper resistance mechanisms in bacteria and Fungi. FEMS Microbiol. Rev. 14:121–137.

Chapnick, S.D., W.S. Moore, and K.H. Nealson. 1982. Microbially mediated manganese oxidation in a freshwater lake. Limnol. Oceanogr. 27:1004–1014.

Childers, S.E., S. Ciufo, and D.R. Lovley. 2002. *Geobacter metallireducens* accesses insoluble Fe(III) oxide by chemotaxis. Nature 416:767–769.

Colmer, A.R., K.L. Temple, and M.E. Hinkle. 1950. An iron-oxidizing bacterium from the acid drainage of some bituminous coal mines. J. Bacteriol. 59:317–328.

Costerton, J.W., D. Lewandowski, D.E. Caldwell, D.R. Korber, and H.M. Lappin-Scott. 1995. Microbial biofilms. Annu. Rev. Microbiol. 49:711–745.

Dean, W.E., and S.K. Ghosh. 1978. Factors contributing to the formation of ferromanganese nodules in Oneida Lake, New York. J. Res. U.S. Geol. Surv. 6:231–240.

Dudal, Y., and F. Gerard. 2004. Accounting for natural organic matter in aqueous chemical equilibrium models: A review of the theories and applications. Earth Sci. Rev. 66:199–216.

Dunbar, J., S.M. Barns, L.O. Ticknor, and C.R. Kuske. 2002. Empirical and theoretical bacterial diversity in four Arizona soils. Appl. Environ. Microbiol. 68:3035–3045.

Ehrlich, H.L., W.C. Ghiorse, and G.L. Johnson. 1972. Distribution of microbes in manganese nodules from the Atlantic and Pacific oceans. Dev. Ind. Microbiol. 13:57–65.

Emerson, D. 2000. Microbial oxidation of Fe(II) and Mn(II) at circumneutral pH, p. 31–52. *In* D.R. Lovley (ed.) Environmental microbe-metal interactions. ASM Press, Washington, DC.

Evangelou, V.P. 1998. Environmental soil and water chemistry. John Wiley & Sons, New York.

Fredrickson, J.K., and J.M. Zachara. 2008. Electron transfer at the microbe–mineral interface: A grand challenge in biogeochemistry. Geobiology 6:245–253.

Ghannoum, M., and G.A. O'Toole (eds.) 2001. Microbial biofilms. ASM Press, Washington, DC.

Gillette, N.J. 1961. Oneida lake pancakes. N.Y. State Conserv. 41:21.

Gorbushina, A.A. 2007. Life on the rocks. Environ. Microbiol. 9:1613–1631.

Gorby, Y.A., S. Yanina, J.S. McLean, K.M. Rosso, D. Moyles, A. Dohnalkova, T.J. Beveridge et al. 2006. Electrically conductive bacterial nanowires produced by *Shewanella oneidensis* strain MR-1 and other microorganisms. Proc. Natl. Acad. Sci. USA 103:11358–11363.

Haferburg, G., and E. Kothe. 2007. Microbes and metals: Interactions in the environment. J. Basic Microbiol. 47:453–467.

Harder, E.C. 1919. Iron-depositing bacteria and their geologic relations, p. 1–89. U. S. Geological Survey Professional Paper 113.

Harrison, J.J., H. Ceri, and R.J. Turner. 2007. Multimetal resistance and tolerance in microbial biofilms. Nat. Rev. Microbiol. 5:928–938.

Hopkins, F.G. 1913. On the dynamic side of biochemistry. Br. Med. J. 2:713–717.

Kazy, S.K., S.F. D'Souza, and P. Sar. 2009. Uranium and thorium sequestration by a *Pseudomonas* sp.: Mechanism and chemical characterization. J. Hazard. Mater. 163:65–72.

Klapper, A., C. Pasquero, K.O. Konhause, and D.K. Newman. 2005. Deposition of banded iron formations by anoxygenic phyototrophic Fe(II)-oxidizing bacteria. Geology 33:865–868.

Langmuir, D. 1998. Aqueous environmental geochemistry. Prentice-Hall, Upper Saddle River, NJ.

Lloyd, J.R. 2003. Microbial reduction of metals and radionuclides. FEMS Microbiol. Rev. 27:411–425.

Lovley, D.R. 1993. Dissimilatory metal reduction. Annu. Rev. Microbiol. 47:263–290.

Lovley, D.R. (ed.) 2000. Environmental microbe-metal interactions. ASM Press, Washington, DC.

Lovley, D.R., D.E. Holmes, and K.P. Nevin. 2004. Dissimilatory Fe(III) and Mn(IV) reduction. Adv. Microb. Physiol. 49:219–286.

Lower, S.K., M.F. Hochella, Jr., and T.J. Beveridge. 2001. Bacterial recognition of mineral surfaces: Nanoscale interactions between Shewanella and alpha-FeOOH. Science 292:1360–1363.

Lu, Z.H., P. Cobine, C.T. Dameron, and M. Solioz. 1999. How cells handle copper: A view from microbes. J. Trace Elem. Exp. Med. 12:347–360.

Lu, W.B., J.J. Shi, C.H. Wang, and J.S. Chang. 2006. Biosorption of lead, copper and cadmium by an indigenous isolate *Enterobacter* sp. J1 possessing high heavy-metal resistance. J. Hazard. Mater. B 134:80–86.

Lu, Z.H., and M. Solioz. 2002. Bacterial copper transport. Adv. Protein Chem. 60:93–121.

Lundgren, D.G., and W. Dean. 1979. Biogeochemistry of iron, p. 211–251. *In* P.A. Trudinger and D.J. Swaine (eds.) Biogeochemical cycling of mineral-forming elements. Vol. 3. Elsevier Scientific Publishing Co., Amsterdam, the Netherlands.

Luttge, A., L. Zhang, and K.H. Nealson. 2005. Mineral surfaces and their implications for microbial attachment: Results from Monte Carlo simulations and direct surface observations. Am. J. Sci. 305:766–790.

Maguire, M.E. 1992. MgtA and MgtB: Prokaryotic P-type ATPases that mediate Mg^{2+} influx. J. Bioenerg. Biomembr. 24:319–328.

Mansfeld, F. 2007. The interaction of bacteria and metal surfaces. Electrochim. Acta 52:7670–7680.

McBride, M.B. 1994. Environmental chemistry of soils. Oxford University Press, New York.

McLean, J.E., and B.E. Bledsoe. 1992. Behavior of metals in soils. EPA Groundwater Issue, EPA/540/S-92/018. USEPA, Washington, DC.

Mergeay, M., S. Monchy, T. Vallaeys, V. Auquier, A. Benotmane, P. Bertin, S. Taghavi, J. Dunn, D. van der Lelie, and R. Wattiez. 2003. *Ralstonia metallidurans*, a bacterium specifically adapted to toxic metals: Towards a catalogue of metal-responsive genes. FEMS Microbiol. Rev. 27:385–410.

Mills, A.L., and D.K. Powelson. 1996. Bacterial interactions with surfaces in soils. *In* M. Fletcher (ed.) Bacterial adhesion. Wiley-Liss, New York.

Molisch, H. 1910. Die Eisenbakterien. Gustav Fischer, Jena, Germany.

Nealson, K.H., and B. Little. 1997. Breathing manganese and iron: Solid-state respiration. Adv. Appl. Microbiol. 45:213–239.

Newman, D.K., and R. Kolter. 2000. A role for excreted quinones in extracellular electron transfer. Nature 405:94–97.

Nies, D.H. 1992. Resistance to cadmium, cobalt, zinc, and nickel in microbes. Plasmid 27:17–28.

Nies, D.H. 1999. Microbial heavy-metal resistance. Appl. Microbiol. Biotechnol. 51:730–750.

Nies, D.H. 2003. Efflux-mediated heavy metal resistance in prokaryotes. FEMS Microbiol. Rev. 27:313–339.

Noel-Georis, I., T. Vallaeys, R. Chauvaux, S. Monchy, P. Falmagne, M. Mergeay, and R. Wattiez. 2004. Global analysis of the *Ralstonia metallidurans* proteome: Prelude for the large-scale study of heavy metal response. Proteomics 4:151–179.

Omelon, C.R., W.H. Pollard, and F.G. Ferris. 2007. Inorganic species distribution and microbial diversity within high arctic cryptoendolithic habitats. Microb. Ecol. 54:740–752.

Oremland, R.S., M.J. Herbel, J.S. Blum, S. Langley, T.J. Beveridge, P.M. Ajayan, T. Sutto, A.V. Ellis, and S. Curran. 2004. Structural and spectral features of selenium nanospheres produced by Se-respiring bacteria. Appl. Environ. Microbiol. 70:52–60.

Pecheyran, C., C.R. Quetel, M.M. Lecyer, and O.F.X. Donard. 1998. Simultaneous determination of volatile metal (Pb, Hg, Sn, In, Ga) and nonmetal species (Se, P, As) in different atmospheres by cryofocusing and detection by ICPMS. Anal. Chem. 70:2639–2645.

Pierson, B.K., and M.N. Parenteau. 2000. Phototrophs in high iron microbial mats: Microstructure of mats in iron depositing hot springs. FEMS Microb. Ecol. 32:181–196.

Pringsheim, E.G. 1949. The filamentous bacteria *Sphaerotilus*, *Leptothrix*, *Cladothrix*, and their relation to iron and manganese. Philos. Trans. R. Soc. London, Ser. B 233:453–482.

Quintelas, C., B. Fernandes, J. Castro, H. Figueiredo, and T. Tavares. 2008. Biosorption of Cr(VI) by three different bacterial species supported on granular activated carbon—A comparative study. J. Hazard. Mater. 153:799–809.

Raize, O., Y. Argaman, and S. Yannai. 2004. Mechanisms of biosorption of different heavy metals by brown marine macroalgae. Biotechnol. Bioeng. 87:451–458.

Reguera, G., K.D. McCarthy, T. Mehta, J.S. Nicoll, M.T. Tuominen, and D.R. Lovley. 2005. Extracellular electron transfer via microbial nanowires. Nature 435:1098–1101.

Reguera, G., K.P. Nevin, J.S. Nicoll, S.F. Covalla, T.L. Woodard, and D.R. Lovley. 2006. Biofilm and nanowire production leads to increased current in *Geobacter sulfurreducens* fuel cells. Appl. Environ. Microbiol. 72:7345–7348.

Roberts, J.L. 1947. Reduction of ferric hydroxide by strains of *Bacillus polymyxa*. Soil Sci. 63:135–140.

Rosso, K.M., J.M. Zachara, J.K. Fredrickson, Y.A. Gorby, and S.C. Smith. 2003. Nonlocal bacterial electron transfer to hematite surfaces. Geochim. Cosmochim. Acta 67:1081–1087.

Sar, P., and S.F. D'Souza. 2001. Biosorptive uranium uptake by a strain: Characterization and equilibrium studies. J. Chem. Technol. Biotechnol. 76:1286–1294.

Sheng, X., Y.P. Ting, and S.O. Pehkonen. 2008. The influence of ionic strength, nutrients and pH on bacterial adhesion to metals. J. Colloid Interface Sci. 321:256–264.

Silver, S., and M. Mergeay. 2009. Introduction to a special Festschrift issue celebrating the microbiology of *Cupriavidus metallidurans* strain CH34. Antonie Leeuwenhoek 96:113–114.

Silver, S., and L.T. Phung. 1996. Bacterial heavy metal resistance: New surprises. Annu. Rev. Microbiol. 50:753–789.

Sparks, D.L. 1995. Environmental soil chemistry. Academic Press, San Diego, CA.

Sposito, G. 1989. The chemistry of soils. Oxford University Press, New York.

Srinivasiah, S., J. Bhavsar, K. Thapar, M. Liles, T. Schoenfeld, and K.E. Wommack. 2008. Phages across the biosphere: Contrasts of viruses in soil and aquatic environments. Res. Microbiol. 159:349–357.

Stokes, J.L. 1953. Studies on the filamentous sheathed iron bacterium. *Sphaerotilus natans*. J. Bacteriol. 67:278–291.

Stoodley, P., K. Sauer, D.G. Davies, and J.W. Costerton. 2002. Biofilms as complex differentiated communities. Annu. Rev. Microbiol. 56:187–209.

Stumm, W., and J.J. Morgan. 1996. Aquatic chemistry. 3rd edn. John Wiley & Sons, New York.

Teitzel, G.M., and M.R. Parsek. 2003. Heavy metal resistance of biofilm and planktonic *Pseudomonas aeruginosa*. Appl. Environ. Microbiol. 69:2313–2320.

Thomassen, E., G. Gielen, M. Schutz, G. Schoehn, J.P. Abrahams, S. Miller, and M.J. van Raaij. 2003. The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold. J. Mol. Biol. 331:361–373.

Torsvik, V., J. Goksoyr, and F.L. Daae. 1990. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. 56:782–787.

Torsvik, V., and L. Ovreas. 2002. Microbial diversity and function in soil: From genes to ecosystems. Curr. Opin. Microbiol. 5:240–245.

Torsvik, V., L. Ovreas, and T.F. Thingstad. 2002. Prokaryotic diversity—Magnitude, dynamics, and controlling factors. Science 296:1064–1066.

Ueshima, M., B.R. Ginn, E.A. Haack, J.E.S. Szymanowski, and J.B. Fein. 2008. Cd adsorption onto *Pseudomonas putida* in the presence and absence of extracellular polymeric substances. Geochim. Cosmochim. Acta 72:5885–5895.

Varentsov, I.M. 1972. Geochemical studies on the formation of iron-manganese nodules and crusts in recent basins. Usta Mineral. Petrogr. Szeged. 20:363–381.

von Canstein, H., J. Ogawa, S. Shimizu, and J.R. Lloyd. 2008. Secretion of flavins by *Shewanella* species and their role in extracellular electron transfer. Appl. Environ. Microbiol. 74:615–623.

Waksman, S.A. 1953. Sergei N. Winogradsky: His life and work; the story of a great bacteriologist. Rutgers University Press, New Brunswick, NJ.

Walker, J.J., and N.R. Pace. 2007. Endolithic microbial ecosystems. Annu. Rev. Microbiol. 61:331–347.

Wandersman, C., and P. Delepelaire. 2004. Bacterial iron sources: From siderophores to hemophores. Annu. Rev. Microbiol. 58:611–647.

Waters, M.S., C.A. Sturm, M.Y. El-Naggar, A. Luttge, F.E. Udwadia, D.G. Cvitkovitch, S.D. Goodman, and K.H. Nealson. 2008. In search of the microbe/mineral interface: Quantitative analysis of bacteria on metal surfaces using vertical scanning interferometry. Geobiology 6:254–262.

Wayne, R., and J.B. Neilands. 1975. Evidence for common binding sites for ferrichrome compounds and bacteriophage phi 80 in the cell envelope of *Escherichia coli*. J. Bacteriol. 121:497–503.

Williamson, K., S. Srinivasiah, and K.E. Wommack. 2012. Microbiota, p. xxx. *In* P.M. Huang, Y. Li, and M.E. Sumner (eds.) Handbook of soil science: Properties and processes. Vol. 1. Taylor & Francis, Boca Raton, FL.

Yang, F., D.A. Pecina, S.D. Kelly, S.-H. Kim, M. Kenneth, K.M. Kemner, D.T. Long, and T.L. Marsh. In press. Biosequestration via cooperative binding of copper by *Ralstonia pickettii*. Environ. Technol. 31:1045–1060.

26.3 Microbially Mediated Processes: Decomposition

E.G. Gregorich

H.H. Janzen

> *In its life, its death, and its decomposition an organism circulates its atoms through the biosphere over and over again.*
>
> **Vladimir Vernadsky**
> *The Biosphere, 1926 (1998, p. 56)*

26.3.1 Overview

Decomposition is the breakdown of complex organic materials, ultimately into inorganic constituents. The process is mediated mostly by microorganisms, which derive energy and nutrients from the process. Consequently, decomposition is influenced by any factor that governs biological processes: primarily water, temperature, and radiation, but also chemical and physical attributes of soil, such as pH, texture, mineralogy, and structure.

Decomposition is critical for continuing life on earth, because it allows the recycling and reuse of carbon (C) and nutrients by biota at local and global scales (Figure 26.5). Although decay is sometimes seen as destructive, it is really a regenerative process,

releasing previously fixed C back to air for renewed photosynthesis, and liberating nitrogen (N), phosphorus, sulfur, and micronutrients to nourish new generations of plants and other biota. If decomposition were to stop, elements essential for life would soon be tied up, extinguishing all organisms (Jenkinson, 1981).

26.3.2 Substrates for Decomposition in Soil

The substrates for decomposition in soil form a continuum, from fresh plant residues to humified organic matter. Fresh plant residues comprise mostly resistant polymers (e.g., cellulose and lignin) with few essential nutrients, but also N-rich cytoplasmic materials such as sugars, amino acids, proteins, nucleic acids, and peptides. Decomposition of fresh plant residues is regulated by three interacting factors: soil organisms, the quality of chemical composition of the residues, and the physical/chemical environment (Figure 26.6). These factors control not just the rate of the process, but also the final products of decomposition. Humified organic matter, at the other end of the spectrum, is a complex product of the condensation reactions of microbial metabolites and refractory plant residues. The decomposition of this material is much slower than that of fresh plant litter, because of the stabilizing effects of its inherent recalcitrance, its association with soil minerals, or its inaccessibility within soil aggregates. Thus, while the rate of plant litter decomposition (primary decomposition) is governed by factors dictating
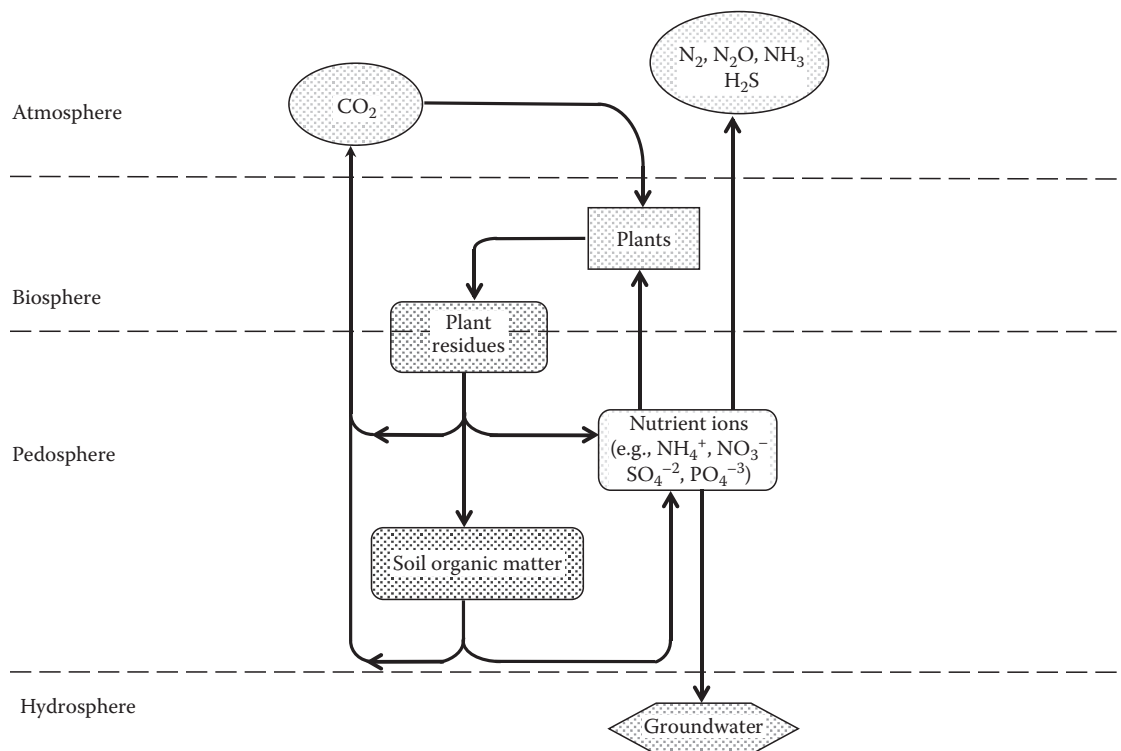
FIGURE 26.5   The process of decomposition of organic residues is central to biogeochemical cycles encompassing the atmosphere, biosphere, pedosphere, and the hydrosphere.
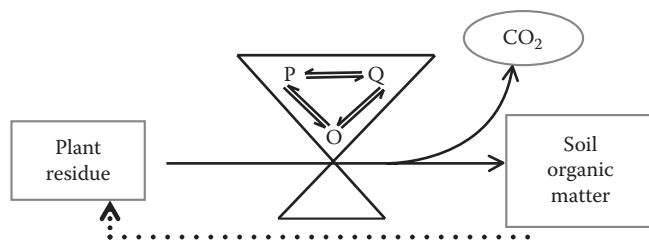
FIGURE 26.6 Decomposition of fresh organic residues in soil is governed by the interaction of three components: the organisms (O), the quality or chemical composition of the organic substrate (Q), and the physical/chemical environment (P). These components not only determine the rate of the process but also to some extent the final products of decomposition. (Adapted from Swift, M.J., O.W. Heal, and J.M. Anderson. 1979. Decomposition in terrestrial ecosystems. University of California Press, Berkeley, CA.)

microbial activity, that of humus decomposition (secondary decomposition) is controlled more by the intrinsic properties and interactions of the material itself.

26.3.3 Priming Effects of Fresh Residue Inputs to Soil

Soils contain vast amounts of organic matter, most of which is refractory humified material not readily decomposable by microorganisms; substrate availability in soils is therefore both scarce and infrequent. Adding fresh residues often stimulates microbial activity, measurable in accelerated carbon dioxide (CO_2) efflux or N mineralization. Some of this enhanced decay may occur at the expense of humified soil organic matter—the so-called priming effect. Because the turnover of soil organic matter is not measured directly, the priming effects may only be "apparent," reflecting activation of microbial metabolism and higher microbial biomass turnover (Kuzyakov et al., 2000). Mechanisms for "real" priming effects remain unclear but may be linked to accelerated native organic matter turnover from relief of nutrient deficiency (especially that of N) by decay of fresh incoming residues. Alternatively, a burst of energy from decomposable compounds in fresh residues may elicit greater synthesis of enzymes used to hydrolyze more recalcitrant compounds in soil organic matter, that is, cometabolism of soil organic matter (Blagodatskaya and Kuzyakov, 2008).

Research on priming suggests that there may be a complex response of soil microorganisms to the addition of labile organic compounds and decomposition of fresh plant residues. The addition of very small amounts of simple, easily decomposable substances (e.g., glucose and amino acids) and more complex compounds (e.g., soil and compost extracts and root exudates) has been shown to result in more C evolved as CO_2 than was present in the original substrate (De Nobili et al., 2001; Mondini et al., 2006). This finding suggests that trace amounts of easily degradable substances of low molecular weight act as trigger molecules or molecular signals, detected by microorganisms as forerunners to a larger pulse of readily available substrate.

26.3.4 Soil Fauna and Microorganisms

The decay of organic substrates in soil proceeds in incremental fashion, with each successive stage mediated by different biota. Soil fauna are important in early phases; although they account for only a small proportion (usually less than 10%) of soil respiration (Juma and McGill, 1986; Anderson, 1991), they physically fragment organic material, thereby increasing the surface area of the substrate and creating favorable habitats for microorganisms. They also mix and move the organic residues into close contact with the soil, increasing their exposure to microbes and their enzymes and also inoculating substrates with microbes. Soil fauna may further stimulate or inhibit microbial growth and activity through such actions as the deposition of mucus or wastes.

Although soil fauna are important agents of litter comminution and redistribution, the primary agents of mineralization are the microbes (Juma and McGill, 1986). Even in the Antarctic Dry Valleys, one of the harshest environments on earth, decomposition and oxidation to CO_2 of simple and complex substrates are microbially mediated (Hopkins et al., 2006).

26.3.5 Classification of Soil Microorganisms Involved in Decomposition

Different theories have been posited to classify microorganisms involved in the decomposition of organic matter on the basis of their nutritional versatility, affinity for substrates, and/or growth characteristics. Winogradsky (1924) proposed a functional distinction of microorganisms based on their activity and response to substrates in soil. *Authochthonous* organisms, which metabolize recalcitrant organic matter, predominate when there is little oxidizable substrate, growing and reproducing only slowly in soil. *Zymogenous* organisms, which metabolize labile organic matter, respond rapidly to fresh organic matter but become dormant when fresh substrate is depleted. Microbes exist in varying degrees of autochthony and zymogeny, comprising a continuum that ensures the most competitive bacteria will predominate with changes in substrate concentration and condition (Killham and Prosser, 2007).

Another way to classify microorganisms is based on ecological theory relating the density of a species to its soil food supply (Odum, 1953). According to this approach, slow-growing *K*-selected species are adapted to a more continuous supply of energy, where competition is intense. In contrast, fast-growing *r*-selected organisms live under conditions of normally sparse food supply and minimal competition but exhibit rapid growth per unit of food to exploit sporadic flushes of substrate. Yet another classification scheme, adapted from aquatic microbiology (Langer et al., 2004), categorizes organisms as *oligotrophic* (adapted to growth under conditions of C/nutrient starvation) and *copiotrophic* (adapted to C/nutrient stress).

Some researchers question the usefulness of classifying microorganisms based on physiology and function (Killham and Prosser, 2007; Kemmitt et al., 2008), suggesting that such divisions may occur within, as well as among, species; that is,

soil microorganisms are facultatively authochthonous or zymogenous, depending on the prevailing conditions and substrate availability. This view proposes that soil microorganisms are flexible, able to metabolize the wide range of substrates entering soil (Killham and Prosser, 2007).

26.3.6 Spatial Variation of Microbial Communities

Much of the soil microbial community may be dormant because of limited mobility or restricted access to food. But it is a diverse community, capable of surviving extreme environmental and food-related stresses. A notable difference between bacteria and fungi is their mode of growth (Coleman and Crossley, 1997). Bacteria depend on thin layers of water in soil for activity and motility. They are usually clustered in colonies, and their movement is mostly episodic, driven by factors such as rainfall, root growth, tillage, and ingestion by fauna. In contrast, fungal hyphae can grow over relatively large distances and penetrate into small spaces where they can decompose organic matter by secreting enzymes and translocating the nutrients back through the hyphae.

26.3.7 Enzymes

Enzymes catalyze many of the reactions necessary for microbial life, as well as for decomposition of litter. In soil, the activity of enzymes (and hence the rate and extent of decomposition) is regulated by many biotic and abiotic components, including living and latent cells, cell debris, clay minerals, humic colloids, and the soil aqueous phase (Burns, 1982). The location of extracellular enzymes in the soil matrix is determined by the size and solubility of its substrate, the species of microorganism, and the physical/chemical nature of organic and inorganic colloids. The binding of enzymes to soil colloids can extend their activity (some mechanism of association with the humic colloids offers protection but yet permits activity) but adsorption to soil surfaces does not always guarantee subsequent activity (Burns, 1982).

Because of the structural heterogeneity of plant tissues, extracellular enzymes are needed to reduce them to constituent monomers, which microorganisms can then consume. The enzymes most commonly linked to decomposition are hydrolases, which catalyze hydrolysis, and transferases, which catalyze the transfer of chemical groups from one molecule to another (Killham and Prosser, 2007). Other enzymes already widely studied include those prominent in C cycling (e.g., amylase, cellulase, lipase, glucosidases, and invertase), those that catalyze N processes (e.g., proteases, amidases, ureases, and deaminases), and those that promote the release of other nutrients (e.g., phosphatase and arylsulfatase).

A global-scale assessment of enzyme activities in 40 different ecosystems showed that the enzymatic potential for hydrolyzing the labile components of SOM is related to substrate availability, soil pH, and the stoichiometry of microbial nutrient demand (Sinsabaugh et al., 2008). The enzymatic potential for oxidizing the recalcitrant fractions of soil organic material, in contrast, is most strongly related to soil pH.

Most of the current methods for enzyme assays are measurements of potential rather than of in situ activity, and, therefore, results obtained from these methods should be carefully considered (Nannipieri et al., 2002). A full discussion of soil enzyme activities, and the location and properties of enzymes in soil is presented by Landi et al. (2009) and Gianfreda et al. (2009).

26.3.8 Factors Affecting Decomposition

26.3.8.1 Abiotic Mechanisms

Although litter decay is often considered a biological process, abiotic mechanisms can also be important; in some instances—notably arid and semiarid environments—abiotic mechanisms predominate, accounting for as much as two-thirds to three-quarters of the total annual loss from litter (Moorhead and Reynolds, 1989).

Photochemical breakdown of plant litter by solar radiation, particularly the ultraviolet (UV) spectrum, may be a particularly important mechanism of decay in some environments (Pancotto et al., 2005; Gallo et al., 2006, 2009). These losses represent a short circuit in the C cycle, diverting photosynthetically fixed C directly to the atmosphere without cycling through soil organic matter pools (Austin and Vivanco, 2006). Aside from direct photochemical breakdown, solar radiation can also affect decay indirectly, through effects on microbial communities or activity. For example, solar radiation may favor populations that tolerate higher temperatures, lower moisture, and higher levels of UV radiation (Caldwell et al., 2007); affect extracellular enzyme activity or plant litter chemical composition and characteristics; or reduce the stability of extracellular enzymes (Sinsabaugh et al., 2002).

Other abiotic processes of litter decay include leaching, which transfers water-soluble compounds away from decomposing organic matter, and physical abrasion, which breaks substrate into smaller pieces, creating more surfaces for microbial colonization. Because of numerous interactions, mechanisms of decomposition cannot be neatly separated into biotic and abiotic processes; the various processes are interdependent, each influencing the other.

26.3.8.2 Residue Quality

Fresh plant litter consists of mixtures of polysaccharides, polyphenols, celluloses, and lignins of variable composition and abundance and various architectural arrangements (Horwath, 2007). Its C concentration is relatively constant (usually between 40% and 50%) but the concentration of other constituents varies widely; for example, N concentration varies from <0.5% in wood to >10% in protein and bacterial cells.

Various indices have been proposed to describe the effect of residue chemical composition on decomposition. The most common of these, the C:N ratio, rests on the premise that N is the nutrient that most limits decomposition. In residues with wide C:N ratios, decomposition may be slowed because N immobilization exceeds mineralization, resulting in a shortage of available N for the decomposer organisms. The threshold C:N ratio,

above which decomposition is suppressed, is often about 20–30 (corresponding to an N concentration of about 15–25 g kg⁻¹).

The C:N ratio, however, is not an inerrant measure of N availability for decomposition (Jenkinson, 1981). Any factor that increases the rate of decomposition, and hence the N demand, tends to increase the threshold N concentration (narrow the threshold C:N ratio). For example, more favorable moisture or temperature regimes, higher rates of residue application, and a higher proportion of readily available C in the substrate all stimulate greater microbial activity, increase N demand, and increase the threshold N concentration. Laboratory studies, generally performed under favorable conditions, may therefore provide biased estimates of threshold C:N ratios or overstate the impact of residue N content field decomposition rates (Dendooven et al., 1990).

Decomposition depends not only on N content, but also on the lignin, polyphenol, and carbohydrate content of residues (Horwath, 2007). Consequently, various other ratios have been proposed as predictors of the decomposition rate based on observed interactions between N and organic constituents such as lignin, carbohydrates, or polyphenols. Berg (1986) proposed that different factors predominate at progressive stages: early on, the decomposition rate is determined by the concentration of nutrients and readily degradable C, while later, the rate is determined by lignin content.

The composition of the microbial community has an effect on decomposition because the biochemical composition of microbial cells varies among microbial groups (e.g., fungi, G⁺, or G⁻ bacteria). As well, species vary in their output of compounds such as phenols and amino sugars, which, even when produced from labile compounds such as glucose, may be more resistant to decomposition than some plant-derived compounds (Voroney et al., 1989).

Another view of decomposition proposes that it is regulated jointly by the chemical composition of the organic matter and the activity of the microbial community (Moorhead and Sinsabaugh, 2006). A central hypothesis of this view is that the relative amounts of lignin and holocellulose (i.e., cellulose + hemicelluloses) define a threshold for the controls on the process of decomposition. The plant residue chemistry is linked to the activity of the heterotrophic microbial community that comprises three primary guilds: a guild of opportunist microorganisms that consumes labile proteins, grows quickly, and has high affinity for soluble substrates; a guild of decomposer specialists that needs external N inputs, grows more slowly, and has high affinity for holocellulose substrates; and a guild of miners that uses oxidative enzymes to breakdown humus, grows very slowly, and is specialized for degrading lignin.

26.3.8.3 Physical–Chemical Environment

26.3.8.3.1 Water

Water potential affects decomposition, because soil microorganisms depend on water for survival and mobility. As well, water potential can influence decomposition indirectly through effects on O_2 supply, amount of soluble substrate, and pH of the soil solution.

Optimum soil water potentials for decomposition of residue are between −0.01 and −0.05 MPa (Parr and Papendick, 1978). With decreasing water potential, decomposition slows because of reduced microbial activity. Bacterial activity becomes limited first, because of reduced cell mobility and limited availability of substrate (Sommers et al., 1981). Fungi tend to tolerate greater water stress and are capable of surviving at very low water potentials (−4 to −10 MPa). At water potentials higher than about −0.01 MPa, decomposition is slowed by low oxygen tension, because oxygen diffuses more slowly in water than in air.

26.3.8.3.2 Oxygen Supply

The complete decomposition of organic substrates to CO_2 and other inorganic constituents depends on an adequate supply of oxygen. Hence, rates of aerobic decomposition decline when oxygen concentration falls below a threshold level, about 10% according to Schachtschabel et al. (1984; cited by Verburg et al., 1995). Whether oxygen depletion affects decomposition, however, depends on the rate of oxygen replenishment relative to that of oxygen consumption at the site of decomposition. Thus, any factor that reduces rates of oxygen diffusion, such as high moisture content, poor soil structure, or increasing depth within soil, tends to limit oxygen availability. Similarly, oxygen deficiency may be induced by high rates of decomposition (i.e., oxygen consumption) at localized sites. Therefore, even well-drained soils may show appreciable anaerobic activity, presumably in local pockets of low oxygen concentration within soil aggregates (Tiedje et al., 1984).

Although fungi, algae, and many bacteria are usually aerobic, the population in a particular soil may include the full range from obligate aerobes to obligate anaerobes. Some fungi may have the ability to survive anaerobic conditions, though anaerobic activity in soil is dominated by bacteria.

26.3.8.3.3 Temperature

The rate of decomposition is highly responsive to temperature, increasing severalfold over the range found in unfrozen soils (0°C–35°C). Laboratory incubations, for example, suggest that the decomposition rate increases about twofold for a 10° increase (i.e., Q_{10} = 2) between 10°C and 40°C (Stanford et al., 1973), though Q_{10} values may range from 1.6 to 3.2 (Anderson, 1991). Near freezing, microbial activity is constrained more by the absence of liquid water than by low temperature per se. Some activity can occur below 0°C, probably in unfrozen water films around soil colloids.

At a regional level, the effects of temperature on decomposition are reflected in the accumulation of organic matter in soil. Transects of soils in the semiarid, semi-humid, and humid regions in the central United States and Canada show that soils in warm climates contain less organic matter than those in cool regions, in part because of faster decomposition (Jenny, 1941).

Evaluating the effect of temperature on a microbial function like decomposition is difficult, because warming usually results in drier soil conditions, and soil temperature and soil water content have opposite effects on microbial activity. In a study involving

21 cropped sites from different climatic regimes across North America, Insam (1990) observed that the proportion of soil C in the microbial biomass (i.e., C_{mic}-to-C_{org} ratio) was relatively high in arid soils but lowest when the precipitation/evaporation ratio (P/E) was equal to 1. Beyond this point, the proportion increased with increasing P/E. This relationship between P/E and C_{mic}-to-C_{org} ratio suggests that a relatively low microbial biomass is maintained with a given C input and soil water conditions favorable for decomposition. When decomposition is retarded during long drought periods, organic matter is metabolized only during brief periods and a higher C_{mic}-to-C_{org} ratio is maintained because the microbial biomass is active for only a short time. High soil water content may have a similar effect on the C_{mic}-to-C_{org} ratio, because saturated or near-saturated conditions also hinder decomposition, particularly at high temperatures.

26.3.8.3.4 pH

Soil pH affects decomposition by controlling the activity of soil microorganisms. Highly acidic or alkaline soil conditions tend to depress the size of the bacterial community, which prefers near-neutral (pH 6–7.5) soil conditions (Alexander, 1977). Fungi are less sensitive to low pH and usually predominate in acidic soils.

Soil pH also governs the activity of extracellular enzymes immobilized in the soil matrix, each of which has an optimal pH. For example, glycosidases have pH optima between 4 and 6; phenol oxidases, lignin peroxidases, and most proteases have optima between 7 and 9 (Sinsabaugh et al., 2008).

Jenkinson (1977) found that organic material decomposed more slowly in strongly acid soils than in neutral soils. After 1 year, soils with pH 3.7 had more residue than soils with pH 6.9. However, in the long term (>5 years), this difference disappeared, suggesting that the effects of soil acidity on decomposition were predominant in the early stages of decomposition.

26.3.8.3.5 Accessibility

26.3.8.3.5.1 Degree of Physical Protection Organic substrates can be physically protected from microbial and enzyme attack in several ways (Jastrow et al., 2007; Krull et al., 2007). First, clay minerals can adsorb large organic molecules directly, reducing their availability to decomposition. Enzymes that attack the organic molecules may be adsorbed to clay surfaces and become inactivated. Secondly, organic material may be trapped within aggregates (especially microaggregates), which shield them from attack. Most of the habitable space in soil may consist of pores so small that they prevent the entry of most bacteria (Juma, 1993).

Evidence for the physical protection of organic matter within aggregates is the flush of CO_2 that is sometimes released when aggregates are crushed (Elliott, 1986). Chemical analysis of organic material located within aggregates indicates that some of it is relatively undecomposed (Golchin et al., 1994). The capacity of a soil to physically protect organic matter may be limited; Hassink (1996) postulated that this protective capacity is saturated in soils that have not been cultivated or that have been under grass for a long time. According to this hypothesis, the

potential for preserving soil organic matter against microbial activity is directly related to the proportion of the soil's potential protective capacity not yet occupied (Stewart et al., 2008).

The effect of physical protection may be mitigated by other factors. For example, Skene et al. (1997) suggested that physical protection may be the primary deterrent to decomposition of high quality substrates, but that chemical protection (i.e., chemical composition) is the predominant limiting factor in low-quality substrates.

Iron and aluminum oxides, hydroxides, and oxyhydroxides (i.e., FeOx and AlOx) influence the decomposition of soil organic matter because they have extensive and reactive surfaces and therefore a high sorption capacity (Tipping, 1981; Wagi and Mayer, 2007). This high sorptive capacity helps explain observations that metal oxides are often better predictors of soil organic C concentration and have a greater influence on organic C concentration than clay minerals (Eusterhues et al., 2005; Wiseman and Puttmann, 2005). Another important feature of metal oxides related to decomposition is their inhibiting effect on microbial mineralization of organic compounds (Jones and Edwards, 1998).

26.3.8.3.5.2 Particle Size Finely divided organic matter usually decomposes more quickly than coarse organic matter because of increased surface area and greater exposure to microbial activity. In one study, for example, decomposition of straw in an 8–10 month period was 26%–47% faster in mesh bags allowing entry of earthworms than in bags that excluded earthworms (Curry and Byrne, 1992). The effect of particle size, however, varies with time: in early stages, decay may be faster for fine particles than for coarse particles early on, but the trend may be reversed later (Angers and Recous, 1997).

26.3.8.3.5.3 Location Plant litter at the soil surface often decomposes more slowly than that within the matrix of the soil, though the effect of placement diminishes with time (Cogle et al., 1987; Coppens et al., 2006). The initial lag in decomposition of residues on the soil surface is probably attributable to reduced contact with decomposer organisms and to less favorable temperature and moisture conditions, and its magnitude may therefore vary with climate. In cool soils that are poorly drained, decomposition may also be slowed when residues are placed deeper within the profile, for example, by moldboard plowing (Angers et al., 1997).

26.3.8.4 Human Controls on Decomposition

Human modification or management of agricultural, range, and forest ecosystems affects decomposition by altering some of the factors described above. One of the most widespread and pronounced effects on decomposition was the original clearing or cultivation of land for intensive agriculture (Watson, 2000). This land use change often resulted in accelerated decomposition because of changes in water content, aeration, temperature, and accessibility of organic matter. Together with reduced C inputs (because of export in harvests), this enhanced decomposition usually caused depletion in soil C stocks.

Even after the effects of original cultivation have waned, rates of decomposition are influenced by the way lands are managed. Management practices such as tillage, selection of crops and cropping sequences, grazing management, and fertilization can alter decomposition rates through their effects on soil structure, water holding capacity, aeration, temperature, and composition and placement of residues (e.g., Smith et al., 2008a; Johnston et al., 2009). Restoring a semblance of the pre-cultivation tillage, such as by reestablishing permanent grasslands, often induces significant soil C accrual (Conant et al., 2001), presumably, in part because of altered decomposition rates and larger C inputs.

Recently, managing lands to enhance soil C reserves has taken on renewed importance, with growing worries about climate change and the recognition that biological C sinks can, at least in the short term, help slow the accrual of atmospheric CO_2. In general, soil C is lost when grasslands, forest, or other native ecosystems are converted to croplands, or by draining, cultivating, or liming highly organic soils. Conversely, soil C increases when former croplands are restored to grasslands, forests, or native vegetation, or by restoring organic soils to their native condition. In managed lands, practices that increase C inputs to the soil (e.g., increased and/or improved residue and manure management) or reduce losses (e.g., reduced tillage, reduced residue removal) help maintain or even increase soil C levels. Since soil disturbance usually stimulates soil C losses through enhanced decomposition and erosion, reduced- or no-till practices are often suggested as a means of increasing the terrestrial C sink (Smith et al., 2008b). How residue is managed (e.g., whether burned or buried with tillage implements) is also crucial for soil C storage.

26.3.8.5 Methods for Measuring Decomposition

Since decomposition comprises various biological, physical, and chemical processes, a broad range of research techniques and tools has been used to study it (Berg and Laskowski, 2006). Generally, research techniques are divided between field (or in situ) and laboratory methods. Depending on the research objectives, the various techniques can be used equally well, with only minor modifications, in both the field and laboratory, although the interpretation of results may differ between them.

26.3.8.5.1 Field Methods

The most widely used method of measuring residue decomposition in the field is the litter bag technique. In this approach, litter is placed in nondegradable mesh bags and incubated in the field, and mass loss is determined by periodic sampling. This technique has a number of limitations (Heal et al., 1997; Beyaert and Fox, 2008). It measures only net changes and ignores the fate of material leaving the bag; this material may be more labile and contain more nutrients than the residue remaining and so play a key role in other ecosystem processes affecting water and atmospheric quality (Figure 26.5). As well, the exclusion of fauna and vegetation, as well as different climatic conditions in the bag from those of the surrounding soil, may affect decomposition.

Another way to evaluate decomposition in the field is to add labeled (e.g., ^{14}C, ^{13}C, or ^{15}N) plant residues to soil. The residues are usually placed in small cylinders of soil and covered with a mesh screen to prevent contamination by unlabeled material. At periodic intervals, soil is removed from the cylinders and analyzed to measure the quantity of the label remaining. This approach permits direct measurement of the fate of plant litter, and analysis by mass spectrometer or nuclear magnetic resonance (^{13}C NMR) allows for quantitative (i.e., kinetics) and qualitative (i.e., biochemical) changes of decomposition to be accurately measured.

The variation in natural abundance of stable C isotope ratios ($^{13}C/^{12}C$) in plants with different photosynthetic pathways can be used to estimate the proportion of soil organic matter derived from distinct C inputs (Ellert and Rock, 2008). The stable C isotope ratio ($\delta^{13}C$) of soil organic matter is similar to that of the vegetation growing on it. Plants with a C_3 photosynthetic pathway have low $\delta^{13}C$ ratios compared to those with C_4 pathway, and, thus, at sites where the vegetation has changed from one photosynthetic pathway type to the other, the $\delta^{13}C$ ratio of soil organic matter changes in a measurable way. On average, the $\delta^{13}C$ of C_4 plants is 14‰ higher than that of C_3 plants. This large isotopic difference is at least 5–15 times greater than the changes that can be induced by C isotope discrimination during decomposition (Boutton, 1996). This approach is limited to sites with vegetation changes of known date but is powerful because it characterizes C produced in situ and can be used to evaluate soil C dynamics over short (1 year) or long (>100 years) time periods (Balesdent and Mariotti, 1996).

Stable isotope probing (SIP) is a powerful tool that links the taxonomic identity of microorganisms to their metabolic function (Dumont and Murrell, 2005). In soil decomposition studies, SIP involves adding a substrate enriched in a stable isotope (e.g., ^{13}C or ^{15}N) to soil. After a period of incubation during which microorganisms assimilate the stable isotope label, cellular components are extracted from the soil and analyzed for the incorporated label. This incorporation of stable isotopes from labeled compounds into biological signature molecules (i.e., biomarkers) allows for selective detection (e.g., PLFAs by isotope ratio mass spectrometry) or isolation (e.g., DNA and RNA by density centrifugation) of the labeled macromolecules from which the identity of microorganisms can be inferred (Manefield et al., 2002).

26.3.8.5.2 Laboratory Methods

Laboratory incubations permit the detailed study of residue decomposition products and generate information about the decomposition processes. They are usually conducted under favorable temperature and moisture regimes and often exclude processes like faunal activity and nutrient leaching, which influence decomposition in the field. Although the simplified and optimal conditions under which they are conducted prevent direct extrapolation of rates to the field, laboratory incubations are nevertheless useful for characterizing the influence of individual factors on decomposition.

Laboratory incubations with ^{13}C-labeled plant material are valuable in evaluating microbial community level changes during decomposition through analysis of ^{13}C PLFAs (Williams et al., 2006; Moore-Kucera and Dick, 2008). Such isotopic techniques accurately quantify the flow of C into the soil microbial community and identify specific groups of microorganisms responsible for decomposition.

A serious research challenge in evaluating the decomposition of soil organic matter is the development of laboratory methods that extract from soil meaningful fractions of soil organic matter, that is, those that are functionally relevant (Olk and Gregorich, 2006). A useful extraction procedure would isolate meaningful fractions that comprise significant proportions of total soil organic matter and respond in quantity and chemical nature to management or an ecosystem process. Such fractions could be used to elucidate the effects of a management practice or ecosystem process on soil organic matter pools that are cycling at relevant rates.

Comprehensive reviews are available for extractions involving particulate organic matter and the light fraction (Wander, 2004; Gregorich et al., 2006), particle size fractions (Christensen, 1992), aggregate classes (Six et al., 2004), and chemical extractants (Stevenson, 1994).

26.3.8.5.3 *Modeling Methods*

Given the importance of decomposition in understanding fluxes of soil C and N, many soil biogeochemical models have sought to describe the process in a variety of mathematical formulations. There are a number of reviews and comparisons of these models in the literature (McGill, 1996; Paustian et al., 1997; Molina and Smith, 1998; Smith et al., 1998; Wutzler and Reichstein, 2008; Manzoni and Porporato, 2009). The formulations can be an equation or suite of equations that mathematically describe how biotic and abiotic factors regulate the flow of C and N toward more stable forms or mineral pools. When decomposition is viewed as an enzyme-catalyzed reaction a Michaelis–Menten equation is used to characterize the C flow from decomposition in terms of both substrate and enzyme concentration. Some models express the rate of decomposition as constant or linear with respect to the substrate or the microbial biomass, that is, simple first-order kinetics. Multiplicative equations are used to describe the decomposition process when the substrate and microbial biomass are considered to be linked. Some biogeochemical models use complex nonlinear formulations with feedbacks based on microbial function. To date, more than half the models developed are based on first-order decomposition kinetics (Manzoni and Porporato, 2009).

26.3.8.6 Decomposition in a Changing World

Human activity can influence decomposition indirectly by altering global conditions. Increased temperature, by itself, increases the rate of decomposition (Kirschbaum, 1995), and incorporating this observation into simulation models has led to the suggestion that increases in global temperature resulting from an enhanced "greenhouse effect" could accelerate decomposition, resulting in depletion of soil organic matter reserves (Jenkinson

et al., 1991; Kirschbaum, 2000; Jones et al., 2005). Conversely, the greenhouse effect could induce increases in plant-derived C inputs that would exceed the increases in decomposition, resulting in a buildup of soil organic matter levels.

These types of feedbacks to climate change have not yet been clearly elucidated yet (Davidson and Janssens, 2006) because rising temperatures are just one part of a complex suite of changes that would affect the C cycle under global climate change. Precipitation (i.e., water availability) and atmospheric CO_2 concentration (i.e., CO_2 fertilization) would also change. While such changes affect belowground heterotrophic processes, they also influence aboveground photosynthetic processes (i.e., net primary productivity); both sets of processes—dissimilatory and assimilatory—need to be studied to understand the net effect (Heimann and Reichstein, 2008). Decomposition and primary production respond not to average climate but to actual weather conditions, so predictions based on gradual changes in average climate variables and atmospheric gas concentrations may be misleading if climate variability and extremes are ignored.

The effect of temperature becomes even more complicated in the context of the decomposition of soil organic matter that contains myriad organic C compounds, each having different structure, molecular weight, and solubility, and thus different kinetic properties and susceptibility to enzymatic degradation. The more complex soil organic matter compounds are characterized by low decomposition rates, high activation energies, and inherently high temperature sensitivity (Davidson and Janssens, 2006). Since soil organic matter comprises mostly more complex and recalcitrant compounds, the behavior of this pool is crucial to estimating the direction and magnitude of a response to an increase in temperature.

Decomposition of plant residues and soil organic matter has been studied now for many decades, yielding broader understanding of this complex process. But coming global changes will impose new stresses on the global C cycle, and the process of decomposition is at the heart of the C cycle's response to these stresses. Consequently, the study of this critical process in ecosystems worldwide will likely become even more urgent in coming decades.

References

Alexander, M. 1977. Introduction to soil microbiology. John Wiley & Sons, New York.

Anderson, J.M. 1991. The effects of climate change on decomposition processes in grassland and coniferous forests. Ecol. Appl. 1:326–347.

Angers, D.A., M.A. Bolinder, M.R. Carter, E.G. Gregorich, C.F. Drury, B.C. Liang, R.P. Voroney, R.R. Simard, R.G. Donald, R.P. Beyaert, and J. Martel. 1997. Impact of tillage practices on organic carbon and nitrogen storage in cool, humid soils of eastern Canada. Soil Tillage Res. 41:191–201.

Angers, D.A., and S. Recous. 1997. Decomposition of wheat straw and rye residues as affected by particle size. Plant Soil 189:197–203.

Austin, A.T., and L. Vivanco. 2006. Plant litter decomposition in a semi-arid ecosystem controlled by photodegradation. Nature 442:552–558.

Balesdent, J., and A. Mariotti. 1996. Measurement of soil organic matter turnover using ^{13}C natural abundance, p. 83–111. *In* T.W. Boutton and S. Yamasaki (eds.) Mass spectrometry of soils. Marcel Dekker, Inc., New York.

Berg, B. 1986. Nutrient release from litter and humus in coniferous forest soils—A mini review. Scand. J. For. Res. 1:359–369.

Berg, B., and R. Laskowski. 2006. Methods in studies of organic matter decay. Adv. Ecol. Res. 38:291–331.

Beyaert, R.P., and C.A. Fox. 2008. Assessment of soil biological activity, p. 527–545. *In* M.R. Carter and E.G. Gregorich (eds.) Soil sampling and methods of analysis. CRC Press, Boca Raton, FL.

Blagodatskaya, E., and Y. Kuzyakov. 2008. Mechanisms of real and apparent priming effects and their dependence on soil microbial biomass and community structure: Critical review. Biol. Fertil. Soils 45:115–132.

Boutton, T.W. 1996. Stable carbon isotope ratios of soil organic matter and their use as indicators of vegetation and climate change, p. 47–82. *In* T.W. Boutton and S. Yamasaki (eds.) Mass spectrometry of soils. Marcel Dekker, New York.

Burns, R.G. 1982. Enzyme activity in soil: Location and a possible role in microbial ecology. Soil Biol. Biochem. 14:423–427.

Caldwell, M.M., J.F. Bornman, C.L. Ballare, S.D. Flint, and G. Kulandaivelu. 2007. Terrestrial ecosystems, increased solar ultraviolet radiation, and interactions with other climate change factors. Photochem. Photobiol. Sci. 6:252–266.

Cheng, W., and D.C. Coleman. 1990. Effect of living roots on soil organic matter decomposition. Soil Biol. Biochem. 22:781–787.

Christensen, B.T. 1992. Physical fractionation of soil and organic matter in primary particle size and density separates. Adv. Soil Sci. 20:1–90.

Cogle, A.L., W.M. Strong, P.G. Saffigna, J.N. Ladd, and M. Amato. 1987. Wheat straw decomposition in subtropical Australia. II. Effect of straw placement on decomposition and recovery of added ^{15}N urea. Aust. J. Soil Res. 25:481–490.

Coleman, D.C., and D.A. Crossley, Jr. 1997. Fundamentals of soil ecology. Academic Press Inc., San Diego, CA.

Conant, R.T., K. Paustian, and E.T. Elliott. 2001. Grassland management and conversion into grassland: Effects on soil carbon. Ecol. Appl. 11:343–355.

Coppens, F., R. Merckx, and S. Recous. 2006. Impact of crop residue location on carbon and nitrogen distribution in soil and in water-stable aggregates. Eur. J. Soil Sci. 57:570–582.

Curry, J.P., and D. Byrne. 1992. The role of earthworms in straw decomposition and nitrogen turnover in arable land in Ireland. Soil Biol. Biochem. 24:1409–1412.

Davidson, E.A., and I.A. Janssens. 2006. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. Nature 440:165–173.

De Nobili, M., M. Contin, C. Mondini, and P.C. Brookes. 2001. Soil microbial biomass is triggered into activity by trace amounts of substrate. Soil Biol. Biochem. 33:1163–1170.

Dendooven, L., L. Verstraeten, and K. Vlassak. 1990. The N-mineralization potential: An undefinable parameter, p. 170–181. *In* R. Merckx, H. Vereecken, and K. Vlassak (eds.) Fertilization and the environment. Leuven University Press, Leuven, Belgium.

Dumont, M.G., and J.C. Murrell. 2005. Stable isotope probing—Linking microbial identity to function. Nat. Rev. 3:499–504.

Ellert, B.H., and L. Rock. 2008. Stable isotopes in soil and environmental research, p. 693–711. *In* M.R. Carter and E.G. Gregorich (eds.) Soil sampling and methods of analysis. CRC Press, Boca Raton, FL.

Elliott, E.T. 1986. Aggregate structure and carbon, nitrogen, and phosphorus in native and cultivated soils. Soil Sci. Soc. Am. J. 50:627–633.

Eusterhues, K., C. Rumpel, M. Kleber, and I. Kögel-Knabner. 2005. Organo-mineral associations in sandy acidic forest soils: Importance of specific surface area, iron oxides and micropores. Eur. J. Soil Sci. 56:753–763.

Gallo, M.E., A. Porras-Alfaro, K.J. Odenbach, and R.L. Sinsabaugh. 2009. Photoacceleration of plant litter decomposition in an arid environment. Soil Biol. Biochem. 41:1433–1441.

Gallo, M.E., R.L. Sinsabaugh, and S.E. Cabaniss. 2006. The role of ultraviolet radiation in litter decomposition in arid ecosystems. Appl. Soil Ecol. 34:82–91.

Golchin, A., J.M. Oades, J.O. Skjemstad, and P. Clarke. 1994. Study of free and occluded particulate organic matter in soils by solid state ^{13}C CP/MAS NMR spectroscopy and scanning electron microscopy. Aust. J. Soil Res. 32:285–309.

Gregorich, E.G., M.H. Beare, U.F. McKim, and J.O. Skjemstad. 2006. Chemical and biological characteristics of physically uncomplexed organic matter. Soil Sci. Soc. Am. J. 70:975–985.

Hassink, J. 1996. Preservation of plant residues in soils differing in unsaturated protective capacity. Soil Sci. Soc. Am. J. 60:487–491.

Heal, O.W, J.M. Anderson, and M.J. Swift. 1997. Plant litter quality and decomposition: An historical overview, p. 3–30. *In* G. Cadisch and K.E. Giller (eds.) Driven by nature: Plant litter quality and decomposition. CABI, Wallingford, U.K.

Heimann, M., and M. Reichstein. 2008. Terrestrial ecosystem carbon dynamics and climate feedbacks. Nature 451:289–292.

Hopkins, D.W., A.D. Sparrow, P.M. Novis, E.G. Gregorich, B. Elberling, and L.G. Greenfield. 2006. Controls on the distribution of productivity and organic resources in Antarctic dry valley soils. Proc. R. Soc. B 273:2687–2695.

Horwath, W. 2007. Carbon cycling and formation of soil organic matter, p. 303–339. *In* E.A. Paul (ed.) Soil microbiology, ecology, and biochemistry. Academic Press, New York.

Insam, H. 1990. Are soil microbial biomass and basal respiration governed by the climatic regime? Soil Biol. Biochem. 22:525–532.

Jastrow, J.D., J.E. Amonette, and V.L. Bailey. 2007. Mechanisms controlling soil carbon turnover and their potential application for enhancing carbon sequestration. Clim. Change 80:5–23.

Jenkinson, D.S. 1977. Studies on the decomposition of plant material in soil V. The effects of plant cover and soil type on the loss of carbon from ^{14}C labelled ryegrass decomposing under field conditions. J. Soil Sci. 28:424–434.

Jenkinson, D.S. 1981. The fate of plant and animal residues in soil, p. 505–561. *In* D.J. Greenland and M.H.B. Hayes (eds.) The chemistry of soil processes. John Wiley & Sons, New York.

Jenkinson, D.S., D.E. Adams, and A. Wild. 1991. Model estimates of CO_2 emissions from soil in response to global warming. Nature 351:304–306.

Jenny, H. 1941. Factors of soil formation. McGraw-Hill, New York.

Johnston, A.E., P.R. Poulton, and K. Coleman. 2009. Soil organic matter: Its importance in sustainable agriculture and carbon dioxide fluxes. Adv. Agron. 101:1–75.

Jones, D.L., and A.C. Edwards. 1998. Influence of sorption on the biological utilization of two simple carbon substrates. Soil Biol. Biochem. 30:1895–1902.

Jones, C., C. McConnell, K. Coleman, P. Cox, P. Fallon, D. Jenkinson, and D. Powlson. 2005. Global climate change and soil carbon stocks: Prediction from two contrasting models for the turnover of organic carbon in soil. Global Change Biol. 11:154–166.

Juma, N.G. 1993. Interrelationships between soil structure/texture, soil biota/soil organic matter and crop production. Geoderma 57:3–30.

Juma, N.J., and W.B. McGill. 1986. Decomposition and nutrient cycling in agro-ecosystems, p. 74–136. *In* M.J. Mitchel and J.P. Nakas (eds.) Microfloral and faunal interactions in natural and agro-ecosystems. Martinus Nijhoff/Dr. W. Junk Publishers, Rotterdam, the Netherlands.

Kemmitt, S.J., C.V. Lanyon, I.S. Waite, Q. Wen, T.M. Addiscott, N.R.A. Bird, A.G. O'Donnell, and P.C. Brookes. 2008. Mineralization of native soil organic matter is not regulated by the size, activity or composition of the soil microbial biomass—A new perspective. Soil Biol. Biochem. 40:61–73.

Killham, K., and J.I. Prosser. 2007. The prokaryotes, p. 119–162. *In* E.A. Paul (ed.) Soil microbiology, ecology, and biochemistry. Academic Press, Amsterdam, the Netherlands.

Kirschbaum, M.U.F. 1995. The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. Soil Biol. Biochem. 27:753–760.

Kirschbaum, M.U.F. 2000. Will changes in soil organic carbon act as a positive or negative feedback on global warming? Biogeochemistry 48:21–51.

Krull, E.S., J.A. Baldock, and J.O. Skjemstad. 2007. Importance of mechanisms and processes of the stabilisation of soil organic matter for modelling carbon turnover. Funct. Plant Biol. 30:207–222.

Kuzyakov, Y., J.K. Friedel, and K. Stahr. 2000. Review of mechanisms and quantification of priming effects. Soil Biol. Biochem. 32:1485–1498.

Langer, U., L. Böhme, and F. Böhme. 2004. Classification of soil microorganisms based on growth properties: A critical view of some commonly used terms. J. Plant Nutr. Soil Sci. 167:267–269.

Manefield, M., A.S. Whiteley, N. Ostle, P. Ineson, and M.J. Bailey. 2002. Technical considerations for RNA-based stable isotope probing: An approach to associating microbial diversity with microbial community function. Rapid Commun. Mass Spectrom. 16:2179–2183.

Manzoni, S., and A. Porporato. 2009. Soil carbon and nitrogen mineralization: Theory and models across scales. Soil Biol. Biochem. 41:1355–1379.

McGill, W.B. 1996. Review and classification of ten soil organic matter (SOM) models, p. 111–133. *In* D.S. Powlson, P. Smith, and J.U. Smith (eds.) Evaluation of soil organic matter models. NATO series I: Global environmental change. Vol. 38. Springer-Verlag, Berlin, Germany.

Molina, J.A.E., and P. Smith. 1998. Modeling carbon and nitrogen processes in soil. Adv. Agron. 62:253–298.

Mondini, C., M.L. Cayuela, M.A. Sanchez-Monedero, A. Roig, and P.C. Brookes. 2006. Soil microbial biomass activation by trace amounts of readily available substrate. Biol. Fertil. Soils 42:542–549.

Moore-Kucera, J., and R.P. Dick. 2008. Application of ^{13}C-labeled litter and root materials for in situ decomposition studies using phospholipids fatty acids. Soil Biol. Biochem. 40:2485–2493.

Moorhead, D.L., and J.F. Reynolds. 1989. Mechanisms of surface litter mass loss in the northern Chihuahan desert: A reinterpretation. J. Arid Environ. 16:157–163.

Moorhead, D.L., and R.L. Sinsabaugh. 2006. A theoretical model of litter decay and microbial interaction. Ecol. Monogr. 76:151–174.

Nannipieri, P., E. Kandeler, and P. Ruggiero. 2002. Enzyme activities and microbiological and biochemical processes in soil, p. 1–33. *In* R.G. Burns and R. Dick (eds.) Enzymes in the environment. Marcel Dekker, New York.

Odum, E.P. 1953. Fundamentals of ecology. W.B. Saunders Company, Philadelphia, PA.

Olk, D., and E.G. Gregorich. 2006. Overview of the symposium proceedings, "Meaningful pools in determining soil carbon and nitrogen dynamics." Soil Sci. Soc. Am. J. 70:967–974.

Pancotto, V.A., O.E. Sala, T.M. Robson, M.M. Caldwell, and A.L. Scopel. 2005. Direct and indirect effects of solar ultraviolet-B radiation on long-term decomposition. Global Change Biol. 11:1982–1989.

Parr, J.F., and R.I. Papendick. 1978. Factors affecting the decomposition of crop residues by microorganisms, p. 101–129. *In* W.R. Oschwald (ed.) Crop residue management systems. Special Publication No. 31. ASA, Madison, WI.

Paustian, K., G.I. Ågren, and E. Bosatta. 1997. Modelling litter quality effects on decomposition and soil organic matter dynamics. *In* G. Cadish and G.E. Giller (eds.) Driven by nature: Plant litter quality and decomposition. CABI, Wallingford, U.K.

Sinsabaugh, R.L., M.M. Carreiro, and D.A. Repert. 2002. Allocation of extracellular enzymatic activity in relation to litter composition, N deposition, and mass loss. Biogeochemistry 60:1–24.

Sinsabaugh, R.L., C.L. Lauber, M.N. Weintraub, B. Ahmed, S.D. Allison, C. Crenshaw, A.R. Contosta et al. 2008. Stoichiometry of soil enzyme activity at global scale. Ecol. Lett. 11:1252–1264.

Six, J., H. Bossuyt, S. Degryze, and K. Denef. 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil Tillage Res. 79:7–31.

Skene, T.M., J.O. Skjemstad, J.M. Oades, and P.J. Clarke. 1997. The influence of inorganic matrices on the decomposition of Eucalyptus litter. Aust. J. Soil Res. 35:73–87.

Smith, P., O. Andren, L. Brussaard, D.M. Dangerfield, K. Ekschmitt, P. Lavelle, and K. Tate. 1998. Soil biota and global change at the ecosystem level: Describing soil biota in mathematical models. Global Change Biol. 4:773–784.

Smith, P., D. Martino, Z. Cai, D. Gwary, H.H. Janzen, P. Kumar, B. McCarl et al.. 2008a. Greenhouse gas mitigation in agriculture. Philos. Trans. R. Soc. B 363:789–813.

Smith, P., D. Martino, Z. Cai, D. Gwary, H. Janzen, P. Kumar, B. McCarl et al. 2008b. Agriculture. *In* B. Metz et al. (eds.) Climate change 2007: Mitigation. Contribution of working group III to the 4th assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, U.K.

Sommers, L.E., C.M. Gilmour, R.E. Wildung, and S.M. Beck. 1981. The effect of water potential on decomposition processes in soils, p. 97–117. *In* J.F. Parr, W.R. Gardner, and L.F. Elliott (eds.) Water potential relations in soil microbiology. SSSA Special Publication No. 9. SSSA, Madison, WI.

Stanford, G., M.H. Frere, and D.H. Schwaninger. 1973. Temperature coefficient of soil nitrogen mineralization. Soil Sci. 115:321–323.

Stevenson, F.J. 1994. Humus chemistry: Genesis, composition, reactions. John Wiley & Sons, New York.

Stewart, C.E., A.F. Plante, K. Paustian, R.T. Conant, and J. Six. 2008. Soil carbon saturation: Linking concept and measurable carbon pools. Soil Sci. Soc. Am. J. 72:379–392.

Swift, M.J., O.W. Heal, and J.M. Anderson. 1979. Decomposition in terrestrial ecosystems. University of California Press, Berkeley, CA.

Tiedje, J.M., A.J. Sexstone, T.B. Parkin, N.P. Revsbech, and D.R. Shelton. 1984. Anaerobic processes in soil. Plant Soil 76:197–212.

Tipping, E. 1981. The adsorption of aquatic humic substances by iron oxides. Geochim. Cosmochim. Acta 45:191–199.

Verburg, P.S.J., D. van Dam, J.C.Y. Marinissen, R. Westerhof, and N. van Breemen. 1995. The role of decomposition in C sequestration in ecosystems, p. 85–112. *In* M.A. Beran (ed.) Carbon sequestration in the biosphere, NATO ASI Series. Series I, Global environmental change. Vol. 33. Springer-Verlag, Berlin, Germany.

Vernadsky, V.I. 1998. The biosphere. Copernicus, Springer-Verlag, New York Inc., NY.

Voroney, R.P., E.A. Paul, and D.W. Anderson. 1989. Decomposition of wheat straw and stabilization of microbial products. Can. J. Soil Sci. 69:63–77.

Wagi, R., and L.M. Mayer. 2007. Sorptive stabilization of organic matter in soils by hydrous iron oxides. Geochim. Cosmochim. Acta 71:25–35.

Wander, M. 2004. Soil organic matter fractions and their relevance to soil function, p. 67–102. *In* F. Magdoff and R.R. Weil (eds.) Soil organic matter in sustainable agriculture. CRC Press, Boca Raton, FL.

Watson et al. (eds.) 2000. Intergovernmental panel on climate change (IPCC). Land use, land-use change, and forestry. Cambridge University Press, Cambridge, U.K.

Williams, M.A., D.D. Myrold, and P.J. Bottomley. 2006. Carbon flow from ^{13}C-labeled straw and root residues into the phospholipid fatty acids of a soil microbial community under field conditions. Soil Biol. Biochem. 38:759–768.

Winogradsky, S. 1924. Sur la microflora autochtone de la terre arable. C.R. Hebd. Seances Acad. Sci. (Paris) D 178:1236–1239.

Wiseman, C.L.S., and W. Püttmann. 2005. Soil organic carbon and its sorptive preservation in central Germany. Eur. J. Soil Sci. 56:65–76.

Wutzler, T., and M. Reichstein. 2008. Colimitation of decomposition by substrate and decomposers: A comparison of model formulations. Biogeoscience 5:749–759.

26.4 Anaerobic Microbially Mediated Processes

Makoto Kimura

Susumu Asakawa

26.4.1 Decomposition Processes of Organic Materials under Anaerobic Conditions

Organisms capable of growing in anaerobic environments can be divided into three groups (Gottschalk and Peinemann, 1992): (1) organisms that are aerobes but can use inorganic oxidized molecules, such as NO_3^-, NO_2^-, MnO_2, and $Fe_2O_3/Fe(OH)_3$ as alternative electron acceptors; (2) organisms that are facultative aerobes but can carry out fermentations in the absence of oxygen and can only carry out the substrate-level phosphorylation under anaerobic conditions; (3) obligately anaerobic bacteria such as sulfate reducers, methanogens, and homoacetogens. In the decomposition of plant residue under waterlogged soil conditions, macromolecules, such as lignocellulose, proteins, and lipids, are generally catabolized to low-molecular-weight organic acids, alcohols, H_2, CO_2, and H_2O by microorganisms belonging to Groups 1 and 2, and these intermediate products are further decomposed to CO_2, CH_4, and H_2O by microorganisms belonging to Group 3.

Schink (1997) summarized the decomposition of polymer compounds and relative microbial decomposers in natural ecosystems (Figure 26.7); the complete conversion of polymeric compounds, for example, cellulose, to CH_4 and CO_2 can only occur by the concerted action of at least four different groups of bacteria,
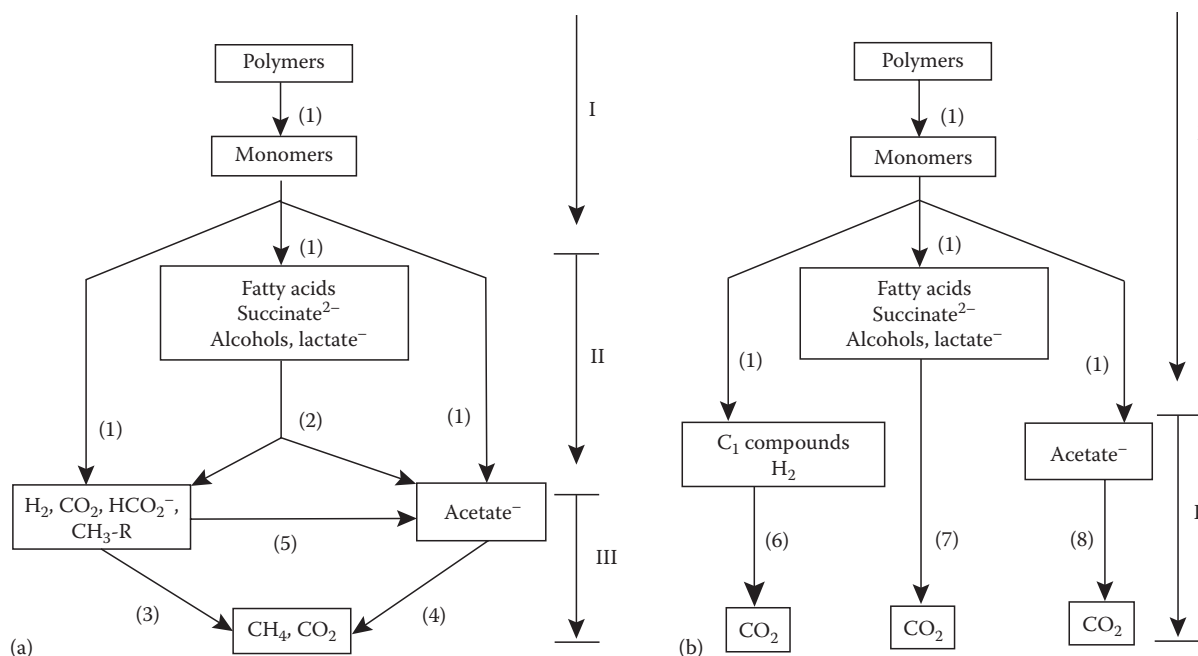
FIGURE 26.7 Scheme for the trophic groups of anaerobes in the (a) Methanogenic and (b) SO_4^{2-}-dependent degradation of complex organic matter. Groups of bacteria involved: (1) primary fermenters; (2) secondary fermenters (obligate syntrophs); (3) hydrogenotrophic methanogens; (4) acetotrophic methanogens; (5) homoacetogens; (6)–(8) SO_4^{2-} reducers. I, II, III: Steps in degradation. (From Kimura, M. 2000. Anaerobic microbiology in waterlogged rice fields, p. 35–138. *In* J.M. Bollag and G. Stotzky (eds.) Soil biochemistry. Vol. 10. Marcel Dekker, New York. With permission from Taylor & Francis Publishers.)

including primary fermenters, secondary fermenters, and two types of methanogens. Polymeric compounds are first degraded by primary fermenters (Group 1) to fermentation products (Step I) as well as to H_2, one-carbon compounds, and acetate. Acetate, H_2, and one-carbon compounds are converted to CH_4 directly by methanogens (Step III, Groups 3 and 4), whereas fatty acids with more than two carbon atoms, alcohols with more than one carbon atom, and branched-chain and aromatic fatty acids are degraded to acetate, H_2, and one-carbon compounds by secondary fermenters (Step II, Group 2). In either metabolic pathways for the degradation of polymeric compounds, H_2-scavenging bacteria (Group 3) have a key role in the overall mineralization, and the syntrophisms between organisms of Groups 2 and 3 are based in many cases on "interspecies H_2 transfer" (transfer of H_2 as an electron carrier between H_2-producing fermentative bacteria and hydrogenotrophic methanogens). Anaerobes of Steps II and III are generally restricted to substrates that they can metabolize, and they depend on fermenters (primary fermenters) (Ehrlich, 1993).

In SO_4^{2-}-rich anoxic habitats, SO_4^{2-} reducers are metabolically versatile and can use all products of primary fermentations and oxidize them to CO_2, simultaneously reducing SO_4^{2-} to S^{2-} (Figure 26.7b). Similar reactions in anoxic habitats can also be conducted by denitrifiers, MnO_2 reducers, and $Fe_2O_3/Fe(OH)_3$ reducers, and polymeric compounds may be metabolized to CO_2 in two steps (i.e., without the involvement of methanogens; Ehrlich, 1993).

26.4.2 H$_2$ Production as an Intermediate Product

Molecular hydrogen (H_2) occupies a unique position in anaerobic metabolism in anoxic soils as an intermediate product of fermentation, and then as a subsequent electron donor for anaerobic respiration. Hydrogen is a common metabolite produced by fermentative microorganisms, and it is used as an electron donor in the enzyme-mediated, energy-yielding reductions by many electron acceptors in soil, such as NO_3^-, MnO_2, $Fe_2O_3/Fe(OH)_3$, SO_4^{2-}, and CO_2. In both aerobic and anaerobic metabolisms of glucose to pyruvate, 2 mol of NADH, 2 mol of ATP, and 2 equivalent $[H^+]$ are produced from 1 mol of glucose (1 glucose $\rightarrow$ 2 pyruvate$^-$ + 2H$^+$ + 2H$_2$). The NADH produced is reoxidized to NAD^+ to continue the reaction, by NADH-linked hydrogenase, which only functions under the low partial pressures of H_2 (NADH + H$^+$ $\rightarrow$ NAD$^+$ + H$_2$). In aerobic microorganisms, O_2 consumes H_2 (2H$_2$ + O$_2$ $\rightarrow$ 2H$_2$O; $\Delta G^{o\prime}$ = −475 kJ mol^{-1} glucose at 25°C and pH 7). Fermentative microorganisms, which cannot use appropriate electron acceptors for scavenging H_2, reduce pyruvate to lactate or ethanol and regenerate NAD$^+$ (2 pyruvate$^-$ + 2H$_2$ $\rightarrow$ 2 lactate$^-$; 2 pyruvate$^-$ + 2H$_2$O + 2H$_2$ $\rightarrow$ 2 ethanol + 2HCO$_3^-$). $\Delta G^{o\prime}$ produced in the pyruvate production reaction is −112 kJ mol^{-1} of glucose at the standard condition (pH 7, 25°C), while the synthesis of 2NADH and 2ATP from 2NAD$^+$ and 2ADP consumes a metabolic energy of 184 kJ mol^{-1} of glucose (Thauer et al., 1977). This apparent

endoergonic reaction becomes exergonic only under conditions of low partial pressure of H_2. In soils, many microbial species can help fermentative microorganisms to regenerate NAD^+ by consuming H_2. The consumption/oxidation of H_2 coupled with appropriate electron acceptors can be associated to growth of fermentative microorganisms. This is called "interspecies H_2 transfer" (Bryant et al., 1967; Reddy et al., 1972). Hydrogen in anaerobic soils can be consumed by denitrifiers, MnO_2 reducers, $Fe_2O_3/Fe(OH)_3$ reducers, SO_4^{2-} reducers, methanogens, and homoacetogens; these reactions keep H_2 at low partial pressure. Consumption of H_2 by "syntrophic" microorganisms permits fermentative microorganisms to further oxidize pyruvate to acetate, producing additional 2 mol of NADH, $2H^+$, and 2 mol of ATP (2 pyruvate$^-$ + $2H_2O \rightarrow$ 2 acetate$^-$ + $2HCO_3^-$ + $2H^+$ + $2H_2$). Thus, the anaerobic decomposition of 1 mol of glucose to acetate produces 4 mol of NADH and 4 mol of ATP (Oremland, 1988).

26.4.3 Denitrification

Denitrification is the process by which NO_3^- and NO_2^- are reduced to nitrogenous gases, mainly to N_2O and N_2, according to the following pathway, $NO_3^- \rightarrow NO_2^- \rightarrow (NO) \rightarrow N_2O \rightarrow N_2$. Microorganisms capable of respiratory denitrification are called denitrifiers. Generally, the major environmental factors that control denitrification are presence of O_2, energy sources and NO_3^- and temperature, pH, and salinity. Waterlogged paddy soils may not experience inhibition of denitrification by acidity because of the pH increase to neutral values due to anaerobic conditions (Ponnamperuma, 1972). The redox potential of waterlogged paddy soils is another environmental factor controlling the reduction rate of NO_3^-.

Almost 130 bacterial species belonging to more than 50 genera can show denitrifying activity. However, half of the identified species are members of only three genera: 28 *Pseudomonas* spp., 13 *Neisseria* spp., and 12 *Bacillus* spp. Generally, only a few denitrifying species are known within a genus. Denitrifiers are predominantly Gram-negative eubacteria, except *Bacillus* spp. (Smith and Tiedje, 1980). Organic compounds are usually oxidized during denitrification, whereas inorganic compounds can only be oxidized by specific groups of microorganisms, for example, sulfur compounds by sulfur-oxidizing bacteria (*Thermothrix*, *Thiobacillus*, *Thiomicrospira*, and *Thiosphaera*), H_2 by hydrogenotrophic bacteria (*Alcaligenes*, *Bradyrhizobium*, *Paracoccus*, *Pseudomonas*, and *Rhizobium*), and NH_4^+ by nitrifying bacteria (*Nitrosomonas* and *Nitrobacter*) (Smith and Tiedje, 1980).

Denitrification coupled with the oxidation of NH_4^+ has been reported by Mulder et al. (1995) in a fluidized bed reactor; 3 mol of nitrate were required to oxidize 5 mol of NH_4^+ ($5NH_4^+$ + $3NO_3^- \rightarrow 4N_2 + 9H_2O + 2H^+$; −297 kJ mol^{-1} NH_4^+). This process, called the "Anammox process," can be carried out by bacteria belonging to the phylum Planctomycetes; three *Candidatus* genera (*Brocadia*, *Scalindua*, and *Kuenenia*) were detected in wastewater treatment plants and aquatic ecosystems (Hayatsu et al., 2008). The anammox process contributes significantly to the N_2 formation in marine environments (Dalsgaard et al., 2003),

freshwater sediments (Penton et al., 2006; Zhang et al., 2007; Dale et al., 2009), and in biofilms (Kindaichi et al., 2007), and, thus, it seems reasonable to hypothesize the presence of the process in anaerobic soils. However, anammox bacteria have not yet been reported in soil environments (Hayatsu et al., 2008).

Nitrous oxide (N_2O), an important greenhouse gas that is predominantly emitted from soil, is an intermediate product in the processes of nitrification and denitrification. Denitrification is the main process that produces N_2O in rice fields under waterlogged conditions. Some soil environmental factors can influence the nitrogenous gases produced (N_2 and N_2O) in the denitrification process. By increasing the concentrations of NO_3^-, NO_2^-, and O_2, it increases the proportion of N_2O produced. By decreasing pH and temperature, it also increases the production of N_2O. On the contrary, a decrease in N_2O proportion is caused by an increase in available carbon, and the mole fraction of N_2O produced decreases in the presence of plants. Conditions that decrease the proportion of N_2O prevail in paddy field soils under waterlogged condition. On the contrary, N_2O production in the nitrification process is predominant under upland soil conditions. It is produced as the by-product in hydroxylamine (NH_2OH) oxidation to NO_2^-, and the production increases by increasing the temperature and application of nitrogenous fertilizers. The N_2O production in the nitrification process is also affected by the kind of crops (Skiba and Smith, 2000).

26.4.4 Dissimilatory Nitrate Reduction to Ammonium

Dissimilatory reduction of nitrate to ammonia (DNRA) $\left(NO_3^- \rightarrow NO_2^- \rightarrow NH_4^+\right)$ is not inhibited by the presence of NH_4^+ or other N compounds. Microbes capable of DNRA include Enterobacteriaceae, or genera such as *Pseudomonas*, *Clostridium*, *Desulfobulbus*, *Desulfovibrio*, and *Bacillus*.

The standard free energy changes of denitrification (NO_3^- + H^+ + $2.5H_2 \rightarrow 0.5N_2 + 3H_2O$) and of DNRA ($NO_3^-$ + $2H^+$ + $4H_2 \rightarrow NH_4^+ + 3H_2O$) are −560.3 and −599.6 kJ per reaction, respectively (Stouthamer, 1988). Thus, the gain of energy from the reduction of each mole of NO_3^- would be similar between denitrification and DNRA, but the gain of ATP by DNRA is less than by denitrification. DNRA is conducted by two groups of microorganisms in soils: One group gets energy directly for cell growth through electron transport phosphorylation of the DNRA processes; one group cannot get energy directly by DNRA and uses this process as an electron sink to reoxidize NADH. In addition, the ATP yield by the first DNRA group is much less than that by denitrifiers due to a difference in the complexity of the respiratory chain (Stouthamer, 1988). DNRA is considered to work as an electron sink in soils.

Tiedje et al. (1982) pointed out the absence of denitrifiers and the predominance of DNRA-microorganisms in the rumen. By considering the relative predominance of DNRA-microorganisms over denitrifiers in several ecosystems with active NO_3^- reduction, they proposed a conceptual model with the dominance of the DNRA process over denitrification in strongly anaerobic habitats, such as the rumen, where the shortage of electron

acceptors is probably the most limiting factor for microbial growth, and DNRA is more favored because it consumes three additional electrons per mole of NO_3^-. On the contrary, in less anaerobic habitats, denitrification is more favored for microbial growth because of the more efficient production of energy per hydrogen supplied in the form of NADH. In general, in water-logged rice fields, denitrification is the dominant process due to the relatively low supply of substrates for the microbial growth by DNRA. Digested sludge, estuarine sediments, and lake sediments present environmental conditions, which are intermediate between those of rumen and those of waterlogged rice fields.

26.4.5 Reduction of MnO_2

Many bacteria, such as *Bacillus*, *Pseudomonas*, *Leptothrix*, and *Micrococcus*, and fungi, such as *Aspergillus*, *Penicillium*, and *Pichia*, can conduct respiration/reduction of MnO_2 (Ghiorse, 1988). Sulfate reducers, such as *Desulfovibrio desulfuricans*, *Desulfomicrobium baculatus*, *Desulfobacterium autotrophicum*, *Desulfuromonas acetoxidans*, and *Geobacter metallireducens*, can oxidize S^0 to SO_4^{2-} when MnO_2 acts as an electron acceptor ($S^0 + 3MnO_2 + 4H^+ \rightarrow SO_4^{2-} + 3Mn^{2+} + 2H_2O$; Lovley and Phillips, 1994). However, several microbial species isolated from paddy soil are also able to reduce MnO_2 in liquid medium by microbial metabolites, indicating that this reaction is widespread among soil microorganisms inhabiting these soils (Kamura and Yoshida, 1972). MnO_2 can also be abiotically reduced by Fe^{2+}. Thus, enzyme-mediated direct reduction (respiratory reduction), indirect reduction by metabolites and chemical reduction are processes reducing MnO_2 in anaerobic soils.

26.4.6 Iron Reduction

Iron reduction in anaerobic soils can involve direct and indirect biological processes. The role of $Fe_2O_3/Fe(OH)_3$ as the terminal electron acceptor was suggested by a significant correlation between the rate of consumption of added glucose and the increase in the amount of Fe^{2+} production in waterlogged paddy soils (Yamanaka and Motomura, 1959). The production of Fe^{2+} was completely inhibited, and a visible change from brownish to grayish soil color due to Fe reduction was not observed when biocides such as NaN_3, $HgCl$, or monoiodacetic acid were added to soil (Yamanaka and Motomura, 1959).

The direct involvement of $Fe_2O_3/Fe(OH)_3$ as a H_2 acceptor in respiration has been shown in bacteria belonging to *Pseudomonas*, *Alteromonas*, *Geobacter*, *Desulfuromonas*, *Pelobacter*, and *Acidiphilium*, which can obtain energy for their growth by the coupling oxidation of H_2 to reduction of $Fe_2O_3/Fe(OH)_3$ (Kimura, 2000). Lovley et al. (1995) suggested that the use of $Fe_2O_3/Fe(OH)_3$ as a terminal electron acceptor can be a characteristic of the *Geobacter–Desulfuromonas–Pelobacter* branch of the δ Proteobacteria. Lovley (1993, 1995) proposed a model for the anaerobic oxidation of organic matter to CO_2 with $Fe_2O_3/Fe(OH)_3$ as the sole electron acceptor; fermentative microorganisms metabolize sugars and amino acids to short-chain fatty acids and H_2. These fermentation products are then oxidized with the reduction of $Fe_2O_3/Fe(OH)_3$.

Microorganisms reducing $Fe_2O_3/Fe(OH)_3$ can use diverse kind of compounds as electron acceptors. For example, $Fe_2O_3/Fe(OH)_3$-reducing *Pseudomonas* sp. strain 200 can reduce NO_3^-, ClO_3^-, MnO_4^-, CrO_7^{2-}, SO_3^{2-}, and $S_2O_3^{2-}$ and thus limit electron flow to $Fe_2O_3/Fe(OH)_3$ (Obuekwe and Westlake, 1982).

Several $Fe_2O_3/Fe(OH)_3$ reducers are present in soils (Ottow, 1969). However, since a few microbial groups can use $Fe_2O_3/Fe(OH)_3$ directly as an electron acceptor, reduction of $Fe_2O_3/Fe(OH)_3$ in soil mainly occurs by indirect microbial reduction. In addition, chemical reduction of $Fe_2O_3/Fe(OH)_3$ by S^{2-} has been shown in soil (Murase and Kimura, 1997). Lovley et al. (1996) found that the $Fe_2O_3/Fe(OH)_3$-reducing microorganisms, *G. metallireducens* and *Shewanella alga*, can use humic substances as an electron acceptor to reduce $Fe_2O_3/Fe(OH)_3$. Reduction of humic substances can also occur by indigenous soil bacteria under anaerobic soil conditions (Lovley et al., 1996). Thus, both direct (enzyme-mediated) and indirect microbial reduction and chemical reduction of $Fe_2O_3/Fe(OH)_3$ can occur in soil under anaerobic conditions.

26.4.7 Sulfate Reduction

Sulfate-reducing bacteria can only be active under strict anaerobic conditions. They include 12 genera of eubacteria, such as *Desulfovibrio*, *Desulfomicrobium*, *Desulfobulbus*, *Desulfobacter*, *Desulfococcus*, *Desulfosarcina*, *Desulfomonile*, *Desulfonema*, *Desulfobotulus*, and *Desulfoarculus* (Gram-negative, nonspore-forming bacteria of the δ subdivision of Proteobacteria), *Desulfotomaculum* (a Gram-positive, sporeforming bacterium), and *Thermodesulfobacterium* (a unique bacterium containing ether lipids). Many of these bacteria can oxidize H_2 and lactate by reducing SO_4^{2-}, and all strains can use not only SO_4^{2-} but also SO_3^{2-}; most of these strains can also reduce $S_2O_3^{2-}$. Acetate is only used by *Desulfobacter* and by some species of *Desulfotomaculum*. Reduction of S^0 is relatively rare for this bacterial group and is restricted to *Desulfovibrio gigas* and a *Desulfomicrobium* sp. (Biebl and Pfennig, 1977; Hansen, 1988). More than 65 compounds are known to be utilized by SO_4^{2-} reducers as energy sources, including H_2, CO, monocarboxylic acids, dicarboxylic acids, alcohols, amino acids, and other miscellaneous compounds (Hansen, 1988).

According to Widdel (1988), SO_4^{2-} reducers are divided into two metabolic groups: (1) species oxidizing their substrates incompletely to acetate ("classical" lactate-oxidizing species, such as *Desulfotomaculum* and *Desulfovibrio*); and (2) species able to oxidize their organic substrates, including acetate, completely to CO_2 (*Desulfobacter*, *Desulfococcus*, *Desulfosarcina*, *Desulfonema*, and *Desulfobacterium*). Therefore, sulfate reducers in the former and latter groups may belong to Groups 1 and 2, and Groups 1, 4, 6, 7, and 8 of Figure 26.7, respectively.

Strains of *Desulfovibrio*, *Desulfomonas*, and *Desulfotomaculum* are the most common SO_4^{2-} reducers in soils. These bacteria only metabolize incompletely a limited number of substrates, such as lactate, ethanol, or dicarboxylic acids to acetate. Therefore,

syntrophy with associated microorganisms is a common way of life for SO_4^{2-} reducers. The syntrophy of SO_4^{2-} reducers with associated microorganisms involves three facets: substrate-yielding, proton-discharging, and proton-scavenging.

Desulfovibrio sulfodismutans disproportionated 1 mol of SO_3^{2-} to 0.75 mol of SO_4^{2-} and 0.25 mol of S^{2-} ($SO_3^{2-} + H^+ \rightarrow 0.75SO_4^{2-} + 0.25HS^-$) and $S_2O_3^{2-}$ to equal amounts of SO_4^{2-} and S^{2-} ($S_2O_3^{2-} + H_2O \rightarrow SO_4^{2-} + HS^- + H^+$) (Bak and Cypionka, 1987; Bak and Pfennig, 1987). Bacteria able to disproportionate inorganic S compounds appear to be widely distributed in nature. Bak and Cypionka (1987) counted by the MPN method 2×10^6 cells of $S_2O_3^{2-}$-disproportionating bacteria mL^{-1} of mud in a freshwater ditch sediment. The disproportionation of $S_2O_3^{2-}$ was probably a hydrolytic process by SO_4^{2-} reducers with the inner and the outer S atoms transformed into SO_4^{2-} and S^{2-}, respectively; $S_2O_3^{2-}$ is a key intermediate in the S cycle, being a main product of anoxic S^{2-} oxidation and a shunt between oxidative and reductive pathways (Jørgensen, 1990).

26.4.8 Sulfur Reducers

Both eubacteria and archaeobacteria (Crenarchaeota) can be S (S^0) reducers (Widdel and Pfennig, 1992). Organic substrates, mainly simple organic acids, or H_2 are used as electron donors (Rabus et al., 2006). The archaeobacterial members are extremely thermophilic, may not have a meaningful role in S^0 metabolism in soil (Rabus et al., 2006), and have been isolated exclusively from solfataric soils (Huber and Prangishvili, 2006). Contrary to SO_4^{2-} reducers, some eubacteria and archaeobacteria conducting respiratory S^0 reduction can grow aerobically (Rabus et al., 2006). A *Pseudomonas mendonica*-like strain and *Alteromonas putrefaciens* can use $Fe_2O_3/Fe(OH)_3$, MnO_2, and NO_3^- as electron acceptors, as well as S^0 and $S_2O_3^{2-}$ (Widdel and Hansen, 1992).

A chemolithotrophic pathway, in which S^0 is oxidized to SO_4^{2-} in the presence of either MnO_2 or FeOOH, was discovered in anaerobic marine enrichment cultures, where S^0 was metabolized disproportionately to H_2S and SO_4^{2-} ($4S^0 + 4H_2O \rightarrow SO_4^{2-} + 3H_2S + 2H^+$) (Thamdrup et al., 1993). *Desulfobulbus propionicus* was found to disproportionate S^0 to SO_4^{2-} and S^{2-} (Lovley and Phillips, 1994). When the sediment from a tidal flat was amended with S^0, the production of both SO_4^{2-} and S^{2-} was consistent with the ratio that was expected from the disproportionation reaction ($3S^0 + 2Fe(OH)_3 \rightarrow SO_4^{2-} + 2FeS + 2H_2O + 2H^+$) (Canfield and Thamdrup, 1996).

In the presence of acetate or ethanol, S^0 is shared as an electron-carrying catalyst in the syntrophism between *Desulfuromonas* and green sulfur bacteria (Pfennig and Biebl, 1976), and as little as 0.25 mM of S^{2-} or S^0 was sufficient to maintain their optimal syntrophic growth (Biebl and Pfennig, 1978).

26.4.9 Methanogenesis and Homoacetogenesis

In recent years, CH_4 has received great attention as one of the greenhouse gases. Rice fields are one of the major emission sources of atmospheric CH_4 with an annual emission of 31–112 Tg $year^{-1}$, a considerable output since the total annual CH_4 emission is 582 Tg $year^{-1}$ (Denman et al., 2007). The CH_4 emitted from rice fields probably derives from soil organic matter, rice plants, and applied organic materials (rice straw) (Kimura et al., 2004). In the waterlogged soil pot experiment, the percentages of total CH_4 released from rice soils were 42% from rice straw, 37%–40% from rhizodeposition, and 18%–21% from soil organic matter when the rice straw was applied (6 g kg^{-1} soil in pot) to soil; and 15%–20% from soil organic matter and remaining 80%–85% from rhizodeposition without rice straw application (Watanabe et al., 1999).

Methanogens are a large and diverse group of microorganisms characterized by the production of large quantities of CH_4 as the major product of their energy metabolism; in addition, they are strict anaerobes and are members of the domain Archaea or archaeobacteria (Whitman et al., 1992). Methanogenic archaea comprise five orders: Methanobacteriales, Methanococcales, Methanomicrobiales, Methanosarcinales, and Methanopyrales; members of Methanobacteriales, Methanococcales, Methanomicrobiales, and Methanopyrales can use H_2 and formate (or certain alcohols) as energy sources, while members of the order Methanosarcinales can use a variety of methyl-containing C-1 compounds (Garcia et al., 2000). Acetate is the major source of CH_4 for *Methanosarcina* and *Methanosaeta* (*Methanothrix*) species of Methanosarcinales. The methyl carbon of acetate is reduced to CH_4, and the carboxyl carbon is oxidized to CO_2. The first isolate of the lineage "Rice cluster I," which is commonly detected in paddy field soil by molecular analysis, has been described as *Methanocella paludicola* in a novel order, Methanocellales (Sakai et al., 2008).

Methanogenic activity usually increases sharply after flooding of rice fields, but the methanogenic population survives well under dry and aerobic conditions of rice fields and their numbers do not fluctuate significantly between the fallow season (drained conditions) and the rice cultivation period (waterlogged conditions). There has been a disjunct between the methanogenic populations measured by the MPN method and its activities in rice fields. Fetzer et al. (1993) speculated that the MPN method only detects viable cells of bacterial aggregates, and O_2 might have killed some of the cells localized on external part of aggregate (i.e., the peripheral cells), whereas cells localized in the internal part (i.e., central cells) may have survived.

As each mole of glucose is metabolized to 2 mol of acetate and 4 mol of NADH ($C_6H_{12}O_6 + 2H_2O \rightarrow 2CH_3COO^- + 2HCO_3^- + 4NADH + 4H^+$), two-thirds of the CH_4 in rice fields is theoretically derived from acetate and one-third from CO_2/H_2. In addition, as the partial pressures of H_2 in waterlogged rice fields in situ usually are very low (1–10 Pa) in comparison with K_m values (2.5–12 μM) of hydrogenotrophic methanogenesis (Kimura, 2000) and concentrations of acetate are generally high enough to meet K_m values of acetotrophic methanogenesis (Rothfuss and Conrad, 1993), CH_4 production from acetate seems to be more favorable in rice fields than hydrogenotrophic methanogenesis. However, the number of microorganisms capable of hydrogenotrophic methanogenesis is often higher and their contribution to CH_4 production is often greater than that of microorganisms capable of acetotrophic methanogenesis in rice fields (Schütz et al., 1989; Mayer and Conrad, 1990).

26.4.10 Homoacetogenesis

Homoacetogens are a microbial group that use energy produced by the formation of acetate from H_2 and CO_2 ($2CO_2 + 4H_2 \rightarrow CH_3COOH + 2H_2O$). There are three types of homoacetogenesis: (1) those growing heterotrophically on organic compounds such as sugars, and using the synthesis of acetate from CO_2 as an electron acceptor for metabolic oxidation reactions; (2) those using H_2 and CO_2 as the sole energy sources, and using acetogenesis for the generation of metabolic energy; and (3) those growing on CO as the sole energy source, and producing both H_2 and the carboxyl group from the reaction of CO with H_2O (Diekert, 1992). All of these homoacetogens are strictly anaerobes containing the key enzyme, carbon monoxide dehydrogenase, as well as an enzyme-bound methylcobalamin. Taxonomically, the homoacetogens are an extremely heterogeneous group and include genera of *Acetobacterium*, *Acetogenium*, *Butyribacterium*, *Clostridium*, *Eubacterium*, *Peptostreptococcus*, and *Sporomusa* (Fuchs, 1986; Diekert, 1992).

Chin and Conrad (1995) studied intermediary metabolism in methanogenic rice soil slurry incubated at 15°C and 30°C and found that the lowest temperature decreased the CH_4 production rate with a transient accumulation of acetate, propionate, caproate, lactate, and isopropanol. The added H_2 was consumed, together with CO_2, mainly by methanogenesis at 30°C and mainly by homoacetogenesis at 15°C. The H_2-producing reactions are more sensitive to low temperature than the H_2-consuming reactions, and H_2 consumption by methanogenesis is more sensitive than H_2 consumption by homoacetogenesis (Chin and Conrad, 1995).

Contribution of homoacetogenesis to CO_2/H_2 metabolism is small in a waterlogged Italian rice field (Rothfuss and Conrad, 1993). Similar small contribution of H_2-oxidizing homoacetogens to the consumption of H_2 was observed in sediments of a eutrophic lake, and less than 2% of the total production of acetate in a eutrophic lake sediment was estimated to be derived from homoacetogenesis (Lovley and Klug, 1983). The ratios of methanogens to homoacetogens in lake sediments were ca. 100–1 (Braun et al., 1979).

26.4.11 Competition and Syntrophism among Microbial Groups in Relation to Sequential Reduction

The type of microbial metabolism in waterlogged rice soils changes sequentially according to the oxidation state, from respiration in the presence of O_2 and NO_3^- (most efficient energy-yielding processes) to SO_4^{2-} reduction and methanogenesis (least efficient energy-yielding processes). The Eh_7 of the soil decreases sequentially, according to the respective metabolic processes, from +0.6 to +0.5 V to −0.2 to −0.3 V after submergence (Table 26.2). "Sequentially" does not mean that a subsequent reduction process only starts after the termination of the previous one, but that the time of onset of the respective reduction process occurs sequentially (Kimura, 2000).

TABLE 26.2 Approximate Redox Potentials (mV) at Which the Respective Reactions Start

Reaction	Takai and Kamura (1966)	Takai (1980)	Patrick and Reddy (1978)
Oxygen consumption	+600 to +500	+600 to +300	+380 to +320
Denitrification	+600 to +500	+400 to +100	+280 to +220
Mn^{2+} formation	+600 to +400	+400 to −100	+280 to +220
Fe^{2+} formation	+500 to +300	+200 to −200	+180 to +150
Sulfate reduction	0 to −190	0 to −200	−120 to −180
Methane production	−150 to −190	−200 to −300	−200 to −280
Hydrogen formation	−150 to −220		

Hydrogen is a common electron donor for all microorganisms involved in the sequential reduction processes in anaerobic soils, as explained before. Formate and acetate are the other common electron donors. Competition for H_2, formate, and acetate is, therefore, expected to occur among bacterial groups in anaerobic soils. Differences in K_m and V_{max} of the Michaelis–Menten kinetics of the respective reduction processes seem to explain the sequential reduction commonly observed in waterlogged rice fields. However, the Michaelis–Menten kinetics is only applicable to the competition between SO_4^{2-} reducers and methanogens in the environments with concentrations of H_2 that are 20- to 100-fold greater than those anaerobic soils. Lovley and Goodwin (1988) found that concentrations of H_2 at steady-state conditions depended primarily on the physiological characteristics of microorganism(s) consuming H_2; in addition, they found that microorganisms catalyzing the oxidation of H_2, with the reduction of a more electrochemically positive electron acceptor, can maintain lower concentrations of H_2 than microorganisms using electron acceptors that are reduced at a lower Eh by H_2. In other words, microorganisms responsible for the reduction have their own specific "threshold partial pressures of H_2," and the specific process only occurs above the inherent threshold partial pressure of H_2. The ambient partial pressure of H_2 remains near the specific threshold partial pressure of H_2 as the process continues, and the partial pressure of H_2 increases to higher levels stepwise when more electrochemically positive electron acceptors are consumed. The threshold partial pressure of H_2 is the pressure at which microorganisms metabolizing H_2 can obtain the minimum amount of energy for their growth. It decreases by increasing Gibbs standard free energy $\Delta G^{o\prime}$ obtained from the metabolism of H_2 by electron acceptors. The $\Delta G^{o\prime}$ values of the various processes expressed per mole of glucose at 25°C and pH 7.0 (Thauer et al., 1977) are as follows:

Denitrification ($4H_2 + 1.6H^+ + 16NO_3^- \rightarrow 0.8N_2 + 4.8H_2O$)	$\Delta G^{o\prime} = -896\,kJ$
MnO_2 reduction ($2H_2 + 12H^+ + 4MnO_2 \rightarrow 4Mn^{2+} + 8H_2O$)	$\Delta G^{o\prime} = -556\,kJ$
Fe(III) reduction ($4H_2 + 8Fe^{3+} \rightarrow 8Fe^{2+} + 8H^+$)	$\Delta G^{o\prime} = -275\,kJ$
Sulfate reduction ($4H_2 + SO_4^{2-} + H^+ \rightarrow HS^- + 4H_2O$)	$\Delta G^{o\prime} = -152\,kJ$
Methanogenesis ($4H_2 + HCO_3^- + H^+ \rightarrow CH_4 + 3H_2O$)	$\Delta G^{o\prime} = -136\,kJ$
Homoacetogenesis ($4H_2 + 2HCO_3^- + H^+ \rightarrow CH_3COO^- + 4H_2O$)	$\Delta G^{o\prime} = -105\,kJ$

The threshold partial pressures of H_2 under anaerobic conditions at which microorganisms can carry out the various sequential reduction processes are <0.05 nM for denitrification, <0.05 nM for MnO_2 reduction, 0.03–0.04 Pa for Fe_2O_3/$Fe(OH)_3$ reduction, 0.17–1.9 Pa for sulfate reduction, 0.6–10.6 Pa for methanogenesis, and 51–94 Pa for homoacetogenesis. The threshold partial concentrations of acetate can range from 0.5 to 2.2 μM for sulfate reduction and from 5.2 to 3700 μM for methanogenesis (Kimura, 2000).

26.4.12 Anaerobic Reoxidation of Electron Acceptors

The theory of sequential reduction processes implies that the concentration of reduced products, such as Mn^{2+}, Fe^{2+}, S^{2-}, and CH_4, in soil can increase as the respective reduction processes proceed under anaerobic conditions. However, when several reduction processes take place concurrently in different microsites of soil, the reduced products produced at more reduced sites may be reoxidized at neighboring, less reduced, sites. Several studies have demonstrated the reoxidation of reduced products in anoxic rice soils.

Yoshida and Kamura (1975) observed the oxidation of Fe^{2+} by MnO_2 in a submerged rice soil, and the amount of oxidized Fe^{2+} was equivalent to that of MnO_2 added. When Fe^{2+} was added as $(NH_4)_2Fe(SO_4)_2$ to aerobic paddy soil samples, MnO_2 was reduced to Mn^{2+} chemically with oxidation of Fe^{2+} to Fe_2O_3/$Fe(OH)_3$ (Murase and Kimura, 1997).

Elemental sulfur (S^0) added to rice soil with $NaNO_3$ was oxidized to SO_4^{2-} with reduction of NO_3^- under in waterlogged conditions (Murase and Kimura, 1997). Garcia-Gil and Golterman (1993) observed the FeS-mediated denitrification with SO_4^{2-} formation in river delta sediments, and both NO_3^- reduction and FeS oxidation followed the Michaelis–Menten kinetics (Dannenberg et al., 1992).

Monosulfide (S^{2-}) was reported to reduce $Fe(OH)_3$ chemically with the formation of polysulfide or S^0 in acid sulfate soils (Pons et al., 1982). Addition of S^{2-} ion as Na_2S to aerobic paddy soil samples chemically reduced MnO_2 and Fe_2O_3/$Fe(OH)_3$ to Mn^{2+} and Fe^{2+} (Murase and Kimura, 1997). Lovley and Phillips (1994) also observed S^0 oxidation to SO_4^{2-} coupled with MnO_2 reduction by many SO_4^{2-} reducers.

More than 70% of CH_4 leached from the soil pot disappeared in the subsoil column, indicating anoxic oxidation of CH_4 in a soil pot experiment with and without rice plants over the entire period of rice cultivation (Murase and Kimura, 1994).

Only few studies on the oxidation of CH_4 in anoxic paddy soils have been reported whereas many studies have shown the anoxic oxidation of CH_4 in fresh and marine waters and sediments, with anaerobic CH_4 oxidizers and SO_4^{2-} reducers working in a syntrophic association. Methane oxidation coupled with SO_4^{2-} reduction is possible thermodynamically ($CH_4 + SO_4^{2-} \rightarrow HCO_3^- + HS^- + H_2O$, $\Delta G^{o\prime} = -16.6$ kJ). The anaerobic CH_4 oxidizers belong to the domain Archaea and are phylogenetically close to methanogens of order Methanosarcinales. They are called "ANME" (anaerobic methane oxidizers) but they have not yet been isolated (Conrad, 2007).
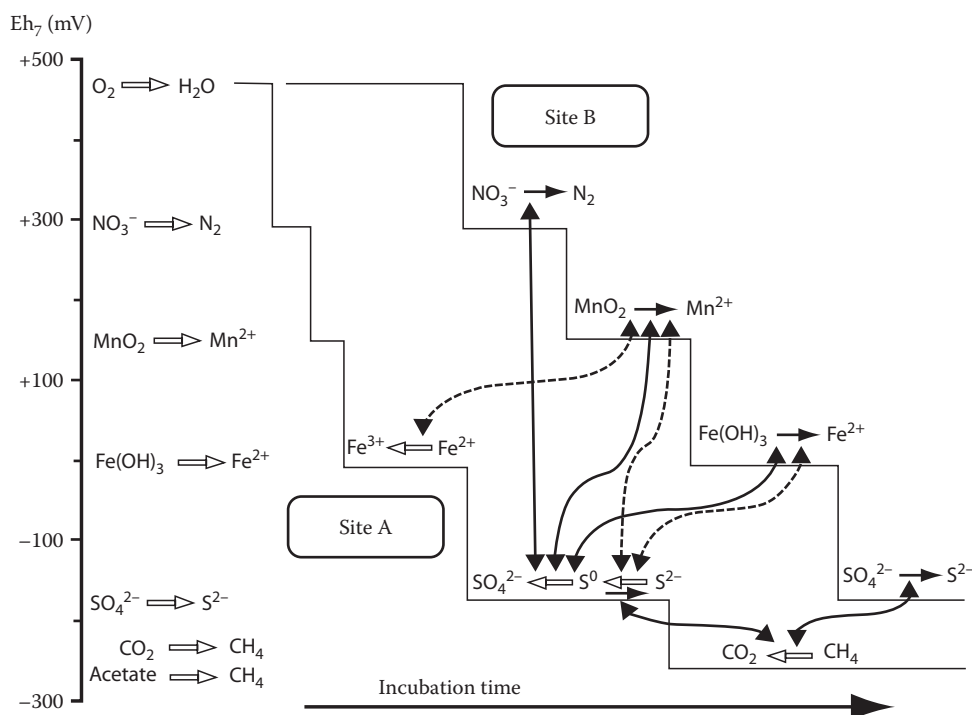
FIGURE 26.8 Conceptual model of reduction processes in submerged rice soils. Solid lines represent biological reactions and dotted lines represent chemical reactions. (From Murase, J., and M. Kimura. 1997. Anaerobic reoxidation of Mn^{2+}, Fe^{2+}, S^0 and S^{2-} in submerged paddy soils. Biol. Fertil. Soils 25:302–306. With permission.)

Results of research mentioned above suggest reconsideration and revision of the prevailing simple reduction sequence model of waterlogged rice soils to a more dynamic model as shown in Figure 26.8, in which two sites with different reduction state are considered (Murase and Kimura, 1997). In the site A, the reduction processes are fast whereas in the site B the processes are slower. Solid and dotted lines indicate microbiological and chemical reoxidation reactions of reduced electron acceptors, respectively, such as CH_4, S^{2-}, S^0, and Fe^{2+}, formed at Site A by SO_4^{2-}, $Fe_2O_3/Fe(OH)_3$ (shown as Fe^{3+} in Figure 26.8), MnO_2, and NO_3^- at Site B. It is reasonable to imagine that these two sites are in communication in soil by water diffusion, water percolation, and water absorption by rice roots. Although it is not yet elucidated how far anaerobic reoxidation occurs in submerged paddy soils, the oxidation–reduction reactions involving "reduced electron acceptors" formed in sites with more reduced states (Site A) and "electron acceptors" of sites with less reduced states (Site B) may balance oxidation–reduction states of waterlogged paddy soil.

References

Bak, F., and H. Cypionka. 1987. A novel type of energy metabolism involving fermentation of inorganic sulphur compounds. Nature 326:891–892.

Bak, F., and N. Pfennig. 1987. Chemolithotrophic growth of *Desulfovibrio sulfodismutans* sp. nov. by disproportionation of inorganic sulfur compounds. Arch. Microbiol. 147:184–189.

Biebl, H., and N. Pfennig. 1977. Growth of sulfate-reducing bacteria with sulfur as electron acceptor. Arch. Microbiol. 112:115–117.

Biebl, H., and N. Pfennig. 1978. Growth yields of green sulfur bacteria in mixed cultures with sulfur and sulfate reducing bacteria. Arch. Microbiol. 117:9–16.

Braun, M., S. Schoberth, and G. Gottschalk. 1979. Enumeration of bacteria forming acetate from H_2 and CO_2 in anaerobic habitats. Arch. Microbiol. 120:201–204.

Bryant, M.P., E.A. Wolin, M.J. Wolin, and R.S. Wolfe. 1967. *Methanobacillus omelianskii*, a symbiotic association of two species of bacteria. Arch. Microbiol. 59:20–31.

Canfield, D.E., and B. Thamdrup. 1996. Fate of elemental sulfur in an intertidal sediment. FEMS Microbiol. Ecol. 19:95–103.

Chin, K.J., and R. Conrad. 1995. Intermediary metabolism in methanogenic paddy soil and the influence of temperature. FEMS Microbiol. Ecol. 18:85–102.

Conrad, R. 2007. Microbial ecology of methanogens and methanotrophs. Adv. Agron. 96:1–63.

Dale, O.R., C.R. Tobias, and B. Song. 2009. Biogeographical distribution of diverse anaerobic ammonium oxidizing (anammox) bacteria in Cape Fear River Estuary. Environ. Microbiol. 11:1194–1207.

Dalsgaard, T., D.E. Canfield, J. Petersen, B. Thamdrup, and J. Acuña-González. 2003. N_2 production by the anammox reaction in the anoxic water column of Golfo Dulce, Costa Rica. Nature 422:606–608.

Dannenberg, S., M. Kroder, W. Dilling, and H. Cypionka. 1992. Oxidation of H_2, organic compounds and inorganic sulfur compounds coupled to reduction of O_2 or nitrate by sulfate-reducing bacteria. Arch. Microbiol. 158:93–99.

Denman, K.L., G. Brasseur, A. Chidthaisong, P. Ciais, P.M. Cox, R.E. Dickinson, D. Hauglustaine et al. 2007. Couplings between changes in the climate system and biogeochemistry, p. 499–587. *In* S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller (eds.) Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, U.K.

Diekert, G. 1992. The acetogenic bacteria, p. 517–533. *In* A. Balows, H. Dworkin, W. Harder, and K.H. Schleifer (eds.) The Prokaryotes. A handbook on the biology of bacteria: Ecophysiology, isolation, identification, applications. 2nd edn. Vol. 1. Springer-Verlag, New York.

Ehrlich, H.L. 1993. Bacterial mineralization of organic carbon under anaerobic conditions, p. 219–247. *In* J.M. Bollag and G. Stotzky (eds.) Soil biochemistry. Vol. 8. Marcel Dekker, New York.

Fetzer, S., F. Bak, and R. Conrad. 1993. Sensitivity of methanogenic bacteria from paddy field soil to oxygen and desiccation. FEMS Microbiol. Ecol. 12:107–115.

Fuchs, G. 1986. CO_2 fixation in acetogenic bacteria: Variations on a theme. FEMS Microbiol. Rev. 39:181–213.

Garcia, J.-L., B.K.C. Patel, and B. Ollivier. 2000. Taxonomic, phylogenetic, and ecological diversity of methanogenic archaea. Anaerobe 6:205–226.

Garcia-Gil, L.J., and H.L. Golterman. 1993. Kinetics of FeS-mediated denitrification in sediments from the Camargue (Rhone delta, southern France). FEMS Microbiol. Ecol. 13:85–92.

Ghiorse, W.C. 1988. Microbial reduction of manganese and iron, p.305–331. *In* A.J.B. Zehnder (ed.) Biology of anaerobic microorganisms. John Wiley & Sons. New York.

Gottschalk, G., and S. Peinemann. 1992. The anaerobic way of life, p. 301–311. *In* A. Balows, H. Dworkin, W. Harder, and K.H. Schleifer (eds.) The prokaryotes. A handbook on the biology of bacteria: Ecophysiology, isolation, identification, applications. 2nd edn. Vol. 1. Springer-Verlag, New York.

Hansen, T.A. 1988. Physiology of sulphate-reducing bacteria. Microbiol. Sci. 5:81–84.

Hayatsu, M., K. Tago, and M. Saito. 2008. Various players in the nitrogen cycle: Diversity and functions of the microorganisms involved in nitrification and denitrification. Soil Sci. Plant Nutr. 54:33–45.

Huber, H., and D. Prangishvili. 2006. Sulfolobales, p. 23–51. *In* M. Dwarkin, S. Falkow, E. Rosenberg, K.H. Schleifer, and E. Stackebrandt (eds.) The prokaryotes. A handbook on the biology of bacteria. 3rd edn. Vol. 3. Springer Science+Business Media, New York.

Jørgensen, B.B. 1990. The sulfur cycle of freshwater sediments: Role of thiosulfate. Limnol. Oceanogr. 35:1329–1342.

Kamura, T., and K. Yoshida. 1972. Manganese reducing micro-organisms in soils and the reduction mechanism of manganese in the culture solution. Reduction mechanism of manganese in paddy soils (III). J. Sci. Soil Manure, Japan. 43:451–455.

Kimura, M. 2000. Anaerobic microbiology in waterlogged rice fields, p. 35–138. *In* J.M. Bollag and G. Stotzky (eds.) Soil biochemistry. Vol. 10. Marcel Dekker, New York.

Kimura, M., J. Murase, and Y. Lu. 2004. Carbon cycling in rice field ecosystem in the context of input, decomposition and translocation of organic materials and the fates of their end-products (CO_2 and CH_4). Soil Biol. Biochem. 36:1399–1416.

Kindaichi, T., I. Tsushima, Y. Ogasawara, M. Shimokawa, N. Ozaki, H. Satoh, and S. Okabe. 2007. In situ activity and spatial organization of anaerobic ammonium-oxidizing (Anammox) bacteria in biofilms. Appl. Environ. Microbiol. 73:4931–4939.

Lovley, D.R. 1993. Dissimilatory metal reduction. Annu. Rev. Microbiol. 47:263–290.

Lovley, D.R. 1995. Microbial reduction of iron, manganese, and other metals. Adv. Agron. 54:175–231.

Lovley, D.R., J.D. Coates, E.L.B. Blunt-Harris, E.J.P. Phillips, and J.C. Woodward. 1996. Humic substances as electron acceptors for microbial respiration. Nature 382:445–448.

Lovley, D.R., and S. Goodwin. 1988. Hydrogen concentrations as an indicator of the predominant terminal electron-accepting reactions in aquatic sediments. Geochim. Cosmochim. Acta 52:2993–3003.

Lovley, D.R., and M.J. Klug. 1983. Methanogenesis from methanol and methylamines and acetogenesis from hydrogen and carbon dioxide in the sediments of a eutrophic lake. Appl. Environ. Microbiol. 45:1310–1315.

Lovley, D.R., and E.J.P. Phillips. 1994. Novel processes for anaerobic sulfate production from elemental sulfur by sulfate-reducing bacteria. Appl. Environ. Microbiol. 60:2394–2399.

Lovley, D.R., E.J.P. Phillips, D.J. Lonergan, and P.K. Widman. 1995. Fe(III) and S^0 reduction by *Alteromonas putrefaciens*. Appl. Environ. Microbiol. 61:2132–2138.

Mayer, H.P., and R. Conrad. 1990. Factors influencing the population of methanogenic bacteria and the initiation of methane production upon flooding of paddy soil. FEMS Microbiol. Ecol. 73:103–112.

Mulder, A., A.A. van de Graaf, L.A. Robertson, and J.G. Kuenen. 1995. Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. FEMS Microbiol. Ecol. 16:177–184.

Murase, J., and M. Kimura. 1994. Methane production and its fate in paddy fields. VII. Electron acceptors responsible for anaerobic methane oxidation. Soil Sci. Plant Nutr. 40:647–654.

Murase, J., and M. Kimura. 1997. Anaerobic reoxidation of Mn^{2+}, Fe^{2+}, S^0 and S^{2-} in submerged paddy soils. Biol. Fertil. Soils 25:302–306.

Obuekwe, C.O., and D.W.S. Westlake. 1982. Effect of reducible compounds (potential electron acceptors) on reduction of ferric iron by *Pseudomonas* species. Microbios. Lett. 19:57–62.

Oremland, R.S. 1988. Biogeochemistry of methanogenic bacteria, p. 641–705. *In* A.J.B. Zehnder (ed.) Biology of anaerobic microorganisms. John Wiley & Sons, New York.

Ottow, J.C.G. 1969. The distribution and differentiation of iron-reducing bacteria in gley soils. Zentralbl. Bakteriol. Parasitenkde Infektionskrankh. Hyg. Zweitenaturwissenschaftliche-Abteilung 123:600–615.

Patrick, W.H., Jr., and C.N. Reddy. 1978. Chemical changes in rice soils, p. 361–379. *In* International Rice Research Institute (ed.) Soils and rice. International Rice Research Institute, Los Banos, Philippines.

Penton, C.R., A.H. Devol, and J.M. Tiedje. 2006. Molecular evidence for the broad distribution of anaerobic ammonium-oxidizing bacteria in freshwater and marine sediments. Appl. Environ. Microbiol. 72:6829–6832.

Pfennig, N., and H. Biebl. 1976. *Desulfuromonas acetoxidans* gen. nov. and sp. nov., a new anaerobic, sulfur-reducing, acetate-oxidizing bacterium. Arch. Microbiol. 110:3–12.

Ponnamperuma, F.N. 1972. The chemistry of submerged soils. Adv. Agron. 24:29–96.

Pons, J.L., N. van Breemen, and P.M. Driessen. 1982. Physiography of coastal sediments and development of potential soil acidity, p. 1–18. *In* J.A. Kittrick, D.S. Fanning, and L.R. Hossner (eds.) Acid sulfate weathering. SSSA, Madison, WI.

Rabus, R., T.A. Hansen, and F. Widdel. 2006. Dissimilatory sulfate- and sulfur-reducing prokaryotes, p. 659–768. *In* M. Dwarkin, S. Falkow, E. Rosenberg, K.H. Schleifer, and E. Stackebrandt (eds.) The prokaryotes. A handbook on the biology of bacteria. 3rd edn. Vol. 2. Springer Science+Business Media, New York.

Reddy, C.A., M.P. Bryant, and M.J. Wolin. 1972. Characteristics of S organism isolated from *Methanobacillus omelianskii*. J. Bacteriol. 109:539–545.

Rothfuss, F., and R. Conrad. 1993. Vertical profiles of CH_4 concentrations, dissolved substrates and processes involved in CH_4 production in a flooded Italian rice field. Biogeochemistry 18:137–152.

Sakai, S., H. Imachi, S. Hanada, A. Ohashi, H. Harada, and Y. Kamagata. 2008. *Methanocella paludicola* gen. nov., sp. nov., a methane-producing archaeon, the first isolate of the lineage "Rice Cluster I," and proposal of the new archaeal order Methanocellales ord. nov. Int. J. Syst. Evol. Microbiol. 58:929–936.

Schink, B. 1997. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol. Mol. Biol. Rev. 61:262–280.

Schütz, H., W. Seiler, and R. Conrad. 1989. Processes involved in formation and emission of methane in rice paddies. Biogeochemistry 7:33–53.

Skiba, U., and K.A. Smith. 2000. The control of nitrous oxide emissions from agricultural and natural soils. Chemosphere 2:379–386.

Smith, M.S., and J.M. Tiedje. 1980. Growth and survival of antibiotic-resistant denitrifier strains in soil. Can. J. Microbiol. 26:854–856.

Stouthamer, A.H. 1988. Dissimilatory reduction of oxidized nitrogen compounds, p. 245–303. *In* A.J.B. Zehnder (ed.) Biology of anaerobic microorganisms. John Wiley & Sons, New York.

Takai, Y. 1980. Microbial dynamics of paddy soils. Fertil. Sci. 3:17–55.

Takai, Y., and T. Kamura. 1966. The mechanism of reduction in waterlogged paddy soil. Folia Microbiol. 11:304–313.

Thamdrup, B., K. Finster, J.W. Hansen, and F. Bak. 1993. Bacterial disproportion of elemental sulfur coupled to chemical reduction of iron or manganese. Appl. Environ. Microbiol. 59:101–108.

Thauer, R.K., K. Jungermann, and K. Decker. 1977. Energy conservation in chemotrophic anaerobic bacteria. Bacteriol. Rev. 41:100–180.

Tiedje, J.M., A.J. Sexstone, D.D. Myrold, and J.A. Robinson. 1982. Denitrification: Ecological niches, competition and survival. Antonie Leeuwenhoek J. Microbiol. 48:569–583.

Watanabe, A., T. Takeda, and M. Kimura. 1999. Evaluation of origins of CH_4 carbon emitted from rice paddies. J. Geophys. Res. 104D:23623–23630.

Whitman, W.B., T.L. Bowen, and D.R. Boone. 1992. The methanogenic bacteria, p. 719–767. *In* A. Balows, H. Dworkin, W. Harder, and K.H. Schleifer (eds.) The prokaryotes. A handbook on the biology of bacteria: Ecophysiology, isolation, identification, applications. 2nd edn. Vol. 1. Springer-Verlag, New York.

Widdel, F. 1988. Microbiology and ecology of sulfate- and sulfur-reducing bacteria, p. 469–585. *In* A.J.B. Zehnder (ed.) Biology of anaerobic microorganisms. John Wiley & Sons, New York.

Widdel, F., and T.A. Hansen. 1992. The dissimilatory sulfate- and sulfur-reducing eubacteria, p. 583–624. *In* A. Balows, H. Dworkin, W. Harder, and K.H. Schleifer (eds.) The prokaryotes. A handbook on the biology of bacteria: Ecophysiology, isolation, identification, applications. 2nd edn. Vol. 1. Springer-Verlag, New York.

Widdel, F., and N. Pfennig. 1992. The genus *Desulfuromonas* and other Gram-negative sulfur-reducing eubacteria, p. 3379–3389. *In* A. Balows, H. Dworkin, W. Harder, and K.H. Schleifer (eds.) The prokaryotes. A handbook on the biology of bacteria: Ecophysiology, isolation, identification, applications. 2nd edn. Vol. 4. Springer-Verlag, New York.

Yamanaka, K., and S. Motomura. 1959. Studies on the gley formation of soils. I. On the mechanism of the formation of active ferrous iron in soils. Soil Plant Food 5:134–140.

Yoshida, K., and T. Kamura. 1975. The model experiments on the role of ferrous iron in manganese reduction (Part 7). The reduction mechanism of manganese in paddy soils. J. Sci. Soil Manure Japan. 46:382–388.

Zhang, Y., X.H. Ruan, H.J. Op den Camp, T.J. Smits, M.S.M. Jetten, and M.C. Schmid. 2007. Diversity and abundance of aerobic and anaerobic ammonium-oxidizing bacteria in freshwater sediments of the Xinyi River (China). Environ. Microbiol. 9:2375–2382.

26.5 Soil Enzymes

Loretta Landi

Giancarlo Renella

Paolo Nannipieri

26.5.1 Introduction

Considerable information on soil enzymes has accumulated since 1899 when Woods postuled that many transformations of organic compounds in soil might be catalyzed by extracellular enzymes originating from soil organisms. Development in soil enzymology occurred after 1950 (Skujins, 1978a), when research concerned the study of origin, production, stabilization, and persistence of soil enzymes and the role of enzyme activities in the soil–plant system. Several review chapters and some books dealing with soil enzyme have been published (Table 26.3). The knowledge of the origin of enzymes, contributing to the measured soil enzyme activity, can allow a better understanding of organisms actively involved in nutrient dynamics in soil with possible insights into the pathways by which energy and nutrients flow through the soil food web.

It is well established that measured enzyme activity is due to activities from enzymes catalyzing the same reaction not only originating from different sources but having different locations in soil (Burns, 1982; Nannipieri, 1994; Gianfreda and Ruggiero, 2006). Activities of intra- or pericellular enzymes and esoenzymes (those attached to the outer cell membrane), and activities of extracellular enzyme stabilized by surface reactive particles or entrapped by humic molecules are thought to be most important. The former group of enzymes is arbitrarily defined as enzymes associated with living and active microbial cells. Activities of free, extracellular enzymes and enzymes of dead cells and cell debris are considered to be short-lived (being easily degraded) and quantitatively not important (Burns, 1982; Nannipieri et al., 2002). It is methodologically difficult to distinguish between extracellular activity of stabilized enzymes from that of enzymes associated with living organisms. Such separation is important because only the activity of enzymes associated with living microbial cells contributes to microbial activity. This methodological problem makes it difficult to understand the meaning of measurements of enzyme activities in soil, which is their role in plant nutrition and soil fertility, in the metabolism of xenobiotics, and their potential applications in the soil–plant system.

26.5.1.1 Kinetic and Classification of Enzymes

Enzymes, the catalysts of biological systems, are proteins able to increase a million fold or more the speed of a chemical reaction with respect to the same nonenzyme-catalyzed reaction; the enzyme-catalyzed reaction has a lower activation energy and therefore has a faster reaction rate (Tabatabai and Singh, 1976; Dick and Tabatabai, 1978). Active enzymes can require "cofactors." These are small molecules (sometimes inorganic ions),

TABLE 26.3   Book Chapters and Books on Soil Enzyme Activities

	References
Review chapters	
Activity, stability, and kinetic properties of enzymes Immobilized on clay minerals and organomineral complexes	Gianfreda and Violante (1995)
Biological significance of enzymes accumulated in soil	Kiss et al. (1975)
Ecological significance of the biological activity in soil	Nannipieri et al. (1990)
Enzyme activities and microbiological and biochemical processes in soil	Nannipieri et al. (2002)
Enzyme activities in soil	Gianfreda and Ruggiero (2006)
Enzymes in soil: Properties, behavior, and potential applications	Gianfreda et al. (2002)
Enzymes in soil: Research and developments in measuring activities	Tabatabai and Dick (2002)
Extracellular enzymes in soil	Skujins (1976)
Extracellular enzyme–substrate interactions in soil	Burns (1983)
Extraction of enzymes from soils	Tabatabai and Fu (1992)
Humus and enzyme activity	Nannipieri et al. (1996b)
Indirect approaches for assessing intracellular arylsulfatase activity in soil	Fornasier (2002)
Influence of natural and anthropogenic factors on enzyme activity in soil	Gianfreda and Bollag (1996)
Kinetics of enzyme reactions in soil environments	Nannipieri and Gianfreda (1998)
Long-term effects of agricultural systems on soil biochemical and microbial parameters. A review	Dick (1992)
Persistence of enzymes in soil	Skujins (1973)
Pesticide effects on enzyme activities in the soil ecosystems	Schäffer (1993)
Potential uses of soil enzymes	Dick and Tabatabai (1992)
Role of stabilized enzymes in microbial ecology and enzyme extraction from soil with potential applications in soil proteomics	Nannipieri (2006)
Soil enzyme activities and biodiversity measurements as integrative microbiological indicators	Dick et al. (1996a)
Soil enzyme activities as integrative indicators of soil health	Dick (1997)
Soil enzymes	Ladd (1985), Tabatabai (1994)
Soil enzymes activities as indicators of soil quality	Dick (1994)
Soil enzymes: Linking proteomics and ecological processes	Allison et al. (2007)
Standardized methods, sampling, and sample pretreatment	Dick et al. (1996b)
The potential use of soil enzymes as indicators of productivity, sustainability, and pollution	Nannipieri (1994)
The role of carbohydrates in the free enzymes in soil	Mayaudon (1986)
Books	
Enzymes in the Environment. Activity, Ecology, and Applications	Burns and Dick (2002)
Enzymology of Disturbed Soils	Kiss et al. (1998)
Soil Enzymes	Burns (1978)

which are needed to complete the catalytical active structure of the enzyme. For example, Ni is essential for the activity of urease (Dalton et al., 1985). The enzyme without the cofactor is an apoenzyme, and the apoenzyme–cofactor complex is a holoenzyme (Lehninger et al., 1993). Cofactor-requiring enzymes are only active in the intracellular environment because it is almost impossible that the substrate, the cofactor, and the enzyme are present in the same soil microsite to carry out the catalyzed reaction in the extracellular soil environment (Skujins, 1978b).

Like all catalysts, enzymes are not altered at the end of the reaction, and generally, they catalyze specific reactions being substrate specific. For example, proteolytic enzymes hydrolyze peptide bonds in proteins containing L-amino acids but not D-amino acids, and maltase hydrolyzes maltose (α-glucoside) to glucose, whereas cellobiase hydrolyzes cellobiose (β-glucoside) to glucose (Eivazi and Tabatabai, 1988; Lehninger et al., 1993). Enzyme activity depends on pH, ionic strength, temperature,

and the presence or absence of inhibitors or activators (Burns, 1978; Tabatabai, 1982). Enzymes are denatured by elevated temperature and extremes of pH.

The kinetic of enzymes in solution can be described by the Michaelis–Menten equation as follows:

$$V = V_{max}[S]/K_m + [S], \tag{26.2}$$

where
   V is the rate of reaction
   V_{max} is the maximum rate at saturating concentration of the substrate
   S is the substrate
   K_m is the Michaelis–Menten constant (the concentration of S at which $V = V_{max}/2$) and indicates the affinity of the enzyme for the substrate (Lehninger et al., 1993)

It is possible to measure both K_m and V_{max} of most soil enzyme activities measured by reliable protocols (Nannipieri and Gianfreda, 1998). The relative kinetic activity is affected by surface-reactive particles surrounding the enzyme. The effect of the soil colloids, where the enzyme is adsorbed, on the kinetics of the enzyme is discussed by Gianfreda et al. (see Chapter 5 of *Handbook of Soil Sciences: Resource Management and Environmental Impacts*) as well as the resistance of these enzyme complexes to thermal and pH denaturation. Enzymes are classified by the International Enzyme Commission, according to a classification scheme of 1964, and revised in 1972. A name and an enzyme classification (EC) number are given by the committee to identify each enzyme, which is classified according to the catalyzed reaction. The EC number consists of four digits corresponding to four levels of classification. The first digit in the EC number represents the first level of classification, a class indicating the type of catalyzed reaction. These are the main classes of the EC:

- EC 1—oxidoreductases (enzymes that catalyze oxido-reduction reactions)
- EC 2—transferases (enzymes that catalyze the transfer of a chemical group)
- EC 3—hydrolases (enzymes that catalyze the hydrolysis of various bonds)
- EC 4—lyases (enzymes that cleave bonds other than by hydrolysis or oxidation)
- EC 5—isomerases (enzymes that catalyze changes within one molecule)
- EC 6—ligases (enzymes that catalyze the formation of bonds by the cleavage of ATP)

The EC scheme can be found at the JCBN Web site: http://www.chem.qmw.ac.uk/iubmb/enzyme/. Each class includes specific subclasses, which include sub-subclasses. Specific enzyme names generally derive from the substrate of the reaction, followed by the suffix -ase. For example, urease catalyzes the hydrolysis of urea to carbon dioxide and ammonium; maltase hydrolyzes maltose to glucose. In soil biochemistry, enzymes are generally grouped as amidases, decarboxylases, dehydrogenases, lipases, phosphatases, to indicate the type of catalyzed reaction, which usually plays an important role in soil functioning (Burns, 1978; Ladd, 1978). The measured enzyme activity depends on several enzyme catalyzing the same reaction and having different origin and location in the soil matrix.

26.5.2 Origin in Soil

Enzymes in soil may be synthetized by microbial, plant and animal cells but microbial cells are considered to be the most important sources (Ladd, 1985; Torsvik et al., 1996; Nannipieri et al., 2003). Enzyme proteins are the main components of microbial cytoplasm. The distinction between activity of enzymes associated with viable microbial cells and activity of enzymes stabilized in soil is important when enzyme activities are measured as indicators of soil quality or microbial activity since microbial activities are among the most sensitive soil properties to

changes in management practices and environmental conditions (Nannipieri et al., 1996b, 2002).

Soil microorganisms need to release extracellular enzymes to degrade or hydrolyze high-molecular-weight organic compounds into small molecules that can be taken up by microbial cells. Therefore, extracellular enzymes catalyze the initial rate-limiting step of decomposition and nutrient mineralization (Asmar et al., 1994; Sinsabaugh and Moorhead, 1994), and this has been incorporated into mathematical models describing degradation of soil organic matter by Schimel and Weintraub (2003).

It is not possible to include the vast bibliography on the origin of enzymes in soil. We shall only discuss the origin of proteases and phosphatases in soil. Soil proteases, which catalyze the first reaction in protein N mineralization (Ladd and Butler, 1972), are derived mainly from heterotrophic soil bacteria (Nannipieri et al., 1983; Asmar et al., 1992). *Pseudomonas fluorescens*, *Bacillus cereus*, *B. mycoides*, and *Flavobacterium–Cytophaga* spp. can release metalloproteases in a broad range of soil (Hayano et al., 1987; Bach and Munch, 2000). Hayano et al. (1987) reported that metalloproteinases were the dominating proteases in extracts from an andosoil collected from a tomato field, and Janssen (2006) identified *Pseudomonas* as the dominant soil bacteria releasing proteases. Geisseler and Horwath (2008) suggested that microorganisms regulate protease synthesis depending on their needs for C and N sources. Saha et al. (2008) found that high protease activity in manure-treated soil was mainly due to an increased microbial growth, and Kandeler et al. (1999) observed that biomass-specific protease activities of different soils were not related to the organic matter content or texture.

It is hypothesized that in soil, acid phosphomonoesterases are produced mainly by plant roots and fungi whereas alkaline phosphomonoesterases are produced by bacteria (Tabatabai, 1994).

Changes in enzyme activities have been compared with changes in bacterial diversity (Renella et al., 2004). During microbial growth stimulated by adding glucose or plant residue to soil, microbial diversity has been determined by PCR denaturing gradient gel electrophoresis (DGGE) after DNA extraction from soil with the aim to evaluate the microbial origin of phosphatases (Renella et al., 2006b). A better understanding of the origin of soil enzymes has been obtained when specific primers have been developed to amplify "functional genes" involved in enzyme synthesis. In this context, the pivotal contribution is that by Metcalfe et al. (2002), who monitored chitinase activity, bacterial diversity, and genes involved in chitinase production of an untreated (control), sludge- or lime- and sludge-treated pasture soil. The treatments increased chitinase activity and actinobacterium-like chitinase sequences but decreased diversity of chitinases. The approach of combining measurements of enzyme assays and genes involved in the enzymes synthesis (Nannipieri et al., 2008) has been used for assessing the origin of proteases. Sakurai et al. (2007) after PCR amplification of DNA by primers specific for the alkaline metalloprotease (*apr-*) and neutral metalloprotease (*npr-*) genes determined benzyloxycarbonyl phenyllanyl leucine (ZFL) hydrolyzing activity of both rhizosphere and bulk soil fertilized with organic and inorganic

fertilizer and bacterial communities. They found that the composition of proteolytic bacterial communities affected the protease activity of soil. Both soil type and season can influence microbial communities synthetizing metalloproteases (npr) (Mrkonjic Fuka et al., 2009). Casein hydrolyzing activity of an agricultural soil was correlated with the abundance of *sub* genes encoding for subtilisin, the second largest family of serine protease (Mrkonjic Fuka et al., 2008a). The abundance of *sub*-gene containing bacteria decreased with soil depth (Mrkonjic Fuka et al., 2008b).

26.5.3 Enzymes in the Rhizosphere

Plants can positively affect enzymatic activity in soil. Roots can release extracellular enzymes into soil or stimulate microbial activity of rhizosphere soil by rhizodepositions; in the latter case, a new synthesis of microbial enzymes can occur (Ladd, 1985; Tarafdar and Jungk, 1987; Whipps, 1990; Dick, 1994; Nannipieri, 1994; Pinton et al., 2001). As a consequence of the plant effect, enzyme activity is generally higher in the rhizosphere than in the bulk soil. Hydrolase activities of rhizosphere are supposed to play an important role in plant nutrition (Nannipieri et al., 2003). For example, the hydrolysis of organic phosphate monoesters by phosphomonoesterases can account for 30%–80% of P taken up by plants in agricultural soils (Tarafdar and Jungk, 1987; Gilbert et al., 1999). Acid phosphatase was secreted in response to P deficiency by epidermal cells of the main tip roots of white lupin and in the cell walls and intercellular spaces of lateral roots (Gilbert et al., 1999).

Most extracellular enzymes have a low mobility in soil due to their molecular size and charge characteristics, and, thus, any secreted enzyme must operate close to the point of secretion but the substrate may be able to diffuse toward the enzyme (Burns, 1982). The mineralization of the different root exudates and their stimulatory effects on microbial activity is difficult to study because of the broad variety of compounds of rhizodeposition. The study of the effects of a single low-molecular-weight organic molecule requires simple systems mimicking the rhizosphere environment (Badalucco and Kuikman, 2001; Falchini et al., 2003; Landi et al., 2006). These systems create a concentration gradient with higher concentration of the exudate near the root surface. Soil at known distances from the root surface was sampled (Badalucco and Kuikman, 2001). The use of this system has confirmed that root exudates are the main factors controlling microbial activity and community structure in the soil rhizosphere (Landi et al., 2006; Renella et al., 2006a). The rapid increase of microbial growth and hydrolase activities observed in rhizosphere soil by Renella et al. (2006a) suggests that most of the measured enzyme activity is due to intracellular and pericellular enzymes and enzymes attached to the outer surface of viable microbial cells. Different root exudates can have different stimulatory effects on microbial growth and on hydrolase activities (Renella et al., 2006a).

Enzyme activities can also be sensitive to perturbations caused by microbial inoculation (Naseby and Lynch, 1997). Increases in urease, β-glucosidase, and protease activities were related to the increase of the rhizosphere microbial population

as a consequence of bacterial inoculation. Increase in acid phosphatase activity was probably due to a direct fungal secretion or an induced secretion by the plant roots (Joner et al., 2000). The highest concentrations of available P in the rhizosphere of two shrub species inoculated with *Glomus claroideum* were due to the hydrolysis of organic P compounds, probably catalyzed by extracellular, fungal, phosphatase activities.

26.5.4 Production and Persistence

Microorganisms can control synthesis of enzymes at different levels, for example, regulation of gene transcription, posttranscriptional modification of mRNA, and protein cytoplasmic processing after protein synthesis. Allison and Vitousek (2005) suggested a regulation at the level of transcription in soil microorganisms (e.g., enzyme induction and enzyme repression). Both synthesis and persistence of enzyme activities have been quantified in studies based on the stimulation of microbial growth by adding easily degradable organic compounds, such as glucose, or plant material to soil (Ladd and Paul, 1973; Nannipieri et al., 1978, 1979, 1983; Dilly and Nannipieri, 2001; Allison and Vitousek, 2005; Renella et al., 2006b). Enzyme production peaks and persistence depend on the amount and quality of substrate (Balasubramanian et al., 1972; Zantua and Bremner, 1976; McCarty et al., 1992; Allison and Vitousek, 2005). The acid phosphomonoesterase and protease activities were significantly higher in forest than in other soils, probably because of the quality of the soil organic matter (Lahdesmaki and Piispanen, 1988) or to the greater microbial growth (Dilly and Nannipieri, 2001). These studies have shown that enzyme production was not precisely synchronized with microbial growth, as peaks of phosphatase, urease, and protease production proceeded or followed the peak in microbial biomass (Ladd and Paul, 1973; Nannipieri et al., 1978; Renella et al., 2007).

Enzyme production can be an inducible response to the presence of complex substrates (Shackle et al., 2000). However, many enzymes are also produced constitutively by microbes, which may allow them to detect complex resources in the environment (Klonowska et al., 2002). These constitutive enzymes generate low concentration of microbially available products that induce additional enzyme synthesis when complex substrates are abundant. Once concentrations of products increase enough, enzyme synthesis becomes suppressed and production returns to constitutive levels (Chróst, 1991). McCarty et al. (1992) observed that the synthesis of microbial urease was not only dependent on the availability of $N – NH_4^+$ but also on the N metabolites produced by the immobilization of $N – NH_4^+$ taken up by microbial cells.

A simple arithmetical expression has been proposed by Renella et al. (2007) to quantify production (Pr) and persistence (Pe) of enzymes in soil when microbial growth was stimulated by adding glucose and a N source to soil. Greater Pr values of alkaline phosphomonoesterase and phosphodiesterase activities occurred in alkaline than in other soils; it was suggested that the phosphomonoesterase activity with a pH optimum close to soil pH is produced in higher quantities and persists longer than that with a pH optimum further away from that of the soil (Renella

et al., 2007). Significant correlations were observed between the Pr and Pe values of some hydrolases and some soil properties (e.g., alkaline phosphomonoesterase Pr with clay content, acid phosphomonoesterase Pe with soil pH) (Renella et al., 2007), while there was no relationship between phosphomonoesterase production and persistence and bacterial community structure (Renella et al., 2006a).

26.5.5 Role of Soil Constituents in Enzyme Stabilization

It is generally assumed that soil factors such as texture, organic matter content, and pH value are important in stabilizing the released enzymes and that association with clays and/or humic substances can protect the extracellular enzymes from microbial decomposition (Burns, 1982; Nannipieri, 1994). Sorption by soil colloids may protect extracellular enzymes from microbial degradation or chemical hydrolysis, and the adsorbed enzyme can retain its activity for long periods (Nannipieri, 1994). The bibliography on state, activity, and resistance of enzymes adsorbed by surface-reactive particles is extensive and is discussed by Gianfreda et al. (see Chapter 5 of *Handbook of Soil Sciences: Resource Management and Environmental Impacts*). The extraction from soil, partial purification, and characterization of humus–enzyme complexes has shown that the humus network surrounding the entrapped enzyme can allow the passages of substrates and products but not that of proteases (Burns, 1982; Nannipieri et al., 1996b). This model is not valid for enzymes acting on high-molecular-weight substrates (Ladd and Butler, 1972) because the immobilized enzymes should also be susceptible to the attack of protease.

26.5.6 Assays of Soil Enzyme Activity

Enzyme assays in soil are based on the determination of the residual substrate or on the production of the reaction product. Some enzyme assays involve adding, to a soil slurry, either a synthetic substrate linked to a fluorescent molecule (fluorophore) or a substrate forming a colored compound (chromophore) with measurement of the increase in fluorescence and optical absorbance over a fixed incubation time (Saiya-Cork et al., 2002). The high sensitivity of fluorimetric enzyme assays allows detecting enzyme activities of small samples (e.g., rhizosphere samples); an automated assay using a 96-well microplate reader has been developed and can process a high number of different samples (Wirth and Wolf, 1992; Marx et al., 2001; Saiya-Cork et al., 2002; Niemi and Vepsäläinen, 2005).

Tabatabai et al. (1995) have set up several enzymes assays. The protocol is generally based on choosing the optimal conditions (pH, temperature, and substrate concentrations) for the enzyme activity and on using several soils with a broad spectrum of properties. Due to the complexity of the soil system, several analytical problems can interfere on the determination of an enzyme activity such as the fixation of NH_4^+ by clays in enzyme assays based on the determination of NH_4^+ released by enzyme activities such as urease activities (Tabatabai, 1994). The present soil enzyme assays have several limitations: they measure potential rather than actual enzyme activity because soil slurries are generally incubated in the presence of a saturate substrate concentration at optimal temperature values, and with a buffer at optimal pH under shaking conditions. These conditions do not occur in situ (Burns, 1982; Nannipieri et al., 2002; Gianfreda and Ruggiero, 2006). Several methods have been proposed to distinguish these two enzyme activities such as the use of plasmolytic agents or gamma-irradiation (Cawse, 1975; Frankenberger and Johanson, 1986). By assuming that the use of the short-term (up to a few hours) enzyme assays only determine the activity of extracellular stabilized enzymes, Klose and Tabatabai (1999a, 1999b) proposed the soil chloroform fumigation to determine intracellular plus extracellular enzyme activities. No one has proved that the present short-term enzyme assays determine the extracellular stabilized enzyme activity. In addition, proteases are active during fumigation and degrade intracellular enzymes, thus causing an underestimation of intracellular enzyme activities (Nannipieri et al., 2002; Renella et al., 2002). The rapid increase in enzyme activities when microbial growth has been stimulated by adding energy and nutrient sources to soil (see above) has been taken as evidence that most of the measured enzyme activity is due to activities of intracellular and pericellular enzymes and esoenzymes of viable microbial cells (Renella et al., 2006a).

Margon and Fornasier (2008) have investigated location and related kinetics of soil enzymes using a rapid fumigation procedure avoiding proteolysis of intracellular enzymes while ensuring an efficient cell lysis.

A different approach was followed by McLaren and Pukite (1973) and Nannipieri et al. (1996a) to measure the activity of extracellular stabilized enzymes. These authors correlated the number of ureolytic microorganisms to urease activity, and the ATP content to phosphomonoesterase activity, respectively. By extrapolating to zero ureolytic number or the ATP content, the positive intercept on the *y*-axis was assumed to give extracellular urease or phosphatase activity. Microbial numbers, the ATP contents, and both enzyme activities were measured throughout when microbial growth was stimulated by adding glucose and a N source to soil. Advantages and limitations of this approach, called the "Physiological Approach" have been discussed by Nannipieri et al. (2002). One of the limits of this method is that synthesis of enzymes can be repressed during microbial growth as it occurs for phosphatase when inorganic P is added with glucose and the N source to soil (Nannipieri et al., 1978).

26.5.7 Enzyme Activities as Indicators of Soil Fertility and Soil Pollution

Oxidoreductases and hydrolases, the most studied enzyme activities of soil, are involved in the oxidation and release of inorganic nutrients from organic matter (Dick and Tabatabai, 1993). Hydrolases are said to play an important role in plant nutrition (Nannipieri et al., 2002), and for this reason, for example, soil sulfatase activity has been taken as an index of S nutrition in soil (Speir, 1984). There is a growing interest in using soil enzymes as indicators of soil fertility since enzymes are sensitive to numerous

factors such as type of amendment, crop rotations, climate, tillage, and edaphic properties (Skujins, 1978b; Naseby and Lynch, 1997; Ndiaye et al., 2001). For example, positive effects of organic amendments have been observed on the soil hydrolase activities involved in N, P, and C cycling of grassland and degraded soils (Kandeler and Eder, 1993; Garcia and Hernández, 2003).

Organic and inorganic pollutants may reduce soil functionality, by inhibiting hydrolase activities (Tyler et al., 1989). Organic pollutants such as pesticides may affect a sensitive microbial population in soil, but may not change or may increase an enzyme activity as a result of the stimulation of the resistant microorganisms. Repeated addition of a single pesticide, or chemically similar pesticides, generally causes selection of resistant microbial population, and the effect on the enzyme activity can be nil or less pronounced that in the case of single application (Nannipieri, 1994).

Various enzymes can catalyze reactions degrading organic pollutants and reducing contamination of soil. The arylesterase enzyme is involved in the degradation of plastics and hydrolysis of toxic metabolites (e.g., paraoxon) and of organophosphates (Emmelot et al., 1964; Wilde and Kekwick, 1964; Primo-Parmo et al., 1996). Therefore, the measurement of enzyme activity in soil may be important for evaluating the response of soil microbial communities to organic contamination and remediation measures, assessing the fate of pesticides and other organic contaminants potentially toxic to microorganisms and soil organisms (Singh and Jain, 2003; Moon and Smith, 2005). The inhibition of some enzyme activities can occur in metal-polluted soils of both long-term field and laboratory experiments (Tyler, 1974; Tabatabai, 1977; Frankenberger and Tabatabai, 1981; Doelman and Haanstra, 1986, 1989; Tyler et al., 1989; Marzadori et al., 1996; Landi et al., 2000; Moreno et al., 2001; Renella et al., 2003, 2004, 2005; Zornoza et al., 2009).

Unlike organic pollutants, inorganic pollutants cannot be degraded and used as a C source, thus stimulating microbial growth. However, resistant microorganisms can grow by using cell debris of killed, sensitive microorganisms. Thus, changes in enzyme activities of soils polluted with inorganic pollutant are also difficult to be interpreted due to direct and indirect effects of the contaminant. Some investigators have employed the ED_{50} value (i.e., the heavy metal concentration at which enzyme activity is half of the uninhibited level) to describe metal inhibition of soil enzymes (Babich et al., 1983; Haanstra and Doelman, 1991; Moreno et al., 2001; Renella et al., 2003). Using this measurement, Doelman and Haanstra (1989) demonstrated that the influence of heavy metals on acid phosphatase activity in soil depended on differences in texture and organic matter content of soil. The negative impact of trace elements on acid phosphatase was most pronounced in sandy soil. The degree of inhibition also depends on the concentration and form of the metal, the contact time of soil with the contaminant, and the assayed enzyme (Tyler, 1981; Renella et al., 2003, 2004). For example, Moreno et al. (1999) reported that high Cd concentration (>800 mg Cd kg[-1]) inhibited dehydrogenase activity, but not β-glucosidase, urease, and phosphomonoesterase activities. Karaca et al. (2002) observed significant reductions of several hydrolase activities by Cd-enriched (50 mg Cd kg[-1]) sewage

sludge, but urease activity was not inhibited. Renella et al. (2005) reported that long-term exposure of soil microflora to high Cd concentrations (40 mg Cd kg[-1]) reduced hydrolase synthesis and metabolic efficiency. Phytostabilization of trace elements may have beneficial effects on soil functionality due to reduction of trace element mobility in contaminated environmental matrices (Izquierdo et al., 2005; Perez de Mora et al., 2005; Kumpiene et al., 2006, 2007, 2009; Mench et al., 2006; Renella et al., 2008).

Soil contamination can reduce enzyme activity (Dick and Tabatabai, 1992; Nannipieri, 1994; Gianfreda and Bollag, 1996) whereas, an efficient phytostabilization of trace elements can increase enzyme activity and be indicative of the sustainable management of the remediated polluted soils (Kumpiene et al., 2009).

Since soil enzymes differ in origin, function, and location (Burns, 1982) and respond to different key environmental signals, it would be useful to condense the information provided by measuring enzyme activities into a single numerical value. Therefore, complex indices calculated by algebraic combinations of different soil biochemical properties (Nannipieri et al., 2002; Gil-Sotres et al., 2005) or multivariate analysis (Blair et al., 1995) such as principal component analysis (PCA) (Sena et al., 2002) and factorial analysis (Melero et al., 2006; Shukla et al., 2006) have been developed. The validity of these indices or these analyses to different soils and environmental conditions is yet to be proven.

26.5.8 Conclusions

Future research is needed to set up enzyme assays distinguishing enzyme activities associated with living cells and sensitive to environmental change and extracellular enzyme activity stabilized by soil colloids. Enzyme activity of soil should be combined with measurements of genes codifying the enzymes and with measurement of expression of these genes at transcriptional (synthesis of mRNA) and translational (synthesis of specific protein) level so as to evaluate the origin of the measured enzyme activities. Finally, bibliography before the electronic searches should be carefully considered since it is extensive as testified by many review chapters and whole books dedicated to the subject (Table 26.3) so as to avoid "to reinvent the wheel."

References

Allison, S.D., T.B. Gartner, K. Holland, M. Weintraub, and R.L. Sinsabaugh. 2007. Soil enzymes: Linking proteomics and ecological process, p. 704–711. *In* C.J. Hurst, R.L. Crawford, J.L. Garland, D.A. Lipson, A.L. Mills, and L.D. Stetzenbach (eds.) Manual of environmental microbiology. 3rd edn. ASM Press, New York.

Allison, S.D., and P.M. Vitousek. 2005. Responses of extracellular enzymes to simple and complex nutrient inputs. Soil Biol. Biochem. 37:937–944.

Asmar, F., F. Eiland, and N.E. Nielsen. 1992. Interrelationship between extracellular enzyme activity, ATP content, total counts of bacteria and CO_2 evolution. Biol. Fertil. Soils 14:288–292.

Asmar, F., F. Eiland, and N.E. Nielsen. 1994. Effect of extracellular enzyme activities on solubilization rate of soil organic nitrogen. Biol. Fertil. Soils 17:32–38.

Babich, H., R.J.E. Bewly, and G. Stotzky. 1983. Application of "ecological dose" concept to the impact of heavy metals in some microbe-mediated ecological processes in soil. Arch. Environ. Contam. 12:421–426.

Bach, H.J., and J.C. Munch. 2000. Identification of bacterial sources of soil peptidases. Biol. Fertil. Soils 31:219–224.

Badalucco, L., and P.J. Kuikman. 2001. Mineralization and immobilization in the rhizosphere, p. 141–196. *In* R. Pinton, Z. Varanini, and P. Nannipieri (eds.) The rhizosphere. Biochemistry and organic substances at the soil–plant interface. Marcel Dekker, New York.

Balasubramanian, A., R. Siddaramappa, and G. Rangaswami. 1972. Effect of organic manuring on the activities of the enzymes hydrolysing sucrose and urea and on soil aggregation. Plant Soil 37:319–328.

Blair, G.J., R.D.B. Lefroy, and L. Lisle. 1995. Soil carbon fractions based on their degree of oxidation, and the development of a carbon management index for agricultural systems. Aust. J. Agric. Res. 46:1459–1466.

Burns, R.G. 1978. Enzyme activity in soil: Some theoretical and practical considerations, p. 295–340. *In* R.G. Burns (ed.) Soil enzymes. Academic Press, London, U.K.

Burns, R.G. 1982. Enzyme activity in soil: Location and a possible role in microbial ecology. Soil Biol. Biochem. 14:423–427.

Burns, R.G. 1983. Extracellular enzyme-substrate interactions in soil, p. 249–298. *In* J.H. Slater, R. Whittenburry, and J.W.T. Wimpemy (eds.) Microbes in their natural environments. University Press, Cambridge, U.K.

Burns, R.G., and R.P. Dick. 2002. Enzymes in the environment. Activity, ecology and applications. Marcel Dekker, New York.

Cawse, P.A. 1975. Microbiology and biochemistry of irradiated soils, p. 213–267. *In* E.A. Paul and A.D. McLaren (eds.) Soil biochemistry. Vol. 3. Marcel Dekker, New York.

Chróst, R.J. 1991. Environmental control of the synthesis and activity of aquatic microbial ecto enzymes, p. 29–59. *In* R.J. Chróst (ed.) Microbial enzymes in aquatic environments. Springer-Verlag, New York.

Dalton, D.A., H.J. Evans, and F.J. Hanus. 1985. Stimulation by nickel of soil microbial urease activity and urease and hydrogenase activities in soybeans grown in a low-nickel soil. Plant Soil 88:245–258.

Dick, R.P. 1992. Long-term effects of agricultural systems on soil biochemical and microbial parameters. A review. Agric. Ecosyst. Environ. 40:25–36.

Dick, R.P. 1994. Soil enzyme activity as indicators of soil quality, p. 107–124. *In* J.W. Doran, D.C. Coleman, D.F. Bezdicek, and B.A. Stewart (eds.) Defining soil quality for a sustainable environment. SSSA Special Publication No. 35. SSSA, Madison, WI.

Dick, R.P. 1997. Soil enzyme activities as integrative indicators of soil health, p. 121–156. *In* C.E. Pankhurst, B.M. Doube, and V.V.S.R. Gupta (eds.) Biological indicators of soil health. CABI, Wallingford, U.K.

Dick, R.P., D.P. Breakwell, and R.F. Turco. 1996a. Soil enzyme activities and biodiversity measurements as integrative microbiological indicators, p. 247–272. *In* J.W. Doran and A.J. Jones (eds.) Methods for assessing soil quality. SSSA Special Publication No. 49. SSSA, Madison, WI.

Dick, W.A., and M.A. Tabatabai. 1978. Inorganic pyrophosphatase activity of soils. Soil Biol. Biochem. 10:59–65.

Dick, W.A., and M.A. Tabatabai. 1992. Potential uses of soil enzymes, p. 95–127. *In* F.B. Metting, Jr. (ed.) Soil microbial ecology: Applications in agricultural and environmental management. Marcel Dekker, New York.

Dick, W.A., and M.A. Tabatabai. 1993. Significance and potential uses of soil enzymes, p. 95–127. *In* F. Blaine (ed.) Soil microbial ecology. Marcel Dekker, New York.

Dick, R.P., D.R. Thomas, and J.J. Halvorson. 1996b. Standardized methods, sampling and sample pretreatment, p. 107–122. *In* J.W. Doran and A.J. Jones (eds.) Methods for assessing soil quality. Special Publication No. 49. SSSA, Madison, WI.

Dilly, O., and P. Nannipieri. 2001. Response of ATP content, respiration rate and enzyme activities in an arable and a forest soil to nutrient additions. Biol. Fertil. Soils 34:64–72.

Doelman, P., and L. Haanstra. 1986. Short- and long-term effects of heavy metals on urease activity in soils. Biol. Fertil. Soils 2:213–218.

Doelman, P., and L. Haanstra. 1989. Short- and long-term effects of heavy metals on phosphatase activity: An ecological dose-response model approach. Biol. Fertil. Soils 8:235–241.

Eivazi, F., and M.A. Tabatabai. 1988. Glucosidases and galactosidases in soils. Soil Biol. Biochem. 20:601–606.

Emmelot, P., C.J. Bos, E.L. Benedetti, and P.H. Rũmke. 1964. Studies on plasma membranes. I. Chemical composition and enzyme content of plasma membranes isolated from rat liver. Biochim. Biophys. Acta. 90:126–145.

Falchini, L., N. Naumova, P.J. Kuikman, J. Bloem, and P. Nannipieri. 2003. CO_2 evolution and denaturing gradient gel electrophoresis profiles of bacterial communities in soil following addition of low molecular weight substrates to simulate root exudates. Soil Biol. Biochem. 36:775–782.

Fornasier, F. 2002. Indirect approaches for assessing intracellular arylsulfatase activity in soil, p. 345–352. *In* A. Violante, P.M. Huang, J.M. Bollag, and L. Gianfreda (eds.) Soil mineral–organic matter–microorganism interactions and ecosystem health. Development in soil science, 28B. Elsevier, London, U.K.

Frankenberger, W.T., Jr., and J.B. Johanson. 1986. Use of plasmolytic agents and antiseptics in soil enzyme assays. Soil Biol. Biochem. 18:209–213.

Frankenberger, W.T., Jr., and M.A. Tabatabai. 1981. Amidase activity in soils. III. Stability and distribution. Soil Sci. Soc. Am. J. 45:333–338.

Garcia, C., and T. Hernández. 2003. Microbial activity in degraded soils under semiarid climate, changes with their rehabilitation, p. 83–97. *In* M.C. Lobo and J.J. Jbáňez (eds.) Preserving soil quality and soil biodiversity. The role of surrogate indicators. IMIA, CSIC, Madrid, Spain.

Geisseler, D., and W.R. Horwath. 2008. Regulation of extracellular protease activity in soil in response to different sources and concentrations of nitrogen and carbon. Soil Biol. Biochem. 40:3040–3048.

Gianfreda, L., and J.M. Bollag. 1996. Influence of natural and anthropogenic factors on enzyme activity, p. 123–193. *In* G. Stotzky and J.M. Bollag (eds.) Soil biochemistry. Vol. 9. Marcel Dekker, New York.

Gianfreda, L., M.A. Rao, F. Saccomandi, F. Sannino, and A. Violante. 2002. Enzymes in soil: Properties, behavior and potential applications, p. 301–328. *In* A. Violante, P.M. Huang, J.-M. Bollag, and L. Gianfreda (eds.) Soil mineral–organic matter–microorganism interactions and ecosystem health. Development in soil science, 28B. Elsevier, London, U.K.

Gianfreda, L., and P. Ruggiero. 2006. Enzyme activities in soil, p. 257–311. *In* P. Nannipieri and K. Smalla (eds.) Nucleic acids and proteins in soil. Vol. 8. Springer, Berlin, Germany.

Gianfreda, L., and A. Violante. 1995. Activity, stability and kinetic properties of enzymes immobilized on clay minerals and organo-mineral complexes, p. 201–209. *In* P.M. Huang, J. Berthelin, J.-M. Bollag, W.B. McGill, and A.L. Page (eds.) Environmental impact of soil component interactions. Metals, other inorganics, and microbial activities. Vol. 2. CRC press, Boca Raton, FL.

Gilbert, G.A., J.D. Knight, C.P. Vance, and D.L. Allan. 1999. Acid phosphatase activity in phosphorus-deficient white lupin roots. Plant Cell Environ. 22:801–810.

Gil-Sotres, F., C. Trasar-Cepeda, M.C. Leiro's, and S. Seoane. 2005. Different approaches to evaluating soil quality using biochemical properties. Soil Biol. Biochem. 37:877–887.

Haanstra, L., and P. Doelman. 1991. An ecological dose-response model approach to short- and long-term effects of heavy metals on arylsulfatase activity of soil. Biol. Fertil. Soils 11:18–23.

Hayano, K., M. Takeuchi, and E. Ichishima. 1987. Characterization of a metalloproteinase component extracted from soil. Biol. Fertil. Soils 4:179–183.

Izquierdo, I., I. Caravaca, M.M. Alguacil, G. Hernandez, and A. Roldan. 2005. Use of microbiological indicators for evaluating success in soil restoration after revegetation of a mining area under subtropical conditions. Appl. Soil Ecol. 30:3–10.

Janssen, P.H. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA genes. Appl. Environ. Microbiol. 72:1719–1728.

Joner, E.J., I.M. Aarle, and M. Vosatka. 2000. Phosphatase activity of extra-radical arbuscular mycorrhizal hyphae: A review. Plant Soil 226:199–210.

Kandeler, E., and G. Eder. 1993. Effects of cattle slurry in grassland on microbial biomass and on activities of various enzymes. Biol. Fertil. Soils 16:249–254.

Kandeler, E., M. Stemmer, and M.H. Gerzabek. 1999. Tillage changes microbial biomass and enzyme activities in particle-size fractions of a hapic chernozem. Soil Biol. Biochem. 31:1253–1264.

Karaca, A., D.C. Naseby, and J.M. Lynch. 2002. Effect of cadmium contamination with sewage sludge and phosphate fertiliser amendments on soil enzyme activities, microbial structure and available cadmium. Biol. Fertil. Soils 35:428–434.

Kiss, S., M. Dragan-Bularda, and D. Radulescu. 1975. Biological significance of enzymes accumulated in soil. Adv. Agron. 27:25–87.

Kiss, S., D. Pasca, and M. Dragan-Bularda. 1998. Enzymology of disturbed soils. Development in soil science 26. Elsevier, Amsterdam, the Netherlands.

Klonowska, A., C. Gaudin, A. Fournel, M. Asso, J. Le Petit, M. Giorgi, and T. Tron. 2002. Characterization of a low redox potential laccase from the basidiomycete C30. Eur. J. Biochem. 269:6119–6125.

Klose, S., and M.A. Tabatabai. 1999a. Arylsulphatase activity of microbial biomass in soils. Soil Sci. Soc. Am. J. 63:569–574.

Klose, S., and M.A. Tabatabai. 1999b. Urease activity of microbial biomass in soils. Soil Biol. Biochem. 31:205–211.

Kumpiene, J., G. Guerri, L. Landi, G. Pietramellara, P. Nannipieri, and G. Renella. 2009. Microbial biomass, respiration and enzyme activities after in situ aided phytostabilization of a Pb- and Cu-contaminated soil. Ecotoxicol. Environ. Saf. 12:115–119.

Kumpiene, J., S. Ore, A. Lagerkvist, and C. Maurice. 2007. Stabilization of Pb- and Cu contaminated soil using coal fly ash and peat. Environ. Pollut. 145:365–373.

Kumpiene, J., S. Ore, G. Renella, M. Mench, A. Lagerkvist, and C. Maurice. 2006. Assessment of zerovalent iron for stabilization of chromium, copper, and arsenic in soil. Environ. Pollut. 144:62–69.

Ladd, J.N. 1978. Origin and range of enzymes in soil, p. 51–96. *In* R.G. Burns (ed.) Soil enzymes. Academic Press, London, U.K.

Ladd, J.N. 1985. Soil enzymes, p. 175–221. *In* D. Vaughan and R.E. Malcom (eds.) Soil organic matter and biological activity. Matinus Nijhoff, Dordrecht, the Netherlands.

Ladd, J.N., and J.H.A. Butler. 1972. Short-term assays of soil proteolytic enzyme activities using proteins and dipeptide derivatives as substrates. Soil Biol. Biochem. 4:19–30.

Ladd, J.N., and E.A. Paul. 1973. Changes in enzyme activity and distribution of acid-soluble amino acid-nitrogen in soil during nitrogen immobilization and mineralization. Soil Biol. Biochem. 5:825–840.

Lahdesmaki, R., and R. Piispanen. 1988. Degradation products and the hydrolytic enzyme activities in the soil humification processes. Soil Biol. Biochem. 20:287–292.

Landi, L., G. Renella, J.L. Moreno, L. Falchini, and P. Nannipieri. 2000. Influence of Cd on the metabolic quotient, L-:D-glutamic acid respiration ratio and enzyme activity:Microbial biomass ratio under laboratory conditions. Biol. Fertil. Soils 32:8–16.

Landi, L., F. Valori, J. Ascher, G. Renella, L. Falchini, and P. Nannipieri. 2006. Root exudate effects on the bacterial communities, CO_2 evolution, nitrogen transformation and ATP content of rhizosphere and bulk soils. Soil Biol. Biochem. 38:509–516.

Lehninger, A.L., D.L. Nelson, and M.M. Cox. 1993. Principles of biochemistry. 2nd edn. Worth Publishers, New York.

Margon, A., and F. Fornasier. 2008. Determining soil enzyme location and related kinetics using rapid fumigation and high-yield extraction. Soil Biol. Biochem. 40:2178–2181.

Marx, M.C., M. Wood, and S.C. Jarvis. 2001. A microplate fluorometric assay for the study of enzyme diversity in soils. Soil Biol. Biochem. 33:1633–1640.

Marzadori, C., C. Ciavatta, D. Montecchio, and C. Gessa. 1996. Effects of lead pollution on different soil enzyme activities. Biol. Fertil. Soils 22:53–58.

Mayaudon, J. 1986. The role of carbohydrates in the free enzymes in soil, p. 263–206. *In* C.H. Fuchsman (ed.) Peat and water. Springer, Berlin, Germany.

McCarty, G.W., D.R. Shogren, and J.M. Bremner. 1992. Regulation of urease production in soil by microbial assimilation of nitrogen. Biol. Fertil. Soils 12:261–264.

McLaren, A.D., and A. Pukite. 1973. Ubiquity of some soil enzymes and isolation of soil organic matter with urease activity, p. 187–193. *In* D. Povoledo and M.L. Goltermaan (eds.) Humic substances and function in the biosphere. Pudoc, Wageningen, the Netherlands.

Melero, S., J.C. Porras, J.F. Herencia, and E. Madejon. 2006. Chemical and biochemical properties in a silty loam soil under conventional and organic management. Soil Tillage Res. 90:162–170.

Mench, M., G. Renella, A. Gelsomino, L. Land, and P. Nannipieri. 2006. Biochemical parameters and bacterial species richness in soils contaminated by sludge-borne metals and remediated with inorganic soil amendments. Environ. Pollut. 144:24–31.

Metcalfe, A.C., M. Krsek, G.W. Gooday, J.I. Prosseer, and E.M. Wellington. 2002. Molecular analysis of a bacterial chitinolytic community in an upland pasture. Appl. Environ. Microbiol. 68:5042–50505.

Moon, A.P., and T.J. Smith. 2005. Organophosphate inhibition of an arylesterase from *Lumbricus terrestris* and reversal by pralidoxime. Soil Biol. Biochem. 37:1211–1213.

Moreno, J.L., C. García, L. Landi, L. Falchini, G. Pietramellara, and P. Nannipieri. 2001. The ecological dose value (ED_{50}) for assessing Cd toxicity on ATP content and dehydrogenase and urease activities of soil. Soil Biol. Biochem. 33:483–489.

Moreno, J.L., T. Hernandez, and C. García. 1999. Effects of cadmium-contaminated sewage sludge compost on dynamics of organic matter and microbial activity in an arid soil. Biol. Fertil. Soils 28:230–237.

Mrkonjic Fuka, M., M. Engel, A. Gattinger, U. Bausenwein, M. Sommer, J.C. Munch, and M. Schloter. 2008a. Factors influencing variability of proteolytic genes and activities in arable soils. Soil Biol. Biochem. 40:1646–1653.

Mrkonjic Fuka, M., M. Engel, F. Haesler, G. Welzl, J.C. Munch, and M. Schloter. 2008b. Diversity of proteolytic community encoding for subtilisin in an arable field: Spatial and temporal variability. Biol. Fertil. Soils 45:185–191.

Mrkonjic Fuka, M., M. Engel, A. Hagn, J.C. Munch, M. Sommer, and M. Schloter. 2009. Changes of diversity pattern of proteolytic bacteria over time and space in an agricultural soil. Microbiol. Ecol. 391–401.

Nannipieri, P. 1994. The potential use of soil enzymes as indicators of productivity, sustainability and pollution, p. 238–244. *In* C.E. Pankhurst, B.M. Double, V.V.S.R. Gupta, and P.R. Grace (eds.) Soil biota: Management in sustainable farming systems. CSIRO, Adelaide, Australia.

Nannipieri, P. 2006. Role of stabilized enzymes in microbial ecology and enzyme extraction from soil with potential applications in soil proteomics, p. 75–94. *In* P. Nannipieri and K. Smalla (eds.) Soil biology. Vol. 8. Nucleic acids and proteins in soil. Springer Verlag, Berlin, Germany.

Nannipieri, P., J. Ascher, M.T. Ceccherini, G. Guerri, G. Renella, and G. Pietramellara. 2008. Recent advances in functional genomics and proteomics of plant associated microbes, p. 215–241. *In* C.S. Nautiyal and P. Dion (eds.) Molecular mechanisms of plant and microbe coexistence. Springer, Heidelberg, Germany.

Nannipieri, P., J. Ascher, M.T. Ceccherini, L. Landi, G. Pietramellara, and G. Renella. 2003. Microbial diversity and soil functions. Eur. J. Soil Sci. 54:655–670.

Nannipieri, P., and L. Gianfreda. 1998. Kinetics of enzyme reactions in soil environments, p. 449–479. *In* P.M. Huang, N. Senesi, and J. Buffle (eds.) Structure and surface reactions of soil particles. Wiley, Chichester, U.K.

Nannipieri, P., S. Grego, and B. Ceccanti. 1990. Ecological significance of the biological activity in soil, p. 293–355. *In* J.-M. Bollag and G. Stozky (eds.) Soil biochemistry. Vol. 6. Marcel Dekker, New York.

Nannipieri, P., R.L. Johnson, and E.A. Paul. 1978. Criteria for measuring microbial growth and activity in soil. Soil Biol. Biochem. 10:223–229.

Nannipieri, P., E. Kandeler, and P. Ruggiero. 2002. Enzyme activities and microbiological and biochemical processes in soil, p. 1–33. *In* R.G. Burns and R. Dick (eds.) Enzymes in the environment. Marcel Dekker, New York.

Nannipieri, P., L. Muccini, and C. Ciardi. 1983. Microbial biomass and enzyme activities: Production and persistence. Soil Biol. Biochem. 15:679–685.

Nannipieri, P., F. Pedrazzini, P.G. Arcara, and C. Piovanelli. 1979. Changes in amino acids, enzyme activities and biomasses during soil microbial growth. Soil Sci. 127:26–34.

Nannipieri, P., I. Sastre, L. Landi, M.C. Lobo, and G. Pietramellara. 1996a. Determination of extracellular neutral phosphomonoesterase activity in soil. Soil Biol. Biochem. 28:107–112.

Nannipieri, P., P. Sequi, and P. Fusi. 1996b. Humus and enzyme activity, p. 293–328. *In* A. Piccolo (ed.) Humic substances in terrestrial ecosystems. Elsevier, Amsterdam, the Netherlands.

Naseby, D.C., and J.M. Lynch. 1997. Rhizosphere soil enzymes as indicators of perturbations caused by enzyme substrate and inoculation of a genetically modified strain of *Pseudomonas fluorescence* on wheat seed. Soil Biol. Biochem. 29:1353–1362.

Ndiaye, E.L., J.M. Sandero, O. McGrath, and R.P. Dick. 2001. Integrative biological indicators for detecting change in soil quality. Am. J. Altern. Agric. 15:26–31.

Niemi, R.M., and M. Vepsäläinen. 2005. Stability of the fluorogenic enzyme substrates and pH optima of enzyme activities in different Finnish soils. J. Microbiol. Methods 60:195–205.

Perez de Mora, A., J.J. Ortega-Calvo, F. Cabrera, and E. Madejon. 2005. Changes in enzyme activities and microbial biomass after "*in situ*" remediation of a heavy metal-contaminated soil. Appl. Soil Ecol. 28:125–137.

Pinton, R., Z. Varanini, and P. Nannipieri. 2001. The rhizosphere biochemistry and organic substances at the soil–plant interface. Marcel Dekker, New York.

Primo-Parmo, S.L., R.C. Sorenson, J. Teiber, and B.N. La Du. 1996. The human serum paraoxonase/arylesterase gene (*PON1*) is one member of a multigene family. Genomics 33:498–507.

Renella, G., D. Egamberdiyeva, L. Landi, M. Mench, and P. Nannipieri. 2006a. Microbial activity and hydrolase activities during decomposition of root exudates released by an artificial root surface in Cd-contaminated soils. Soil Biol. Biochem. 38:702–708.

Renella, G., L. Landi, J. Ascher, M.T. Ceccherini, G. Pietramellara, M. Mench, and P. Nannipieri. 2008. Long-term effects of aided phytostabilization of trace elements on microbial biomass and activity, enzyme activities, and composition of microbial community in the Jales contaminated mine spoils. Environ. Pollut. 152:702–712.

Renella, G., L. Landi, J. Ascher, M.T. Ceccherini, G. Pietramellara, and P. Nannipieri. 2006b. Phosphomonoesterase production and persistence and composition of bacterial communities during plant material decomposition in soils with different pH values. Soil Biol. Biochem. 38:795–802.

Renella, G., L. Landi, and P. Nannipieri. 2002. Hydrolase activities during and after the chloroform fumigation of soil as affected by protease activity. Soil Biol. Biochem. 34:51–60.

Renella, G., M. Mench, L. Landi, and P. Nannipieri. 2005. Microbial activity and hydrolase synthesis in long-term Cd-contaminated soils. Soil Biol. Biochem. 37:133–139.

Renella, G., M. Mench, D. van der Lelie, G. Pietramellara, J. Ascher, M.T. Ceccherini, L. Landi, and P. Nannipieri. 2004. Hydrolase activity, microbial biomass and community structure in long-term Cd-contaminated soils. Soil Biol. Biochem. 36:443–451.

Renella, G., A.L. Reyes Ortigoza, L. Landi, and P. Nannipieri. 2003. Additive effects of copper and zinc on cadmium toxicity to phosphomonoesterase activities and ATP content of soil as estimated by the ecological dose (ED_{50}). Soil Biol. Biochem. 35:1203–1210.

Renella, G., U. Szukies, L. Landi, and P. Nannipieri. 2007. Quantitative assessment of hydrolase production and persistence in soil. Biol. Fertil. Soils 44:321–329.

Saha, S., K.A. Gopinath, B.L. Mina, and H.S. Gupta. 2008. Influence of continuous application of inorganic nutrients to a maize-wheat rotation on soil enzyme activity and grain quality in a rainfed Indian soil. Eur. J. Soil Biol. 44:521–531.

Saiya-Cork, K.R., R.L. Sinsabaugh, and D.R. Zak. 2002. The effects of long term nitrogen deposition on extracellular enzyme activity in an *Acer saccharum* forest soil. Soil Biol. Biochem. 34:1309–1315.

Sakurai, M., K. Suzuki, M. Onodera, T. Shinano, and M. Osaki. 2007. Analysis of bacterial communities in soil by PCR-DGGE targeting protease genes. Soil Biol. Biochem. 39:2777–2784.

Schaffer, A. 1993. Pesticide effects on enzyme activities in the soil ecosystem, p. 273–340. *In* J.-M. Bollag and G. Stozky (eds.) Soil biochemistry. Vol. 8. Marcel Dekker, New York.

Schimel, J.P., and M.N. Weintraub. 2003. The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: A theoretical model. Soil Biol. Biochem. 35:549–563.

Sena, M.M., R.T.S. Frighetto, P.J. Valarini, H. Tokeshi, and R.J. Poppi. 2002. Discrimination of management effects on soil parameters by using principal component analysis: A multivariate analysis case study. Soil Tillage Res. 67:171–181.

Shackle, V.J., C. Freeman, and B. Reynolds. 2000. Carbon supply and the regulation of enzyme activity in constructed wetlands. Soil Biol. Biochem. 32:1935–1940.

Shukla, M.K., R. Lal, and M. Ebinger. 2006. Determining soil quality indicators by factor analysis. Soil Tillage Res. 87:194–204.

Singh, O.V., and R.K. Jain. 2003. Phytoremediation of toxic aromatic pollutants from soil. Appl. Microbiol. Biotechnol. 63:128–135.

Sinsabaugh, R.L., and D.L. Moorhead. 1994. Resource allocation to extracellular enzyme production: A model for nitrogen and phosphorus control of litter decomposition. Soil Biol. Biochem. 26:1305–1311.

Skujins, J.J. 1973. Persistence of enzymes in soil, p. 395–408. *In* E. Ingerson (ed.) Proceedings of symposium on hydrogeochemistry and biogeochemistry. Clarke Co., Washington, DC.

Skujins, J.J. 1976. Extracellular enzymes in soil. CRC Crit. Rev. Microbiol. 4:383–421.

Skujins, J. 1978a. History of abiontic soil enzyme research, p. 1–49. *In* R.G. Burns (ed.) Soil enzymes. Academic Press, London, U.K.

Skujins, J. 1978b. Extracellular enzyme in soil. CRC Crit. Rev. Microbiol. 4:383–421.

Speir, T.W. 1984. Urease, phosphatase, and sulphatase activities of Cook Island and Tongan soils. N.Z. J. Sci. 27:73–79.

Tabatabai, M.A. 1977. Effects of trace elements on urease activity in soils. Soil Biol. Biochem. 9:9–13.

Tabatabai, M.A. 1982. Soil enzymes, p. 903–947. *In* A.L. Page, R.H. Miller, and D.R. Keeney (eds.) Methods of soil analysis, Part 2, 2nd edn., Agronomy 9. ASA, Madison, WI.

Tabatabai, M.A. 1994. Soil enzymes, p. 775–833. *In* R.W. Weaver, S. Angle, and P. Bottomley (eds.) Methods of soil analysis, Part 2: Microbiological and biochemical properties. SSSA, Madison, WI.

Tabatabai, M.A., K. Alef, P. Nannipieri, W.T. Frankenberger, and C. Trarar-Cepeda. 1995. Enzyme activities, p. 311–373. *In* K. Alef and P. Nannipieri (eds.) Methods in applied soil microbiology and biochemistry. Academic Press, London, U.K.

Tabatabai, M.A., and W.A. Dick. 2002. Enzymes in soil: Research and developments in measuring activities, p. 567–596. *In* R.G. Burns and R.P. Dick (eds.) Enzymes in the environment. Activity, ecology and applications. Marcel Dekker, New York.

Tabatabai, M.A., and M. Fu. 1992. Extraction of enzymes from soils, p. 197–227. *In* G. Stozky and J.-M. Bollag (eds.) Soil biochemistry. Vol. 7. Marcel Decker, New York.

Tabatabai, M.A., and B.B. Singh. 1976. Rhodanese activity of soils. Soil Sci. Soc. Am. J. 40:381–385.

Tarafdar, J.C., and A. Jungk. 1987. Phosphatase activity in the rhizosphere and its relation to the depletion of soil organic phosphorus. Biol. Fertil. Soils 3:199–204.

Torsvik, V.L., R. Sørheim, and J. Goksøyr. 1996. Total bacterial diversity in soil and sediment communities—A review. J. Ind. Microbiol. 17:170–178.

Tyler, G. 1974. Heavy metal pollution and soil enzymatic activity. Plant Soil 41:303–311.

Tyler, G. 1981. Heavy metals in soil biology and biochemistry, p. 371–414. *In* E.A. Paul and J.N. Ladd (eds.) Soil biochemistry. Vol. 5. Marcel Dekker, New York.

Tyler, G., A.M. Balsberg-Pahlson, G. Bengtsson, E. Bååth, and L. Tranvik. 1989. Heavy-metal ecology of terrestrial plants, microorganisms and invertebrates. Water Air Soil Pollut. 47:189–215.

Whipps, J.M. 1990. Carbon economy. *In* J.M. Lynch (ed.) The rhizosphere. Wiley, West Sussex, U.K.

Wilde, C.E., and R.G.O. Kekwick. 1964. The arylesterase of human serum. Biochem. J. 91:297–307.

Wirth, S.J., and G.A. Wolf. 1992. Micro-plate colourimetric assay for endoacting cellulase, xylanase, chitinase, 1,3-beta-glucanase and amylase extracted from forest soil horizons. Soil Biol. Biochem. 24:511–519.

Zantua, M.I., and J.M. Bremner. 1976. Production and persistence of urease in soil. Soil Biol. Biochem. 8:369–374.

Zornoza, R., L. Landi, P. Nannipieri, and G. Renella. 2009. A protocol for the assay of arylesterase activity in soil. Soil Biol. Biochem. 41:559–662.

27

Nitrogen Transformations

Richa Anand
The University of British Columbia

Jean-Claude Germon
*Institut National de la
Recherche Agronomique*

Peter M. Groffman
University of Georgia

Jeanette M. Norton
Utah State University

Laurent Philippot
*Institut National de la
Recherche Agronomique*

James I. Prosser
University of Aberdeen

Joshua P. Schimel
University of California

27.1 Biological Nitrogen Fixation

Richa Anand

27.1.1 Introduction

Nitrogen is abundant in the atmosphere, lithosphere, and hydrosphere of the earth. However, 99% of nitrogen is in the form of dinitrogen, which is inert and cannot be used by most living organisms. In order for living cells to use it for synthesis of vital bioorganic molecules like proteins, nucleic acids, and vitamins, molecular nitrogen (dinitrogen) has to be in its reduced or fixed form. Dinitrogen fixation, the process by which dinitrogen is reduced to ammonia, is therefore a very important process for the sustenance of life.

Three processes are responsible for most of the dinitrogen fixation in the biosphere. Atmospheric fixation by lightning contributes approximately 5%–8% (5×10^9 kg N year^{-1}) of total fixed

nitrogen (Myrold and Bottomley, 2007). The enormous energy contained in lightning breaks dinitrogen molecules and enables their atoms to combine with oxygen in the air forming nitrogen oxides that dissolve in rain. These oxides of nitrogen then form nitrates that are carried to the earth in rainfall. Industrial nitrogen fixation occurs through a process called the Haber–Bosch industrial process that was established in 1913. This process uses a catalytic agent (iron with a small amount of aluminum added) at high pressure (as much as 5.06×10^7 Pa) and high temperature (600–800 K) and consumes large amounts of fossil fuel. Ammonia produced through this highly expensive process is combined with other elements to produce nitrogenous fertilizers like urea and ammonium nitrate. Although the use of these fertilizers is inevitable in meeting rising food demand to sustain the growing global population, their indiscriminate use has set off very negative effects on water resources and the environment. Approximately 1.1×10^{11} kg N (Myrold and Bottomley, 2007) are manufactured by ammonia industry annually. Increasing fossil

fuel combustion and burning of forests and grasslands contribute approximately 2.5×10^{10} of anthropogenic reactive nitrogen (Nr) (Galloway and Cowling, 2002) that is proving menacing to the environment. Biological nitrogen fixation (BNF), a natural process by which certain prokaryotic microorganisms fix nitrogen by a highly specialized enzyme complex called nitrogenase, is an environmentally benign source of plant-usable fixed nitrogen. According to estimates, approximately 1×10^{11}–1.4×10^{11} kg N (Myrold and Bottomley, 2007) are fixed from the atmosphere by BNF every year. While appreciating the contributions of all sources of fixed nitrogen, this chapter focuses on the process and applications of BNF.

Farmers since ancient Chinese and Roman civilizations have practiced crop rotation with legumes to increase soil fertility and agricultural productivity. However, the science behind such practice was revealed by Boussingault who experimented with leguminous crops fixing N_2 in 1838. In 1886, Hellriegel and Wilfarth showed definitive evidence for N_2 fixation by microbes in legumes.

27.1.2 Nitrogen-Fixing Organisms

The ability to fix nitrogen is limited to bacteria and archaea only. Within these groups, it is quite widely distributed, revealing considerable phylogenetic diversity among diazotrophic organisms. Young (1992) has prepared a comprehensive listing of nitrogen-fixing bacteria and archaea, under 12 broad phylogenetic groups based on 16S rDNA phylogeny including green sulfur bacteria (*Chlorobium*), Firmicutes (*Paenibacillus*, *Clostridium*), thallobacteria (*Frankia*), cyanobacteria (*Anabaena*, *Nostoc*), and all subdivisions of the proteobacteria (α: *Rhizobium*, *Acetobacter*; β: *Azoarcus*, *Herbaspirillum*; γ: *Klebsiella*; δ: *Desulfovibrio*) and archaea (mostly methanotrophs). Dinitrogen-fixing organisms also exhibit quite diverse physiologies (Figure 27.1).

Diazotrophs are also widely distributed ecologically. They are found free living in soils and water (*Klebsiella*, cyanobacteria), in root nodule symbiosis with legumes (*Rhizobium*), associative symbiosis with grasses (*Azoarcus*, *Gluconacetobacter*), actinorhizal associations with woody plants (*Frankia*), cyanobacterial symbiosis with various plants, and symbiotic associations in termite guts.

27.1.3 Chemistry of Nitrogen Fixation

The overall chemical reaction of biological dinitrogen fixation by nitrogenase is represented by the equation

$$N_2 + 8H^+ + 8e^- + 16MgATP \rightarrow 2NH_3 + H_2 + 16MgADP + 16Pi.$$

(27.1)
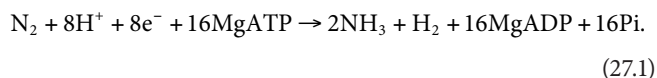
Due to the stark disparity of energy and heat requirements between chemical and biological N fixation, the structure and function of nitrogenase have always been of immense interest to biochemists (Dance, 2007).

Nitrogenase is a complex enzyme comprising two metalloproteins, the Mo–Fe protein, also called dinitrogenase or component I and the Fe protein called dinitrogenase reductase or component II.

The dinitrogenase responsible for the actual reduction of dinitrogenase to ammonia is a heterotetramer composed of two alpha and two beta subunits with an overall molecular weight of 240 kDa. The Mo–Fe protein contains two types of metal centers, the FeMo-cofactor and the P cluster pair, of which the FeMo-cofactor is the active site where dinitrogen binds, whereas the P cluster mediates electron transfer between the Fe protein and the FeMo-cofactor.

The dinitrogenase reductase or Fe protein is a homodimer of two identical subunits, with an overall molecular mass of ~60 kDa. It contains two ATP/ADP molecules and one 4Fe–4S (Kim and Rees, 1994).

Under conditions of molybdenum depletion, alternative nitrogenase systems may be induced. Molecular genetic studies have shown that Mo-independent nitrogenases are quite widely distributed among diazotrophs and vanadium (V) nitrogenases of *Azotobacter chroococcum* and *A. vinelandii*, and the Fe nitrogenases of *A. vinelandii* and *Rhodobacter capsulatus* are the best known for these (Eady, 1996). *Streptomyces thermoautotrophicus* has been recently found to be able to fix dinitrogen, but it harbors a very unusual N_2-fixing system that requires three proteins for nitrogen fixation, a heterotrimeric molybdenum-containing dinitrogenase (St1), a homodimeric manganese-containing superoxide oxidoreductase (St2), and another heterotrimeric molybdenum-containing carbon monoxide dehydrogenase (St3 or CODH). These proteins differ entirely from the known
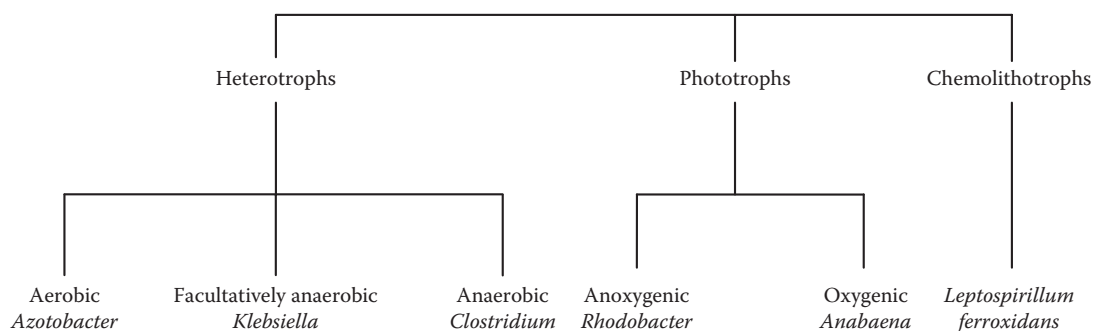
FIGURE 27.1 Diversity of diazotrophs based on physiology.

nitrogenase protein components and are insensitive to O_2. Compared to conventional or alternative nitrogenases, the St nitrogenase also requires less ATP (Qi, 2008).

The overall functioning of nitrogenase can be summarized as a key biochemical cycle also called Fe-protein cycle (Dance, 2007), which involves five steps (Kim and Rees, 1994; Dance, 2007): (1) the reduction of Fe protein by electron carriers such as flavodoxin or ferredoxin, (2) association of the reduced Fe protein (including two MgATP complexes) with the Mo–Fe protein in preparation of electron transfer, (3) hydrolysis of MgATP, which enables transfer of one electron to the Mo–Fe protein (via Fe_4S_4 and the P cluster), (4) electron transfer to dinitrogen and thus its reduction, while it is bound to the active site within the Mo–Fe protein, and (5) dissociation of the two protein molecules, exchange of ATP back into the Fe protein, and re-reduction of the Fe protein. Dance (2008) has reviewed chemical catalysis of nitrogenase in detail, addressing questions such as how N_2 binds the FeMo-cofactor, how exactly electrons travel through the system, and finally how NH_3 leaves the complex.

27.1.4 Genetics of Nitrogen Fixation

The structure and function of nitrogenase are coded by close to 20 genes (*nif* genes) organized in 7 operons (*nif* cluster) spanning over 24 kb. These genes fall into three categories: structural, regulatory, and supplementary and can be housed either in genomic DNA or on plasmids (e.g., sym plasmids in *Sinorhizobium meliloti*). The *nif* cluster of the free-living bacterium *Klebsiella pneumoniae* is the most studied of *nif* genes and serves as a model for understanding the regulation, synthesis, and assembly of nitrogenase. The function and arrangement of these genes are described in Figure 27.2.

Genes governing alternative and V-dependent nitrogenases are called *anf* and *vnf* genes, respectively. These genes share extensive homology with *nif* genes with some differences in sequence and arrangement. Eady (1996) has discussed detailed structure, function, and genetics of these alternative nitrogenases.

Other plant and bacterial genes involved in establishing plant–diazotroph interactions such as symbiosis are also known. Of these, the legume genes called *nod* genes are required for the early steps in nodule formation (Debellé et al., 2002). Other genes that are essential for nitrogen fixation by rhizobia but have no homologue in *K. pneumoniae* are known as "*fix*" genes. The most commonly known of these are *fix ABCX*, which are believed to be involved in electron transport, *fix NOPQ*, which are involved in bacteroid respiration under conditions of low oxygen present in nodules, and *fix GHIS* that are involved in redox (Fischer, 1994).

27.1.5 Methods of Studying Biological Nitrogen Fixation

27.1.5.1 Isolation of Diazotrophs from Plants and Soil

Nitrogen-free media with multiple carbon sources are central to the isolation of most culturable diazotrophs. These media used in semisolid state are ideal for simulating the microaerophillic

Q	Mo uptake or processing
B	FeMoCo synthesis
A	Positive regulator
L	Negative regulator
F	Flavodoxin synthesis
M	Dinitrogenase reductase processing
Z	FeMoCo synthesis
W	FeMoCo synthesis
V	Homocitrate synthesis
S	Metal center biosynthesis
U	Metal center biosynthesis
X	Unknown
N	FeMoCo synthesis
E	FeMoCo synthesis
Y	FeMoCo insertion into dinitrogenase
T	FeMoCo insertion into dinitrogenase
K	Dinitrogenase β subunit
D	Dinitrogenase α subunit
H	Dinitrogenase reductase
J	Pyruvate-flavodoxin oxido reductase

FIGURE 27.2 Arrangement and functions of various *nif* genes in *K. pneumonia*. Arrows represent the seven operons in which *nif* genes are organized and the direction of translation. (Modified from Madigan, M.T., J.M. Martinko, and J. Parker (eds.). 2000. Metabolic diversity, p. 586–590. *In* Brock's biology of microorganisms. Prentice Hall, Upper Saddle River, NJ.)

conditions ideal for nitrogen fixation. Examples of some such media are yeast extract mannitol (YEM) medium, Vincent's minimal medium (Vincent, 1970), and combined carbon medium (Rennie, 1981). The use of intact pieces (0.5–1.0 cm) of roots and other plant parts has been suggested for isolating diazotrophs adhering to the surface of tissues (Bashan et al., 1993). For isolation of all diazotrophs from a plant, parts are washed and macerated and serially diluted before incubation on semisolid, nitrogen-free medium. Endophytic bacteria are isolated by surface sterilizing tissues before grinding and incubating on medium (Cavalcante and Döbereiner, 1988). Similarly, serial dilutions of rhizospheric and nonrhizospheric soils are used to isolate culturable diazotrophs from soil (Seldin et al., 1983).

Complete communities of culturable as well as unculturable diazotrophs associated with soils and plants can also be profiled using various molecular techniques that utilize the polymorphism in the *nifH* gene, which is believed to be quite conserved across diazotrophs (Zehr et al., 2003), to determine the diversity in the *nifH* gene pool in the target environmental sample. This is done

by amplifying the *nifH* gene by the polymerase chain reaction (PCR) using environmental DNA as a template. The PCR product, which is a pool of almost similar amplicons of various *nifH* genes in the sample, is subjected to subsequent analyses by cloning and sequencing, restriction fragment length polymorphism (RFLP) (Poly et al., 2001), terminal-RFLP (T-RFLP; Deslippe et al., 2005), or by denaturing gradient gel electrophoresis (DGGE; Rosado et al., 1998). These techniques are limited to detecting the presence of diazotrophs in the community but do not provide any information on the frequency of a particular species. Macro- and micro-arrays may reveal both, the presence and frequency of different N_2-fixing prokaryotes (Steward et al., 2004; Moisander et al., 2006).

27.1.5.2 Quantification

The most commonly used methods for measurement of nitrogen fixation are nitrogen balance (also called the nitrogen accretion or nitrogen difference method), the acetylene reduction assay (ARA), xylem solute analysis (ureide production), and stable isotope methods (^{15}N isotope dilution, natural abundance, and incorporation). The major considerations for selection are sensitivity, duration, sample type, and whether relative or absolute rates are required (Myrold and Tiedje, 1986). A more detailed review with examples of application has been done by Danso (1995). The nitrogen balance method is the oldest and simplest method that measures BNF as the difference between the total N contents of plants that fix nitrogen versus those that do not. It is highly dependent and thus disadvantaged by its underlying assumption that both nitrogen fixing and nonfixing control plants absorb equal amount of N from soil. This assumption is hard to justify due to differences in root morphology and other physiological attributes.

The xylem solute method of measuring BNF is based on the determination of the composition of nitrogen compounds in plant tissues or the N flowing through the xylem sap to the shoot. The underlying idea is to differentiate between nitrogen absorbed from soil, which is predominantly nitrate and fixed nitrogen occurring in the form of N solutes, primarily amides and ureides (McClure and Israel, 1979). In the presence of increasing levels of soil nitrate, plants get less dependent on nitrogen fixation and the ureide content of xylem solute decreases. Although the method has sometimes been used to even make quantitative measurements of nitrogen fixation (Herridge, 1984), it is severely hindered by the fact that only a very small proportion of nitrogen-fixing plants export fixed nitrogen in the form of ureides (Van Kessel et al., 1988).

The ARA described by Hardy et al. (1968) is a very popular technique that is used to indirectly measure BNF by estimating enzyme activity based on electron flux through nitrogenase. It is based on the ability of nitrogenase to reduce acetylene ($CH\equiv CH$) to ethylene. Samples to be assayed are incubated in a gas tight chamber that is injected with 0.03–0.1 (v/v) acetylene. The gas collected from the chamber after the end of incubation is assayed for ethylene production using a gas chromatograph fitted with Porapak N or P column. It is a simple, low cost, and sensitive assay that can measure BNF in bacterial cultures, detached nodules, plant parts, or even whole plants. The acceptance of ARA as a sole basis of interpretations of BNF has been hindered by several problems including the short-term nature of the assays, the doubtful validity of always using a conversion ratio of 3 and the autoinhibition of acetylene conversion to ethylene (Danso, 1995).

The stable isotope methods using ^{15}N are more widely used and accepted over all other methods of BNF measurement. These methods are based on the principle that the soil or medium in which plants are grown has a distinctly different $^{15}N/^{14}N$ ratio than the almost constant 0.3663% ratio present in the atmosphere. Therefore, plants incorporating fixed nitrogen from the atmosphere will have a $^{15}N/^{14}N$ ratio different from the substrate they are growing on. When N-fixing plants are grown in air labeled with ^{15}N, they are expected to have an enhanced ratio as compared to substrate (^{15}N incorporation method). Where available soil N is labeled with ^{15}N, a reduction in the ratio is expected when plants incorporate fixed nitrogen from air. This method is also called isotope dilution. Isotope dilution methods rely either on the inherent higher $^{15}N/^{14}N$ ratios of growth substrates than that of the atmosphere (natural abundance) or on ^{15}N enrichment of the substrate via labeled fertilizer. In both cases, a reduced $^{15}N/^{14}N$ ratio indicates nitrogen fixation. However, the enrichment method is more common because of the clearer distinction between substrate and atmospheric ratios, which allows easier detection using relatively less costly equipment (Danso et al., 1993).

27.1.6 Nitrogen Fixation in Agriculture and Forestry

Biological N_2 fixation contributes to productivity both directly, where the fixed N_2 is harvested in grain or other food for human or animal consumption, or indirectly, by contributing to the maintenance or enhancement of soil fertility in the agricultural system by adding N to the soil (Giller and Cadisch, 1995). Among symbiotic N_2-fixing systems, nodulated legumes have been used in cropping systems for centuries. They can serve a multitude of purposes in sustainable agriculture. They are used as primary sources of food, fuel, fiber, and fertilizer, or, secondarily, to enrich the soil, preserve moisture, and prevent soil erosion. The *Rhizobium*–legume symbiosis was the focus of agronomic practices and nitrogen fixation research during the nineteenth century and for most part of the twentieth century. However, since the latter half of the twentieth century, researchers have started focusing on extending nitrogen fixation to include nonlegumes as well. Efforts have been made to induce paranodules on roots and inoculate them with effective rhizobial strains (Cocking et al., 1995). The process involves application of auxins, cell wall–degrading enzymes, or genetic modification of bacteria that are not very practical for field application (Bruijn et al., 1995). Exploration of naturally occurring diazotrophic associations in nonlegumes has proven to be a better approach (James and Olivares, 1998). Sugarcane research in Brazil presents a valuable example of this approach. The observation that sugarcane had always been grown successfully in Brazil with little or no nitrogen inputs led Döbereiner (1961) to suggest that sugarcane might be meeting its nitrogen demands with the help of naturally associated nitrogen-fixing bacteria.

Early efforts to explore diazotrophs in sugarcane fields resulted in the isolation of bacteria of the genus *Beijerinckia* in high numbers, with selective enrichment in the rhizosphere and especially on the root surface (Döbereiner, 1961). Diazotrophes of the genera *Erwinia*, *Azotobacter*, *Derxia*, *Azospirillum*, and *Enterobacter* (Purchase, 1980; Gracioli and Ruschel, 1981) were also found from the roots, stems, and even leaves of sugarcane. However, none of these bacteria seemed to occur in large enough numbers to account for the 38–77 Kg N ha^{-1} of mean annual BNF input reported in various studies (Oliveira et al., 1994; Boddey et al., 1995). In 1988, Cavalcante and Döbereiner found an extraordinary diazotroph, *Gluconacetobacter diazotrophicus*, in large numbers within the roots and stems of sugarcane. It is a small, Gram negative, aerobic rod-like bacterium that can grow on very high concentrations of sugar and has the capability to fix N (>100n mol C$_2$H$_2$ mL^{-1} ha^{-1}) at a pH as low as 3 (Stephan et al., 1991) and oxygen levels of 4 kPa (Reis et al., 1990). Its ability to fix nitrogen is also not affected by the presence of high levels of nitrate (Boddey et al., 1991). The discovery of a bacterium possessing such unique properties adapted to fixing nitrogen under conditions very specific to the interiors of a sugarcane plant has encouraged researchers to believe that such systems might be in place for other nonlegumes waiting to be discovered. Some other examples of diazotrophic bacteria associated with other nonlegumes are *Herbaspirillum*, *Azospirillum*, and *Klebsiella* with corn (Chelius and Triplett, 2001), *Herbaspirillum*–rice (Baldani et al., 1986), *Azospirillum*–wetland rice (Döbereiner and Pedrosa, 1987), *Azoarcus*–kallar grass (Reinhold-Hurek et al., 1993). Such associations are also found in horticultural plants, for example, *Burkholderia* with tomato (Caballero-Mellado et al., 2007), *K. variicola* with banana (Martínez et al., 2003), *Azotobacter* with pepper (Govedarica et al., 1997).

Studies revealing associations between endophytic diazotrophic bacteria and forest trees, particularly lodgepole pine are underway in our laboratory and have been described in Anand et al. (2006). These studies are expected to help answer long-standing questions about the presence of unexplained nitrogen in some forest ecosystems (Fenn et al., 1998; Binkley et al., 2000).

Despite much advancement in our basic knowledge of associative nitrogen fixation, the extent of true benefits from it in the field remains to be established especially in temperate agriculture (Peoples, 2006).

27.1.7 Biological Nitrogen Fixation: Economy and Environment

The economic and environmental costs of the heavy use of chemical N fertilizers in agriculture are a cause of global concern. The manufacture of nitrogenous fertilizers is heavily dependent on the already dwindling and disputed reserves of fossil fuel. Their excessive application has adverse effects on human health (methemoglobinemia in infants, cancer, respiratory illness) and the environment (nitrate leaching into water sources, eutrophication, plant toxicity, ozone depletion; Bohlool et al., 1992). BNF is an economically feasible and environmentally benign process that can be used alone or in conjunction with nitrogenous fertilizers to sustain food production particularly in the poor and developing countries where more than 78% of the world's population resides (Khush, 2001). The Food and Agricultural Organization emphasizes on the importance of beans and other legumes in the diets of people in these parts of the world. The contribution of BNF to the social and economic success of Brazil through bean and sugarcane cultivation (Döbereiner, 1997) is an important model for the potential of this nitrogen input in the developing countries in the tropics and subtropics.

Another important economic and environmental application of BNF is in the cultivation of energy crops like sugarcane, for the production of biofuels (Döbereiner, 1997). The elimination of N fertilizer for the production of biofuel crops, such as sugarcane, represents the key to high-energy balances because N fertilizer is produced by reduction of atmospheric N$_2$ to NH$_4$, using natural gas as an energy source (Reis et al., 2007). Brazil where subsidies on nitrogen fertilizers are traditionally low and sugarcane is grown with the help of natural and inoculant-aided BNF (Boddey et al., 1995) has been really successful in producing and utilizing these biofuels. Baldani et al. (2002) have reported that as many as three million cars in Brazil run on 95% hydrous ethanol and all gasoline sold in Brazil contains 20%–24% ethanol as biofuel.

Nitrogen-fixing bacteria have also been found to be useful for bioremediation of industrial waste and hydrocarbon spills. Ikechukwu and Onwurah (1999) found that *Azotobacter* played a role in bioremediation of soil polluted with crude oil by providing other bacteria with nitrogen and through some cometabolic activities. Some diazotrophs isolated from petrochemical sludge have the ability to degrade organic contaminants like polyethylene glycol, naphthalene, and hexadecane (Naumova et al., 2009).

27.1.8 Factors Limiting Biological Nitrogen Fixation

Bohlool et al. (1992) have categorized the constraints to practical application of BNF systems as environmental, biological, methodological, and sociocultural.

Major environmental constraints are concerned with conditions affecting the microbe, the host, or their symbiotic interaction. These include soil acidity, aluminum and manganese toxicity, sulfur and phosphorus deficiency (law of minimum), salinity, and soil aeration. A thorough understanding of these factors and how they affect different BNF systems is required to overcome these limitations. A major obstacle to successful establishment and effective performance of introduced N$_2$-fixing systems is also competition from native organisms.

A variety of biological constraints may influence the expression of BNF in all nitrogen-fixing systems. Both plants and bacteria are subject to biological constraints, such as disease and predation that can directly or indirectly affect the amount of N fixed. In general, the amount of nitrogen fixed is directly related to the growth potential of the host in a particular system. When growth is limited, for example, by disease, nitrogen fixation will be reduced accordingly.

Identification, preparation, and application of diazotrophic inoculants are major methodological constraints to widespread field application of BNF systems. The scale of production, the availability of suitable carrier material, and shelf life of the finished product limit the use of inoculants, especially in developing countries. Various cultural, educational, economic, and political factors also affect fuller implementation of BNF. Small and subsistence farmers lack the awareness and means to utilize this technology, especially in developing countries, while subsidies on nitrogen fertilizers (Bohlool et al., 1992) in both developed and developing countries, are a huge disincentive to the use of BNF, for medium and large farms. Giller and Cadisch (1995) have discussed these constraints and ideas for overcoming them, with examples from Asia, Africa, Europe, and North America.

References

Anand, R., L. Paul, and C. Chanway. 2006. Research on endophytic bacteria: Recent advances with forest trees, p. 89–106. *In* B.J.E. Schulz, C.J.C. Boyle, and T.N. Sieber (eds.) Microbial root endophytes. Springer, Berlin, Germany.

Baldani, J.I., V.L.D. Baldani, L. Seldin, and J. Döbereiner. 1986. Characterization of *Herbaspirillum seropedicae* gen. nov., sp. nov., a root-associated nitrogen-fixing bacterium. Int. J. Syst. Bacteriol. 36:86–93.

Baldani, J.I., V.M. Reis, V.L.D. Baldani, and J. Döbereiner. 2002. A brief story of nitrogen fixation in sugarcane—Reasons for success in Brazil. Funct. Plant Biol. 29:417–423.

Bashan, Y., G. Holguin, and R. Lifshitz. 1993. Isolation and characterization of plant growth-promoting rhizobacteria, p. 331–345. *In* B.R. Glick and J.E. Thompson (eds.) Methods in plant molecular biology and biotechnology. CRC Press, Boca Raton, FL.

Binkley, D., Y. Son, and D.W. Valentine. 2000. Do forests receive occult inputs of nitrogen? Ecosystems 3:321–331.

Boddey, R.M., O.C. Oliveira, S. Urquiaga, V.M. Reis, F.L. Olivares, V.L.D. Baldani, and J. Döbereiner. 1995. Biological nitrogen fixation associated with sugar cane and rice: Contributions and prospects for improvement. Plant Soil 174:195–209.

Boddey, R., S. Urquiaga, V. Reis, and J. Döbereiner. 1991. Biological nitrogen fixation associated with sugar cane. Plant Soil 137:111–117.

Bohlool, B.B., J.K. Ladha, D.P. Garrity, and T. George. 1992. Biological nitrogen fixation for sustainable agriculture: A perspective. Plant Soil 141:1–11.

Bruijn, F.J., Y. Jing, and F.B. Dazzo. 1995. Potential and pitfalls of trying to extend symbiotic interactions of nitrogen-fixing organisms to presently non-nodulated plants, such as rice. Plant Soil 174:225–240.

Caballero-Mellado, J., J. Onofre-Lemus, P. Estrada-de los Santos, and L. Martinez-Aguilar. 2007. The tomato rhizosphere, an environment rich in nitrogen-fixing *Burkholderia* species with capabilities of interest for agriculture and bioremediation. Appl. Environ. Microbiol. 73:5308–5319.

Cavalcante, V., and J. Döbereiner. 1988. A new acid-tolerant nitrogen-fixing bacterium associated with sugarcane. Plant Soil 108:23–31.

Chelius, M.K., and E.W. Triplett. 2001. The diversity of archaea and bacteria in association with the roots of *Zea mays* L. Microb. Ecol. 41:252–263.

Cocking, E.C., S.I. Kothari, C.A. Batchelor, S. Jain, G. Webster, J. Jones, J. Jotham, and M.R. Davey. 1995. Interaction of rhizobia with non-legume crops for symbiotic nitrogen fixation nodulation, p. 197–205. *In* I. Fendrik, M. del Gallo, J. Vanderleyden, and M. de Zamaroczy (eds.) Azospirillum VI and related microorganisms: Genetics, physiology, ecology. Springer-Verlag, New York.

Dance, I. 2007. Elucidating the coordination chemistry and mechanism of biological nitrogen fixation. Chem. Asian J. 2:936–946.

Dance, I. 2008. The chemical mechanism of nitrogenase: Calculated details of the intramolecular mechanism for hydrogenation of η^2-N_2 on FeMo-co to NH_3. Dalton Trans. 5977–5991.

Danso, S.K.A. 1995. Assessment of biological nitrogen fixation. Nutr. Cycl. Agroecosyst. 42:33–41.

Danso, S.K.A., G. Hardarson, and F. Zapata. 1993. Misconceptions and practical problems in the use of ^{15}N soil enrichment techniques for estimating N_2 fixation. Plant Soil 152:25–52.

Debellé, F., L. Moulin, B. Mangin, J. Dénarié, and C. Boivin. 2002. *Nod* genes and Nod signals and the evolution of the *Rhizobium* legume symbiosis. Acta Biochim. Pol. 48:359–365.

Deslippe, J.R., K.N. Egger, and G.H.R. Henry. 2005. Impacts of warming and fertilization on nitrogen-fixing microbial communities in the Canadian High Arctic. FEMS Microbiol. Ecol. 53:41–50.

Döbereiner, J. 1961. Nitrogen-fixing bacteria of the genus *Beijerinckia Derx* in the rhizosphere of sugar cane. Plant Soil 15:211–216.

Döbereiner, J. 1997. Biological nitrogen fixation in the tropics: Social and economic contributions. Soil Biol. Biochem. 29:771–774.

Döbereiner, J., and F.O. Pedrosa. 1987. Nitrogen-fixing bacteria in nonleguminous crop plants, p. 155. *In* Brock/Springer Series in contemporary biosciences. Science Tech Publishers, Madison, WI.

Eady, R.R. 1996. Structure–function relationships of alternative nitrogenases. Chem. Rev. 96:3013–3030.

Fenn, M.E., M.A. Poth, J.D. Aber, J.S. Baron, B.T. Bormann, D.W. Johnson, A.D. Lemly, S.G. McNulty, D.F. Ryan, and R. Stottlemyer. 1998. Nitrogen excess in North American ecosystems: Predisposing factors, ecosystem responses, and management strategies. Ecol. Appl. 8:706–733.

Fischer, H.M. 1994. Genetic regulation of nitrogen fixation in rhizobia. Microbiol. Mol. Biol. Rev. 58:352–386.

Galloway, J.N., and E.B. Cowling. 2002. Reactive nitrogen and the world: 200 years of change. Ambio 31:64–71.

Giller, K.E., and G. Cadisch. 1995. Future benefits from biological nitrogen fixation: An ecological approach to agriculture. Plant Soil 174:255–277.

Govedarica, M., M. Nada, J. Mirjana, D. Milošev, and D. Simonida. 1997. Diazotrophs and their activity in pepper. Acta Hortic. 462:725–732.

Gracioli, L.A., and A.P. Ruschel. 1981. Microorganisms in the phyllosphere and rhizosphere of sugarcane, p. 87–101. *In* P.B. Vose and A.P. Ruschel (eds.) Associative nitrogen fixation. CRC Press, Boca Raton, FL.

Hardy, R.W.F., R.D. Holsten, E.K. Jackson, and R.C. Burns. 1968. The acetylene–ethylene assay for N_2 fixation: Laboratory and field evaluation. Plant Physiol. 43:1185–1207.

Herridge, D.F. 1984. Effects of nitrate and plant development on the abundance of nitrogenous solutes in root-bleeding and vacuum-extracted exudates of soybean. Crop Sci. 24:73–179.

Ikechukwu, N., and E. Onwurah. 1999. Role of diazotrophic bacteria in the bioremediation of crude oil-polluted soil. J. Chem. Technol. Biotechnol. 74:957–964.

James, E.K., and F. Olivares. 1998. Infection and colonization of sugar cane and other graminaceous plants by endophytic diazotrophs. Crit. Rev. Plant Sci. 17:77–119.

Jones, A.L. *Phaseolus* bean: Post-harvest operations. FAO, Rome, Italy. Available online at: http://www.fao.org/inpho/content/compend/text/ch04.htm

Khush, G. 2001. Challenges for meeting the global food and nutrient needs in the new millennium. Proc. Nutr. Soc. 60:15–26.

Kim, J., and D.C. Rees. 1994. Nitrogenase and biological nitrogen fixation. Biochemistry 33:389–397.

Madigan, M.T., J.M. Martinko, and J. Parker (eds.). 2000. Metabolic diversity, p. 586–590. *In* Brock's biology of microorganisms. Prentice Hall, Upper Saddle River, NJ.

Martínez, L., J. Caballero-Mellado, J. Orozco, and E. Martínez-Romero. 2003. Diazotrophic bacteria associated with banana (*Musa* spp.). Plant Soil 257:35–47.

McClure, P.R., and D.W. Israel. 1979. Transport of nitrogen in the xylem of soybean plants. Plant Physiol. 64:411–416.

Metting, F.B. 1993. Structure and physiological ecology of soil microbial communities, p. 3–25. *In* F.B. Metting (ed.) Soil microbial ecology: Applications in agricultural and environmental management. Marcel Dekker, New York.

Moisander, P.H., L. Shiue, G.F. Steward, B.D. Jenkins, B.M. Bebout, and J.P. Zehr. 2006. Application of a *nifH* oligonucleotide microarray for profiling diversity of N_2-fixing microorganisms in marine microbial mats. Environ. Microbiol. 8:1721–1735.

Myrold, D.D., and P.J. Bottomley. 2007. Biological N inputs, p. 365–388. *In* E.A. Paul (ed.) Soil microbiology, ecology and biochemistry. Elsevier, Burlington, MA.

Myrold, D.D., and J.M. Tiedje. 1986. Simultaneous estimation of several nitrogen cycling rates using [15]N: Theory and application. Soil Biol. Biochem. 18:559–568.

Naumova, R.P., T.V. Grigoryeva, A.A. Rizvanov, J.V. Gogolev, N.V. Kudrjashova, and A.V. Laikov. 2009. Diazotrophs originated from petrochemical sludge as a potential resource of waste remediation. World Appl. Sci. J. 6:154–157.

Oliveira, O.C., S. Urquiaga, and R.M. Boddey. 1994. Burning cane: The long term effects. Int. Sugar J. 96:272–275.

Peoples, M.B. 2006. Biological nitrogen fixation: Contributions to agriculture, p. 162–165. *In* R. Lal (ed.) Encyclopedia of Soil Science. CRC Press, Boca Raton, FL.

Poly, F., L.J. Monrozier, and R. Bally. 2001. Improvement in the RFLP procedure for studying the diversity of *nifH* genes in communities of nitrogen fixers in soil. Res. Microbiol. 152:95–103.

Purchase, B.S. 1980. Nitrogen fixation associated with sugarcane, p. 173–176. Proc. S. Afr. Sugar Technol. Assoc. June. Natal, South Africa.

Qi, C. 2008. Perspectives in biological nitrogen fixation research. J. Integr. Plant Biol. 50:786–798.

Reinhold-Hurek, B., T. Hurek, M. Gillis, B. Hoste, M. Vancanneyt, K. Kersters, and J. De Ley. 1993. *Azoarcus* gen. nov., nitrogen-fixing proteobacteria associated with roots of kallar grass (*Leptochloa fusca* (L.) Kunth), and description of two species, *Azoarcus indigens* sp. nov. and *Azoarcus communis* sp. nov. Int. J. Syst. Bacteriol. 43:574–584.

Reis, V., S. Lee, and C. Kennedy. 2007. Biological nitrogen fixation in sugarcane, p. 213–232. *In* C. Elmerich and W.E. Newton (eds.) Associative and endophytic nitrogen-fixing bacteria and cyanobacterial associations. Springer, Amsterdam, the Netherlands.

Reis, V.M., S. Urquiaga, M.A. Paula, and J. Döbereiner. 1990. Infection of sugarcane by *Acetobacter diazotrophicus* and other diazotrophs. Nitrogen fixation: Achievements and objectives. Proc. 8th Int. Congr. Nitrogen fixation. May 20–26, 1990. Knoxville, TN.

Rennie, R.J. 1981. A single medium for the isolation of acetylene reducing (dinitrogen fixing) bacteria from soil. Can. J. Microbiol. 27:8–14.

Rosado, A.S., G.F. Duarte, L. Seldin, and J.D. Van Elsas. 1998. Genetic diversity of *nifH* gene sequences in *Paenibacillus azotofixans* strains and soil samples analyzed by denaturing gradient gel electrophoresis of PCR-amplified gene fragments. Appl. Environ. Microbiol. 64:2770–2779.

Seldin, L., J.D.V. Elsas, and E.G.C. Penido. 1983. Bacillus nitrogen fixers from Brazilian soils. Plant Soil 70:243–255.

Stephan, M.P., M. Oliveria, K.R.S. Teixeira, G. Martinez-Drets, and J. Döbereiner. 1991. Physiology and dinitrogen fixation of *Acetobacter diazotrophicus*. FEMS Microb. Lett. 77:67–72.

Steward, G.F., B.D. Jenkins, B.B. Ward, and J.P. Zehr. 2004. Development and testing of a DNA macroarray to assess nitrogenase (*nifH*) gene diversity. Appl. Environ. Microbiol. 70:1455–1465.

Van Kessel, C., J.P. Roskoski, and K. Keane. 1988. Ureide production by N_2-fixing and non-N_2-fixing leguminous trees. Soil Biol. Biochem. 20:891–897.

Vincent, J.M. 1970. A manual for the practical study of root-nodule bacteria, p. 176. *In* IBP Handbook No.15. Blackwell Scientific Publications, Oxford, U.K.

Young, J.P.W. 1992. Phylogenetic classification of nitrogen-fixing organisms, p. 43–86. *In* G. Stacey and H.J. Evans (eds.) Biological nitrogen fixation. Chapman and Hall, New York.

Zehr, J.P., B.D. Jenkins, S.M. Short, and G.F. Steward. 2003. Nitrogenase gene diversity and microbial community structure: A cross-system comparison. Environ. Microbiol. 5:539–554.

27.2 Nitrogen Mineralization–Immobilization Turnover

Jeanette M. Norton

Joshua P. Schimel

27.2.1 Introduction

The N cycle is complex and seemingly becoming more so with new processes being discovered regularly, but the core of the cycle remains the transformations between organic and inorganic nitrogen (N) (Schimel and Bennett, 2004). These transformations regulate critical ecosystem processes including plant growth, trace gas emissions, and NO_3^- leaching into ground and surface waters. They impact either directly or indirectly almost all the processes involved in the N cycle. Research focused on a wide variety of topics is needed to adequately assess the dynamics of the transformation of N between organic and inorganic forms. However, the best approach to understanding these processes is not always obvious and will vary depending on the specific study objectives.

Mineralization is the general term for the conversion of organic N to inorganic N as either ammonium (NH_4^+) or nitrite/nitrate (NO_2^-/NO_3^-), ammonification is the conversion of organic N to the NH_4^+ form, while nitrification (see Section 27.3) is the oxidation of reduced N (NH_4^+ or organic N) to NO_2^-/NO_3^-. Immobilization is the assimilation of inorganic N to organic N generally mediated by microorganisms. Mineralization–immobilization turnover (MIT) refers to the combined transformations between organic and inorganic N that accompanies the growth and death of the soil biota. To what extent MIT in soils necessitates passage through the NH_4^+ pool remains a subject of debate (Schimel and Bennett, 2004), as both plants and microorganisms can take up and use small N-containing compounds (i.e., amino acids, amines, etc.) without having to release the N as NH_4^+. An overview of these soil N cycle transformations is given in Figure 27.3.

27.2.2 Biochemistry and Physiology of Mineralization

As microorganisms decompose organic matter, including the remains of plants, animals, and microorganisms, they use as much

of the N as they need, while they release the remainder as smaller molecules or NH_4^+. This means that N mineralization is not a single physiological process, but rather, the aggregation of multiple physiological processes that result in the production of N waste materials (Ladd and Jackson, 1982).

Mineralization of organic N is basically a sequence of enzymatic reactions. The first step in this process is often to break down the large polymers that contain the bulk of N in the biota. This depolymerization is carried out by extracellular enzymes (primarily microbial in origin) including the proteinases, chitinases, kinases, amidases, and amidohydrolases (Ladd and Jackson, 1982). The substrates for depolymerization include a broad range of proteins, microbial cell wall constituents (amino sugars and their polymers, chitin, and peptidoglycan), and nucleic acids. Nitrogen is also released from the various N compounds found in soil organic matter including a considerable amount of heterocyclic and phenolic N, which are processed slowly in most soils.

Exoenzymes decrease the size and complexity of organic N molecules facilitating their further decomposition or assimilation; this depolymerization is often the rate-limiting step in the mineralization process (Kemmitt et al., 2008; Wallenstein and Weintraub, 2008; Jones et al., 2009). Some of the specific enzymes, which are involved in mineralization and their reactions, are summarized in Table 27.1. Since proteins and peptides are major sources of mineralizable N, the enzymes that can degrade these are produced abundantly and in wide diversity by microorganisms. Two examples of proteinases are trypsin, a specific enzyme that cleaves peptides at lysine and arginine residues, and subtilisin, an enzyme with broad specificity for peptide bonds that hydrolyzes a number of peptide amides. Fungi produce many proteolytic enzymes that may be particularly important for N acquisition during growth on woody substrates (Kudryavtseva et al., 2008).

The release of N from nonpeptide C–N bonds in amino acids and urea is mediated by amidohydrolases, dehydrogenases such as glutamate dehydrogenase (GDH) and transaminases. Peptidoglycan hydrolase acts on the link between *N*-acetylmuramoyl residues and L-amino acid residues in cell wall glycopeptides, thus playing a central role in the decomposition and turnover of microbial biomass. The polymers of amino sugars that are important components of bacterial and fungal cell walls are hydrolyzed by glycosidases. Other amidohydrolases, which are considered to be important for N mineralization in soils, include L-asparaginase, L-glutaminase, and amidases. Chitinase acts upon chitin and chitodextrins, major
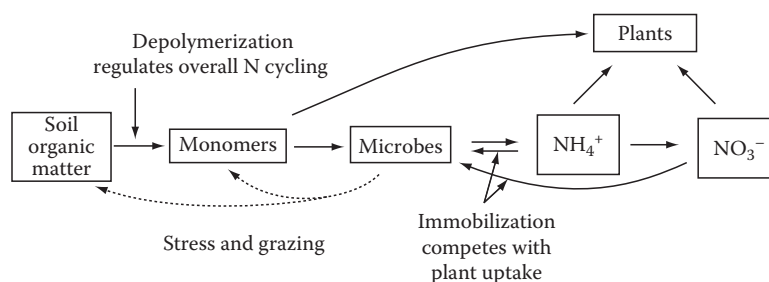
FIGURE 27.3 Overview of soil N cycle with emphasis on nitrogen mineralization–immobilization turnover. (From Schimel, J.P., and J. Bennett. 2004. Nitrogen mineralization: Challenges of a changing paradigm. Ecology 85:591–602.)

TABLE 27.1 Representative Enzymes Involved in N Mineralization in Soil Microorganisms

Enzyme	EC Designation	Action
Proteinases/peptidases	3.4	Hydrolyze peptide bonds
Trypsin	3.4.21.4	Hydrolyzes at Arg, Lys
Chymotrypsin	3.4.21.1	Hydrolyzes at aromatic amino acids
Subtilisin (microbial alkaline proteinase)	3.4.21.14	Alkaline proteinase
Carboxyl proteinases	3.4.23	Acid proteinases similar to pepsin or rennin
Microbial metalloproteinases	3.4.24	Neutral proteinases containing Zn
Amidohydrolases	3.5.1	Hydrolysis C–N bond in linear amides
L-Asparaginase	3.5.1.1	Hydrolysis of L-asparagine
L-Glutaminase	3.5.1.2	Hydrolysis of L-glutamine
Amidase	3.5.1.4	Hydrolysis of aliphatic amides
Urease	3.5.1.5	Hydrolysis of urea to CO_2 and NH_3
Peptidoglycan hydrolase	3.5.1.28	Hydrolysis of peptidoglycan linkages
Amidinohydrolases	3.5.3	Hydrolysis C–N bond of linear amidines
Arginase	3.5.3.1	Hydrolysis of arginine to ornithine and urea
Transaminases	2.6.1	Transfers amine to α-ketoglutarate
Dehydrogenases that deaminate	1.4	
Glutamate dehydrogenase	1.4.1.3	Deaminates glutamate
Glycosidases	3.2	Important in hydrolysis of amino sugars and their polymers
Chitinase	3.2.1.14	Hydrolyze chitin linkages
Muramidase (lysozyme)	3.2.1.17	Hydrolyze muropeptide (peptidoglycan) or mucopolysaccharide
Nucleases	3.1.11-16	Degrade DNA and RNA
DNase	3.1.11	Exodeoxyribonuclease
RNase	3.1.13	Exoribonuclease

Source: Chang, A., M. Scheer, A. Grote, I. Schomburg, and D. Schomburg. 2009. BRENDA, AMENDA and FRENDA the enzyme information system: New content and tools in 2009. Nucleic Acids Res. 37:D588–D592.

compounds in fungi, exoskeleton of insects, and crustaceans (Alef and Nannipieri, 1995). The observation that free nucleic acids are degraded rapidly substantiates the ubiquitous presence of the nucleases in the soil environment (Pietramellara et al., 2009).

There are two approaches to measuring the activity of specific extracellular enzymes. One is to inhibit microbial activity and then to measure the production of the enzymatic products—this is sometimes considered an "actual" enzyme rate, although there are serious constraints on the approach. The other is to add specific substrates and to measure the reaction rate—this is usually described as an enzyme potential. As with most soil assays, they each have their own strengths and limitations and neither is perfect for all applications (Burns, 1982; Nannipieri et al., 2002; Wallenstein and Weintraub, 2008).

To measure "actual" protease activity, microbial amino acid uptake is inhibited, usually by toluene, and then the rate of amino acids accumulation is measured (Tabatabai, 1994). That rate is considered the protease activity. The strength of this approach is that it does not require an added substrate, and so there is some hope that the measured rate actually does reflect the rate of proteolysis that occurs in the soil. However, the use of toluene has potential to cause artifacts (Nannipieri et al., 2002), inhibition may not be complete, and amino acids may adsorb to clays and organic matter, which would suggest that this might be a minimum, rather than an actual, proteolysis rate. Additionally, adding a toxin to soil might release free amino acids from the soil biomass, which would elevate the apparent rate. This latter issue would likely be a particular problem in dry soils as bacteria accumulate amino acids as a mechanism to osmoregulate and retain water under dry conditions (Kunte, 2006). This method is also subject to variability and error from free soil amino acids, extraction problems, etc.

The alternative approach to measure proteolysis is to add a known substrate for reaction by soil enzymes. The simplest version of this approach is to add saturating levels of casein or some other protein to a standard "actual" proteolysis assay to determine a potential rate. This approach allows you to evaluate what fraction of the proteolysis potential is actually being met (Weintraub and Schimel, 2005). Protease assays based on the addition of defined low-molecular-weight substrates (i.e., *N*-benzoylargininamide) and product measurement are also commonly used (Ladd and Butler, 1972; Rejsek et al., 2008). A contrasting approach is to add a defined substrate that when reacted produces a colored or fluorescent product (Marx et al., 2001; Saiya-Cork et al., 2002). These assays measure the enzymatic potential or V_{max} under the conditions chosen, not necessarily the process rate in situ. Details for enzyme assays of L-asparaginase, L-glutaminase, amidase, and urease in soil are given in Tabatabai (1994). The arginine ammonification potential has also been used as a laboratory indicator of a soil's ability to process organic N and as a general indicator of the size of the soil microbial biomass (Alef and Kleiner, 1986; Fuller and Scow, 1996). The release of NH_4^+ from arginine is measured in a shaken slurry short-term incubation. This enzyme assay for argininase may be useful as an index for gross mineralization (Bonde et al., 2001). The activity of urease in soil has

been widely studied because urea is a major N fertilizer (most important N fertilizer in world agriculture; Glibert et al., 2006) and because of the substantial inputs of urea from animal wastes. While it is clear that the release of amino acids and other forms of dissolved organic nitrogen (DON) is central to soil organic N turnover; there is no overall consensus on the enzymes or activities that are the best indicators of this depolymerization dynamic.

27.2.3 Biochemistry and Physiology of Immobilization

The incorporation of N into the microbial biomass and soil organic matter occurs via numerous enzymatic and abiotic pathways. This section focuses on the biotic reactions responsible for N assimilation or immobilization. The utilization of soil N includes various steps of extracellular depolymerization or hydrolysis, cell uptake, deamination, and entry into metabolic pathways. Microbes will assimilate both organic and inorganic N sources. Generally, the preferred inorganic N source for assimilation by most bacteria and fungi is NH_3/NH_4^+ although NO_3^- is also used under appropriate conditions (Marzluf, 1997; Myrold and Posavatz, 2007). The uncharged NH_3 may enter microbial cells by rapid diffusion across cytoplasmic membranes. However, specialized ammonia/ammonium transport proteins are widely distributed in archaea, bacteria, and fungi (Merrick and Edwards, 1995; Von Wirén and Merrick, 2004); and genes encoding putative ammonium transport functions are found in many environmental samples (Ogilvie et al., 2008). Intracellular NH_4^+ assimilation by soil microorganisms is predominantly accomplished through the synthesis of glutamate and/or glutamine via two enzymatic pathways: The GS–GOGAT system comprises glutamine synthetase/glutamate synthase (formerly glutamine:2-oxoglutarate aminotransferase hence GOGAT) and GDH pathways (Figure 27.4). The amination of aspartate by ammonium to form asparagine by asparagine synthetase is another possible pathway for assimilation (Hopkins et al., 1995; Barton, 2005). The glutamate and glutamine formed by these pathways serve as central N donors for transamination reactions yielding the amino acids and nucleotides that are the building blocks for proteins and nucleic acids. At the lower NH_4^+ concentrations typical of most soils, the GS–GOGAT system is operative (Schimel and Firestone, 1989; Landi et al., 1995; Paul and Clark, 1996) while GDH pathway functions to immobilizes N under relatively high NH_4^+ concentrations (i.e., >1 mM). The higher affinity GS–GOGAT system requires the input of energy in the form of ATP (White, 2007).

GS plays a central role in the uptake and regulation of N assimilation in soil microorganisms (McCarty and Bremner, 1992a, 1992b; McCarty, 1995). First characterized in the enteric bacteria, the common form of GS (GSI) is composed of 12 identical 55 kDa subunits and is highly conserved in a wide diversity of bacteria. A distinct form of GS (GSII) has been found in *Rhizobium* and *Agrobacterium* (Merrick and Edwards, 1995; Forchhammer, 2007). The NH_4^+ incorporating activity of GS is controlled by covalent modification (adenylation) of the individual subunits, resulting in a range of activity states for

the enzyme. The adenylation of the subunits is controlled by a complex regulatory cascade that senses the intracellular ratio of glutamine to 2-ketoglutarate (Merrick and Edwards, 1995), so that under N sufficiency the enzyme becomes progressively less active. Reviews of global nitrogen regulation (Ntr systems) and N transport in bacteria are given by Merrick and Edwards (1995) and Forchhammer (2007).

GS activity is involved in the regulation of several aspects of N metabolism in soil microbial communities including urease activity and assimilatory nitrate reduction (ANR) (McCarty, 1995). Both urease and ANR are normally repressed in the presence of NH_3/NH_4^+, the preferred N source. McCarty (1995) used an inhibitor of GS, L-methionine sulfoximine (MSX) to show that the production of L-glutamine by GS activity represses these alternative pathways for N utilization. GDH can function bidirectionally to immobilize or mineralize NH_4^+ (Figure 27.4). The direction of catalysis is controlled by N availability, GDH typically functions to immobilize N only under high NH_3 availability due to the relatively high K_m (1 mM) for this enzyme. In *Escherichia coli*, GDH is a hexamer of identical 50 kDa subunits that are encoded by the *gdhA* gene (Merrick and Edwards, 1995).

Nitrate may be immobilized directly by both bacteria and fungi; this process is known as ANR (Paul and Clark, 1996; Myrold and Posavatz, 2007). Energy is required for the transport of NO_3^- across the cell membrane prior to reduction to NH_4^+ by the ANR system. The enzymes responsible for reduction are assimilatory nitrate reductase and assimilatory nitrite reductase (Figure 27.4c). The synthesis of the ANR enzymes, nitrate and nitrite reductases, is regulated by the global Ntr system and inducible by nitrate (Merrick and Edwards, 1995) but repressed by NH_4^+ (Rice and Tiedje, 1989; Recous et al., 1990; McCarty and Bremner, 1992b) or glutamine (McCarty, 1995).

Investigations in soils have shown that both direct assimilation of amino acids (i.e., leucine and glycine) and rapid turnover with subsequent NH_4^+ assimilation can occur in parallel (Barak et al., 1990; Barraclough, 1997; Bardgett et al., 2003; Garnier et al., 2003; Kemmitt et al., 2008). The diverse mechanisms and controls over DON assimilation into the soil microbial biomass is a subject for current investigation (Jones et al., 2004; Schimel and Bennett, 2004; Kemmitt et al., 2008; Manzoni et al., 2008; Myrold and Bottomley, 2008).

27.2.4 Diversity of Enzymatic Functions

The most common ways of studying MIT processes have been through enzyme assays, quantifying nutrient pools, and measuring transformation rates (Kelly, 2003). Such studies have increased our understanding of the role of microbes in regulating these processes but do not give a complete view of the complexity of microbial contributions to soil function (Zak et al., 2006). Specific enzyme activities (such as urease, glutaminase, and protease activity) do not identify the microbial species directly involved in the measured process, leaving the link between the composition of the microbial community and the production of key enzymes poorly understood (Nannipieri et al., 2003; Colloff et al., 2008; Wallenstein
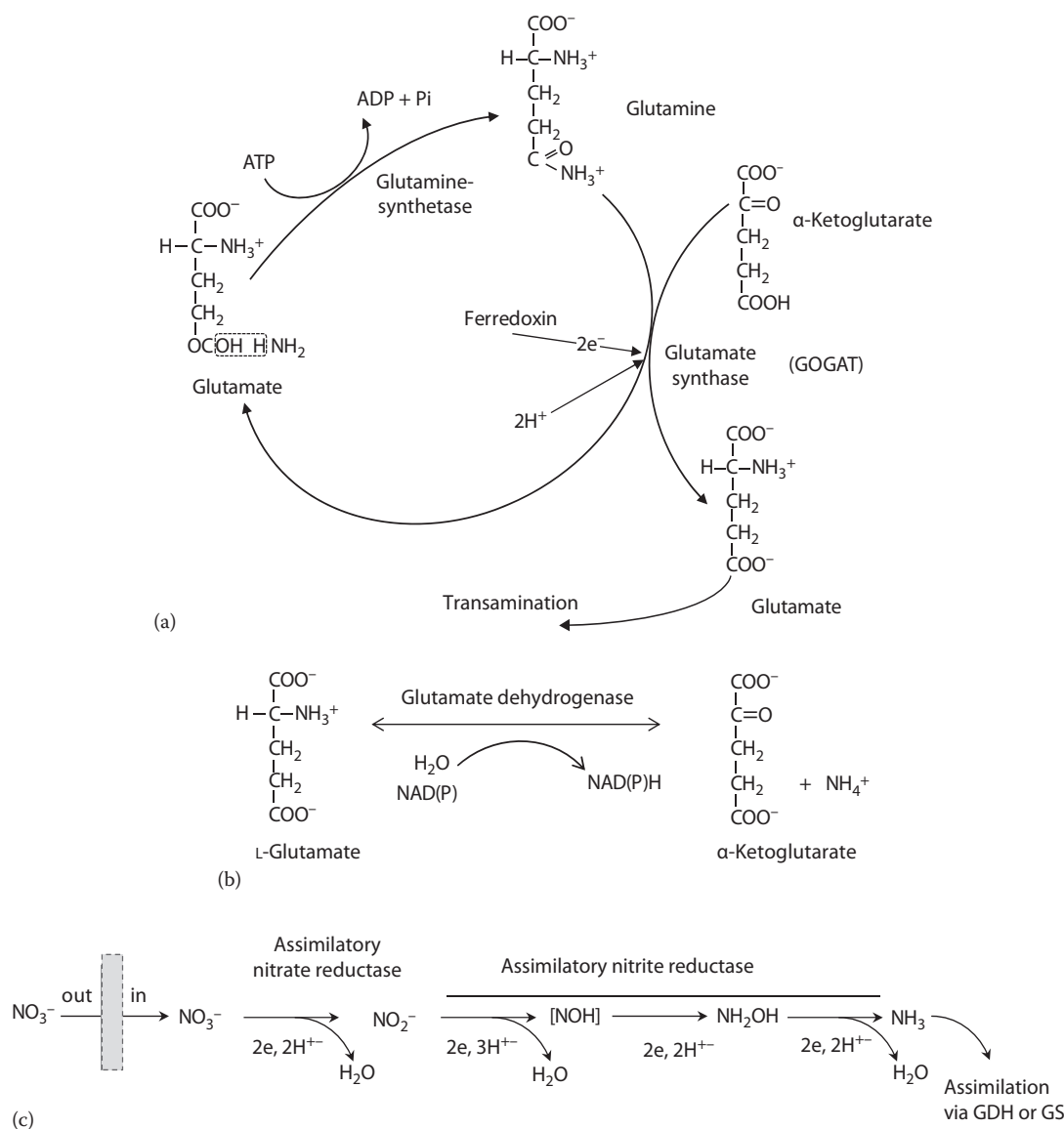
FIGURE 27.4 Major nitrogen assimilation pathways in bacteria. (a) The high affinity (<1 mM) assimilation of ammonia via the glutamine synthetase–glutamate synthase enzymes, (b) low affinity (>1 mM) reversible glutamate dehydrogenase system (From Paul, E.A., and F.E. Clark. 1996. Soil microbiology. Academic Press, San Diego, CA.), and (c) ANR via nitrate binding and transport, nitrate reduction (nitrate reductase), and nitrite reduction (nitrite reductase). (From White, D.C. 2007. The physiology and biochemistry of prokaryotes. 3rd edn. Oxford University Press, New York.)

and Weintraub, 2008). Metagenomic studies in soil environments and the availability of complete genome sequences for microorganisms have revealed a vast genetic diversity of enzyme-encoding genes for the enzymatic processes involved in MIT. The BRENDA searchable database (Chang et al., 2009) is an extensive resource for linking enzymes to their functions, protein sequences, and literature citations. Molecular tools are being developed targeting some of the key enzymes in MIT. For example, extracellular peptidases from proteolytic bacteria were the targets for a study detecting genes in isolates and from soil DNA (Bach et al., 1999, 2001; Bach and Munch, 2000). PCR primers and probes were developed for subsets of the neutral and alkaline metallopeptidases and subtilisin peptidases (Bach et al., 2001). Recent progress in characterizing the diversity of the urease-encoding genes and peptides in

important soil bacteria has allowed these genes to be amplified from DNA extracted from the soil and groundwater (Koper et al., 2004; Gresham et al., 2007). The diversity of the serine proteases (i.e., trypsin) in fungi has been examined for a broad range of cultured fungi (Hu and Leger, 2004) and this analysis may form the basis for targeting subsets of genes encoding fungal enzymes involved in depolymerization. Functional gene microarrays such as the GeoChip, contain >24,000 oligonucleotide probes covering >10,000 genes in >150 functional groups involved in nitrogen, carbon, and sulfur cycling, metal reduction and resistance, and organic contaminant degradation (He et al., 2007). Over 1400 of these were related to N mineralization functions. The most recent version of this microarray contains over 70,000 gene variants in more than 400 functional categories (Zhou et al., 2010) with comparatively

more probes related to MIT processes. Understanding the diversity of organisms responsible for any specific MIT process active in the environment will require further ecophysiological investigations to complement these emerging molecular tools.

27.2.5 Substrate Quality Effects on Mineralization–Immobilization Turnover

Substrates for decomposition vary in their chemical constitution and complexity. These differences affect the rate at which microorganisms process the material and the amount of N that is required or released during microbial growth on the substrate. The substrates, which are decomposed, include both materials that are inputs to the soil such as plant debris, animal manure, and industrial wastes and those that are internal to the soil system such as soil organic matter and microbial decay products. The availability of the substrate for decomposition, the efficiency with which microbes assimilate organic carbon into biomass (substrate use efficiency or yield coefficient), and the C/N ratio of the microbial biomass determine the requirement for N during substrate decomposition (Janssen, 1996; Myrold and Bottomley, 2008). If excess N is present in the substrate relative to the microbial demand, then net N mineralization will result, conversely when the N required by microorganisms is greater than that present in the substrate then microorganisms will immobilize inorganic N from the soil solution or slow their growth and have a widened C/N ratio until N becomes available. While the actual situation during decomposition is extremely complex with numerous fractions of various composition and availability, there is a history of the use of simple characteristics as indicators for substrate quality and in estimation of the N demand during decomposition (Janssen, 1996; Myrold and Bottomley, 2008). Table 27.2 gives some representative C/N ratios for various organic materials and Figure 27.5 shows the relationship between the critical C/N ratio, microbial C/N ratio, and the

FIGURE 27.5 Model of the critical C/N ratio of substrates that results in no net mineralization of nitrogen with different microbial C yield coefficients and C/N ratio of the microbial biomass produced. (After Myrold, D.D. and Bottomley, P.J., Nitrogen mineralization and immobilization, in: W. Raun and J.S. Schepers (eds.), *Nitrogen in Agricultural Soils*, ASA, Madison, WI, pp. 157–172, 2008.)

microbial yield coefficient. Generally, substrates with C/N ratios less than 20/1 or N contents greater than 1.5% N result in net N mineralization while those higher than 30/1 or with N contents less than 1.5% tend to immobilize soil inorganic N. This oversimplification does not represent complex substrates with large proportions of resistant materials such as lignin; these materials may mineralize N at C/N ratios up to 50/1 due to the differential decay rates for distinct fractions of the material (Paul and Clark, 1996; Myrold and Bottomley, 2008). As decomposition of a substrate proceeds, the C/N ratio of the material remaining in the soil (a combination of the remaining substrate and the microbial products) will decrease and N will eventually be mineralized.

Models of C and N turnover in soil have often used the lignin to N ratio of the substrate as an indicator of substrate quality (Melillo et al., 1982; Parton et al., 1987, 1988; Cabrera et al., 2008). Substances with high lignin to N ratios are considered to have more structural materials, which degrade more slowly, with lower microbial efficiencies (Paustian et al., 1992). In contrast, some recent studies did not find the lignin to N ratio to be correlated to decomposition or N mineralization from soils amended with a variety of plant materials (Jensen et al., 2005; Bruun et al., 2006). Particularly when materials contain substantial amounts of other complex polyphenolics, such as tannins, the lignin:N ratio may be inadequate as an indicator of decomposition and mineralization patterns. Simulation models of MIT are further discussed in Section 27.2.7.

One important substrate for mineralization is the microbial biomass itself. Predators and grazers (protozoa, nematodes, etc.) may consume large amounts of bacteria and fungi with relatively low C/N ratios and therefore these consumers subsequently excrete substantial amounts of N (Clarholm, 1985; Coleman et al., 2004). Protozoa and microphagous nematodes are important grazers of microflora in the belowground ecosystems and generally increase microbial turnover, promote N mineralization, and may change the composition of the microbial community (Bonkowski et al., 2000;

TABLE 27.2 Approximate C/N Ratios in Various Organic Materials

Organic Material	C/N Ratio
Soil microorganisms	8/1
Soil organic matter	10/1
Sewage sludge	9/1
Alfalfa residues	16/1
Farmyard manure	20/1
Corn stover	60/1
Grain straw	80/1
Oak	200/1
Pine	300/1
Crude oil	400/1
Conifer sawdust	625/1

Source: Hyvonen, R., G.I. Agren, and O. Andren. 1996. Modelling long-term: Carbon and nitrogen dynamics in an arable soil receiving organic matter. Ecol. Appl. 6:1345–1354; Havlin, J.L., J.D. Beaton, S.L. Tisdale, and W.L. Nelson. 1999. Soil fertility and fertilizers. 6th edn. Prentice Hall, Upper Saddle River, NJ.

Raynaud et al., 2006). The complex belowground food webs and the importance of the microbial loop are discussed in Chapter 26.

27.2.6 Process Rates and Kinetics

The rate and extent of MIT reactions determine the availability of N to plants and the potential for N losses for many ecosystems. However, because the amount of inorganic N in the soil is the result of several opposing productive and consumptive processes, our ability to predict its availability remains challenging.

Generally, mineralization is the conversion of organic N to inorganic N while immobilization is the reverse process of assimilation of inorganic N into the biotic biomass and hence organic N. However, in actual use, these terms have several different meanings that are related to how they are determined. The net mineralization rate is the rate of inorganic N accumulation and is equal to the rate of conversion of organic N to NH_4^+/NH_3 and NO_3^- minus their consumption. Net mineralization refers to the case where the inorganic N pool is increasing with time because gross rates of mineralization are higher than rates of combined immobilization plus losses. The gross mineralization rate is the actual rate of conversion of organic N to NH_4^+/NH_3. The gross mineralization rate is also equal to the rate of NH_4^+ accumulation plus the rate of NH_4^+ consumption plus losses. Determining gross rates requires the use of isotopes of N to separate simultaneous production and consumption. The potential mineralization rate is the net rate of mineralization in a soil system with plant uptake and leaching prevented (although the potential rate is operationally defined by the actual method chosen for its determination). Previously, many investigators assumed that measured net mineralization rates in the absence of plant uptake and leaching were adequate indicators of the importance and rate of mineralization in the environment. This assumption neglected the potential role of inorganic N assimilation by microorganisms and denitrification. Significant rates of inorganic N assimilation are often measured simultaneously with net mineralization (Powlson and Barraclough, 1993; Schimel and Bennett, 2004). Net mineralization also neglects the turnover and uptake of simple organic compounds by both plants and microbes, processes that may supply much of the biotic N demand in some ecosystems (Schimel and Bennett, 2004).

Immobilization is defined as the assimilation of inorganic N (NH_4^+/NO_3^-) into the soil organic matter (i.e., microbial biomass). Net immobilization refers to the case where the inorganic N pool is decreasing due to higher gross rates of immobilization than gross rates of mineralization. Net immobilization may also encompass some abiotic processes, such as NH_4^+ fixation into clays, although these rarely appear to be a large fraction of total immobilization. The gross immobilization rate is the actual rate of conversion of inorganic N to organic N. Gross NH_4^+ immobilization is only a portion of the total NH_4^+ consumption that also includes nitrification and plant root uptake. Inorganic N consumption includes NH_4^+ and NO_3^- immobilization by heterotrophic microorganisms, plant root uptake, nitrate leaching, denitrification, ammonia volatilization, and dissimilatory nitrate reduction to ammonia (DNRA) as well as abiotic reactions of NH_3 and NO_2^- with organic matter.

The distinction between total consumption of inorganic N, NH_4^+ consumption and immobilization becomes particularly important in the determination of gross immobilization rates.

27.2.6.1 Measurements of Mineralization and Immobilization Rates

Numerous measurements of soil mineralization rates have been made in ecosystems throughout the globe, the majority of measurements are net rates; fewer estimates are available for gross mineralization/immobilization rates on undisturbed soil cores (Murphy et al., 2003; Booth et al., 2005). The basic principles and application of N isotopes for determination of gross mineralization and immobilization rates in soils were developed several decades ago (Kirkham and Bartholomew, 1954, 1955; Jansson, 1958) but innovations in mass spectrometry including direct combustion isotope ratio mass spectrometry (Hauck et al., 1994; Mulvaney, 2008) made the use of ^{15}N for estimating gross rates more accessible. Determining NH_4^+, NO_2^-, and NO_3^- concentrations in soils (Bundy and Meisinger, 1994; Hart et al., 1994b) is the starting point for almost all the rate measurements described as follows. Soil should be extracted as soon as possible after sampling as inorganic N levels change rapidly in disturbed samples. The following discussion is focused on methods for determining mineralization and immobilization rates, starting from the simpler laboratory measurements of net mineralization progressing to the field measurements of gross mineralization and immobilization rates. Investigators should consider carefully the justification for using the more complex and expensive determination of gross rates of N transformations unless the goal is to understand the controls and activity of soil microorganisms (Murphy et al., 2003).

27.2.6.1.1 *Laboratory Methods for Net and Potential Mineralization Rates*

Laboratory incubation experiments have a long history of being used to examine net mineralization in the absence of plant uptake and under conditions that minimize gaseous and leaching losses (Standford and Smith, 1972). These experiments give an estimate of the mineralizable N pool and are most useful as indices of the N supplying potential of closely related soils or a soil that has received various amendments or treatments (Griffin, 2008). The basic premise to all the laboratory incubation methods is that initial soil inorganic N determinations are made, the soil is incubated under a defined temperature, and moisture regime and the amount of inorganic N produced during a specified time period are determined. Typically, the soils are sieved and homogenized before incubation; however, intact cores may be more easily related to field mineralization rates (Cabrera and Kissel, 1988). The artificial conditions of constant temperature and controlled moisture and the reliance on net rates minimize the relation of mineralization potential measurements to gross mineralization rates in the field (Stark, 2000). Many variations of the laboratory incubation method are described in the literature (Hart et al., 1994b), the various methods may give contrasting results with different soils (Benbi and Richter, 2002). Interpretations may involve modeling of the time course of released inorganic N to

determine the pool size of the mineralizable or labile N according to various models (most typically, first-order rates). Recent analyses using simulation modeling more realistically represents mineralizable N as several pools with different mineralization rates (Ma and Shaffer, 2001; Benbi and Richter, 2002).

27.2.6.1.2 Field Measurements of Net Mineralization

Several methods are available for determining net mineralization and nitrification in the field; all include a barrier to prevent root uptake of NH_4^+ and NO_3^- and to prevent or capture the NO_3^- in the mass flow of water. The buried bag and soil core methods are described as suitable procedures for field estimates of net mineralization (Hart et al., 1994b). Typically intact soil cores are used to minimize disturbance artifacts (i.e., increased mineralization) caused by sieving and mixing, but this approach increases the number of replicates needed to encompass site variability (Hart et al., 1994b). Both the buried bag and the solid core method may alter the moisture status of the soil inside the container as compared to the external soil (Griffin, 2008). In solid core methods, soil is incubated intact inside PVC or steel cylinders either with the top covered to prevent water entry or opened but with the bottom of the core containing a ring filled with mixed bed ion exchange resin to catch leachate (Hart et al., 1994b). Although net mineralization rates have been related to N loss from forest ecosystems and plant available N in agroecosystems, they may not be well correlated to gross rates of N processes (Davidson et al., 1992; Hart et al., 1994a).

27.2.6.1.3 Gross Rates of Mineralization and Immobilization: The Use of Isotopes

The measurement of actual or gross rates of mineralization and immobilization requires the use of isotopes of N, the stable isotope ^{15}N is used almost exclusively because the radioactive isotope, ^{13}N has an extremely short half-life (~10 min) that is not conducive for ecological studies. In general, ^{15}N can be used in two contrasting approaches to examine MIT: (1) tracer methods in which ^{15}N is added as the substrate for the process of interest and (2) pool dilution methods in which ^{15}N is added to the product pool of the transformation of interest. Mineralization rates have been assessed by tracer techniques in which ^{15}N-labeled plant material or other ^{15}N organic substrates are added to the source pool and the fate of the added ^{15}N is monitored. In the pool dilution approach, ^{15}N is added and the dilution of its enrichment by the production of $^{14}NH_4^+$ by mineralization is monitored over time. Immobilization rates may be estimated by the tracer method or by a combination of the isotope dilution and tracer techniques. The tracer approach for mineralization involves the preparing and using ^{15}N-labeled organic substrates and allows the investigator to assess the mineralization potential of the material and follow the fate of the added ^{15}N. The data from these studies are particularly useful for parameterizing simulation models of MIT. Methods for labeling plants and the microbial biomass with ^{15}N are described by Wolf et al. (1994) and the preparation of organic materials for analysis is described (Hauck et al., 1994). The tracer approach for estimating immobilization rates involves adding $^{15}NH_4^+$ or $^{15}NO_3^-$ and then measuring the appearance of ^{15}N in the sink pool, usually the microbial biomass or soil organic matter (Powlson and Barraclough, 1993). In N-limited systems, the addition of substrate may stimulate microbial assimilation and therefore overestimate immobilization rates (Hart et al., 1994b). The tracer approach is particularly useful for estimating immobilization rates in the presence of several consumptive processes (i.e., plant uptake, nitrification, ammonia volatilization) and is often used in combination with the pool dilution approach described below to answer mechanistic questions on MIT (Schimel et al., 1989; Norton and Firestone, 1996)—when you add $^{15}NH_4^+$ to the soil for a pool dilution study, you can also use it as a tracer for measuring flows into NO_3^-, microbial biomass, and plants (Figure 27.6).
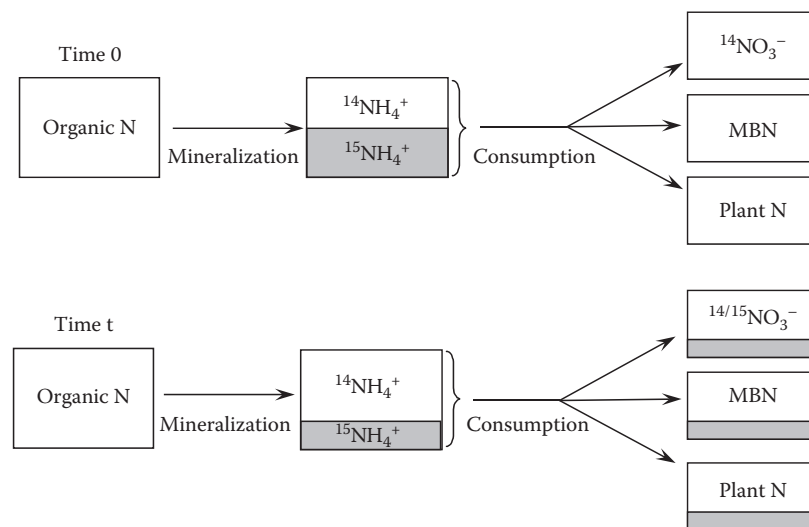
FIGURE 27.6 Combined pool dilution and tracer approach for estimation of mineralization and immobilization rate. At the beginning of incubation (time 0) $^{15}NH_4^+$ is added and initial pool sizes and enrichments are determined. At end of the incubation (time t) the final pool sizes and enrichments are determined. The mineralization and NH_4^+-consumption rates are determined by pool dilution calculation, NH_4^+ immobilization rate determined by tracer analysis. Assumptions include that there is no recycling within the time of the incubation.

The pool dilution approach for estimating mineralization and NH_4^+ consumption rates has been used for numerous laboratory and field studies (Murphy et al., 2003; Booth et al., 2005; Hooker and Stark, 2008). The pool dilution approach is ideally suited for measurements of gross mineralization rates, but the estimates of NH_4^+ immobilization are less reliable because adding the $^{15}NH_4^+$ substrate may stimulate uptake and there are several additional processes that consume NH_4^+. Nitrification, frequently the largest consumptive process, may be determined from parallel NO_3^- pool dilution experiments or nitrification may be blocked using acetylene. Plant uptake can be calculated from the ^{15}N content of the plant tissue or blocked by preventing root uptake with a barrier method. Ammonia volatilization, which is important only in alkaline soils, can be measured by including an acid trap for the volatilized NH_3 in the experiment or calculated by difference. It is recommended that total recoveries of ^{15}N in the plant–soil system be determined to detect any unexplained loss of ^{15}N. The specific recommendations for sampling and techniques for field pool dilution experiments in intact cores without plants are discussed and background and calculations for estimating mineralization and assimilation in the plant–soil system are given (Davidson et al., 1991; Hart et al., 1994b; Stark, 2000; Murphy et al., 2003). Representative rates of NH_4^+ production and consumption determined by pool dilution from agricultural and wildland ecosystems have been reviewed (Booth et al., 2005).

27.2.7 Simulation Models of Mineralization–Immobilization Turnover

Kinetic and simulation models of nitrogen turnover in the plant–soil system can be useful for understanding the interactions of the multiple processes involved. Often the combination of experimental and modeling approaches provides a synthesis and identifies gaps in our knowledge. When models fail to adequately express the realities of the observed system this is often the first step toward understanding of a neglected process. An example is the identification of the importance of simultaneous microbial immobilization and mineralization for simulating short-term inorganic N dynamics and the role of soil heterogeneity (Manzoni et al., 2008). Nitrogen process models have become more crucial for predicting NO_3^- leaching after application of organic materials and mineral fertilizers; it is no longer sufficient to predict plant needs and economic returns, environmental impacts of excess N must be considered. Regional and global carbon models require information on N availability and the potential for N to limit carbon sequestration.

Existing models vary in the number of organic matter pools represented and their definition, the number of distinct pools involved in MIT, the role of the microbial biomass and its turnover, and whether nitrification and denitrification are considered. In general, most models link the cycling of N to the C cycle processes with C availability controlling microbial biomass growth and turnover (Paustian et al., 1992; Cabrera et al., 2008). Current N turnover models have been critically reviewed and compared for their treatment of the above factors. A general consensus is that improved submodels dealing with the maintenance and growth of the microbial biomass and particularly seasonal variation in the microbial biomass N are needed to adequately predict inorganic N dynamics (Ma and Shaffer, 2001; Cabrera et al., 2008). While our conceptual models and the computer simulations of MIT in soils may never fully converge, the process of quantitatively formulating the complex processes in the soil environment improves our predictive ability and identifies our misconceptions.

27.2.8 Concluding Remarks

The nitrogen cycle is the global biogeochemical cycle that humans have most dramatically perturbed—industrial N fixation is now roughly equivalent to total natural fixation (Vitousek et al., 1997; Schlesinger, 2009). This perturbation has resulted in eutrophication of waterways, dead zones in coastal waters, an enhanced greenhouse effect, and local smog production. Managing N cycling in the future to minimize these adverse impacts, while optimizing food, fuel, and fiber production is likely to require a more sophisticated understanding of the processes that impact N availability and transport. Mineralization is the most critical process, as it regulates plant uptake, nitrification, leaching, and trace gas emissions. Yet, MIT is inherently complex because it is a web of processes involving every organism in the soil. It couples the activity of extracellular enzymes, microbial metabolism and growth, food web dynamics, and the physical structuring of these processes within the soil matrix. Our current ability to effectively model this interaction remains crude and inadequate for many situations because there are many cases where the simplifying assumptions of biogeochemical models do not match the reality of the soil. We need new models that more effectively capture the complex dynamics of mineralization but do so in a way that does not require modeling all the specific components directly—such models would be excessively unwieldy. Understanding the role of functional redundancy in the enzymes and biota will be essential to our perception of system function. To reach this new level in modeling N mineralization, we need a more integrated understanding of how all the processes of mineralization interact to regulate the overall turnover of organic and inorganic N in soils. Nitrogen mineralization and immobilization process many times the net inputs and outputs of inorganic N from the soil system. Identification of the controlling factors on the individual processes (Booth et al., 2005), not their combined outcome, will be necessary to develop an improved mechanistic understanding of nitrogen flow through the soil ecosystem.

References

Alef, K., and D. Kleiner. 1986. Arginine ammonification, a simple method to estimate microbial activity potentials in soils. Soil Biol. Biochem. 18:233–235.

Alef, K., and P. Nannipieri. 1995. Methods in applied soil microbiology and biochemistry. Academic Press, London, U.K.

Bach, H.J., D. Errampalli, K.T. Leung, H. Lee, A. Hartmann, J.T. Trevors, and J.C. Munch. 1999. Specific detection of the gene for the extracellular neutral protease of *Bacillus cereus* by PCR and blot hybridization. Appl. Environ. Microbiol. 65:3226–3228.

Bach, H.J., A. Hartmann, M. Schloter, and J.C. Munch. 2001. PCR primers and functional probes for amplification and detection of bacterial genes for extracellular peptidases in single strains and in soil. J. Microbiol. Methods 44:173–182.

Bach, H.J., and J.C. Munch. 2000. Identification of bacterial sources of soil peptidases. Biol. Fertil. Soils 31:219–224.

Barak, P., J.A.E. Molina, A. Hadas, and C.E. Clapp. 1990. Mineralization of amino-acids and evidence of direct assimilation of organic nitrogen. Soil Sci. Soc. Am. J. 54:769–774.

Bardgett, R.D., T.C. Streeter, and R. Bol. 2003. Soil microbes compete effectively with plants for organic-nitrogen inputs to temperate grasslands. Ecology 84:1277–1287.

Barraclough, D. 1997. The direct or MIT route for nitrogen immobilization: A N-15 mirror image study with leucine and glycine. Soil Biol. Biochem. 29:101–108.

Barton, L.L. 2005. Structural and functional relationships in prokaryotes. Springer Science + Business Media, Inc., New York.

Benbi, D.K., and J. Richter. 2002. A critical review of some approaches to modelling nitrogen mineralization. Biol. Fertil. Soils 35:168–183.

Bonde, T.A., T.H. Nielsen, M. Miller, and J. Sorensen. 2001. Arginine ammonification assay as a rapid index of gross N mineralization in agricultural soils. Biol. Fertil. Soils 34:179–184.

Bonkowski, M., W.X. Cheng, B.S. Griffiths, G. Alphei, and S. Scheu. 2000. Microbial-faunal interactions in the rhizosphere and effects on plant growth. Eur. J. Soil Biol. 36:135–147.

Booth, M.S., J.M. Stark, and E. Rastetter. 2005. Controls on nitrogen cycling in terrestrial ecosystems: A synthetic analysis of literature data. Ecol. Monogr. 75:139–157.

Bruun, S., J. Luxhoi, J. Magid, A. De Neergaard, and L.S. Jensen. 2006. A nitrogen mineralization model based on relationships for gross mineralization and immobilization. Soil Biol. Biochem. 38:2712–2721.

Bundy, L.G., and J.J. Meisinger. 1994. Nitrogen availability indices, p. 951–984. *In* R.W. Weaver, S. Angle, P. Bottomly, D. Bezdicek, S. Smith, A. Tabatabai, and A. Wollum (eds.) Methods of soil analysis. Part 2. Microbiological and biochemical properties. SSSA, Madison, WI.

Burns, R.G. 1982. Enzyme activity in soil: Location and a possible role in microbial ecology. Soil Biol. Biochem. 14:423–427.

Cabrera, M.L., and D.E. Kissel. 1988. Potentially mineralizable nitrogen in disturbed and undisturbed soil samples. Soil Sci. Soc. Am. J. 52:1010–1015.

Cabrera, M.L., J.A. Molina, and M. Vigil. 2008. Modeling the nitrogen cycle, p. 695–730. *In* J.S. Schepers and W.R. Raun (eds.) Nitrogen in agricultural systems. ASA, CSSA, SSSA, Madison, WI.

Chang, A., M. Scheer, A. Grote, I. Schomburg, and D. Schomburg. 2009. BRENDA, AMENDA and FRENDA the enzyme information system: New content and tools in 2009. Nucleic Acids Res. 37:D588–D592.

Clarholm, M. 1985. Interactions of bacteria, protozoa and plants leading to mineralization of soil-nitrogen. Soil Biol. Biochem. 17:181–187.

Coleman, D.C., D.A. Crossley, Jr., and P.F. Hendrix. 2004. Fundamentals of soil ecology. 2nd edn. Elsevier Academic Press, Boston, MA.

Colloff, M.J., S.A. Wakelin, D. Gomez, and S.L. Rogers. 2008. Detection of nitrogen cycle genes in soils for measuring the effects of changes in land use and management. Soil Biol. Biochem. 40:1637–1645.

Davidson, E.A., S.C. Hart, and M.K. Firestone. 1992. Internal cycling of nitrate in soils of a mature coniferous forest. Ecology 73:1148–1156.

Davidson, E.A., S.C. Hart, C.A. Shanks, and M.K. Firestone. 1991. Measuring gross nitrogen mineralization, immobilization and nitrification by N-15 isotopic pool dilution in intact soil cores. J. Soil Sci. 42:335–349.

Forchhammer, K. 2007. Glutamine signalling in bacteria. Front. Biosci. 12:358–370.

Fuller, M.E., and K.M. Scow. 1996. Effects of toluene on microbially-mediated processes involved in the soil nitrogen cycle. Microb. Ecol. 32:171–184.

Garnier, P., C. Neel, C. Aita, S. Recous, F. Lafolie, and B. Mary. 2003. Modelling carbon and nitrogen dynamics in a bare soil with and without straw incorporation. Eur. J. Soil Sci. 54:555–568.

Glibert, P.M., J. Harrison, C. Heil, and S. Seitzinger. 2006. Escalating worldwide use of urea—A global change contributing to coastal eutrophication. Biogeochemistry 77:441–463.

Gresham, T.L.T., P.P. Sheridan, M.E. Watwood, Y. Fujita, and F.S. Colwell. 2007. Design and validation of ureC-based primers for groundwater detection of urea-hydrolyzing bacteria. Geomicrobiol. J. 24:353–364.

Griffin, T.S. 2008. Nitrogen availability, p. 613–646. *In* J.S. Schepers and W.R. Raun (eds.) Nitrogen in agricultural systems. ASA, CSSA, SSSA, Madison, WI.

Hart, S.C., G.E. Nason, D.D. Myrold, and D.A. Perry. 1994a. Dynamics of gross nitrogen transformations in an old-growth forest-the carbon connection. Ecology 75:880–891.

Hart, S.C., J.M. Stark, E.A. Davidson, and M.K. Firestone. 1994b. Nitrogen mineralization, immobilization and nitrification, p. 985–1018. *In* R.W. Weaver, S. Angle, P. Bottomly, D. Bezdicek, S. Smith, A. Tabatabai, and A. Wollum (eds.) Methods of soil analysis. Part 2. Microbiological and biochemical properties. SSSA, Madison, WI.

Hauck, R., J.J. Meisinger, and C.S. Mulvaney. 1994. Practical considerations in the use of nitrogen tracers in agricultural and environmental research, p. 907–943. *In* R.W. Weaver, S. Angle, P. Bottomly, D. Bezdicek, S. Smith, A. Tabatabai, and A. Wollum (eds.) Methods of soil analysis. Part 2. Microbiological and biochemical properties. SSSA, Madison, WI.

Havlin, J.L., J.D. Beaton, S.L. Tisdale, and W.L. Nelson. 1999. Soil fertility and fertilizers. 6th edn. Prentice Hall, Upper Saddle River, NJ.
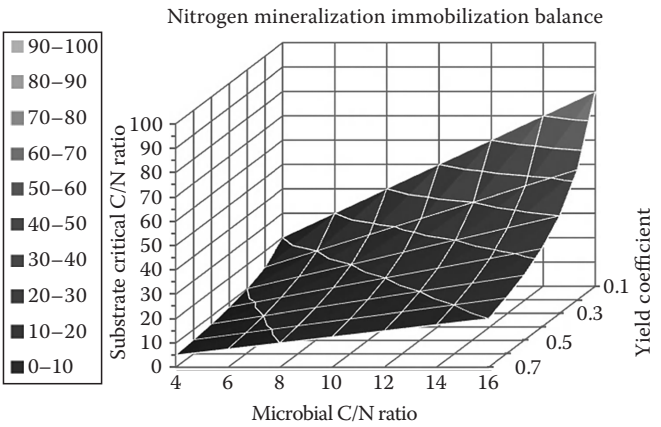
He, Z.L., T.J. Gentry, C.W. Schadt, L.Y. Wu, J. Liebich, S.C. Chong, Z.J. Huang et al. 2007. GeoChip: A comprehensive microarray for investigating biogeochemical, ecological and environmental processes. ISME J. 1:67–77.

Hooker, T.D., and J.M. Stark. 2008. Soil C and N cycling in three semiarid vegetation types: Response to an in situ pulse of plant detritus. Soil Biol. Biochem. 40:2678–2685.

Hopkins, D.W., L. Anderson, and S.E. Scott. 1995. N and C mineralization in soil amended with the N immobilization inhibitor, methionine sulfoximine. Soil Biol. Biochem. 27:377–379.

Hu, G., and R.J.S. Leger. 2004. A phylogenomic approach to reconstructing the diversification of serine proteases in fungi. J. Evol. Biol. 17:1204–1214.

Hyvonen, R., G.I. Agren, and O. Andren. 1996. Modelling long-term: Carbon and nitrogen dynamics in an arable soil receiving organic matter. Ecol. Appl. 6:1345–1354.

Janssen, B.H. 1996. Nitrogen mineralization in relation to C:N ratio and decomposability of organic materials. Plant Soil 181:39–45.

Jansson, S.L. 1958. Tracer studies on nitrogen transformations in soil with special attention to mineralisation–immobilization relationships. Ann. R. Agric. Coll. Sweden 24:101–361.

Jensen, L.S., T. Salo, F. Palmason, T.A. Breland, T.M. Henriksen, B. Stenberg, A. Pedersen, C. Lundstrom, and M. Esala. 2005. Influence of biochemical quality on C and N mineralisation from a broad variety of plant materials in soil. Plant Soil 273:307–326.

Jones, D.L., K. Kielland, F.L. Sinclair, R.A. Dahlgren, K.K. Newsham, J.F. Farrar, and D.V. Murphy. 2009. Soil organic nitrogen mineralization across a global latitudinal gradient. Global Biogeochem. Cycles 23:GB1016.

Jones, D.L., D. Shannon, D.V. Murphy, and J. Farrar. 2004. Role of dissolved organic nitrogen (DON) in soil N cycling in grassland soils. Soil Biol. Biochem. 36:749–756.

Kelly, J.J. 2003. Molecular techniques for the analysis of soil microbial processes: Functional gene analysis and the utility of DNA microarrays. Soil Sci. 168:597–605.

Kemmitt, S.J., D. Wright, D.V. Murphy, and D.L. Jones. 2008. Regulation of amino acid biodegradation in soil as affected by depth. Biol. Fertil. Soils 44:933–941.

Kirkham, D., and W.V. Bartholomew. 1954. Equations for following nutrient transformations in soil utilizing tracer data. Soil Sci. Soc. Am. Proc. 18:33–34.

Kirkham, D., and W.V. Bartholomew. 1955. Equations for following nutrient transformations in soil, utilizing tracer data: II. Soil Sci. Soc. Am. Proc. 19:189–192.

Koper, T.E., A.F. El-Sheikh, J.M. Norton, and M.G. Klotz. 2004. Urease-encoding genes in ammonia-oxidizing bacteria. Appl. Environ. Microbiol. 70:2342–2348.

Kudryavtseva, O.A., Y.E. Dunaevsky, O.V. Kamzolkina, and M.A. Belozersky. 2008. Fungal proteolytic enzymes: Features of the extracellular proteases of xylotrophic basidiomycetes. Microbiology 77:643–653.

Kunte, H.J. 2006. Osmoregulation in bacteria: Compatible solute accumulation and osmosensing. Environ. Chem. 3:94–99.

Ladd, J.N., and J.H.A. Butler. 1972. Short-term assays of soil proteolytic enzyme activities using proteins and dipeptide derivatives as substrates. Soil Biol. Biochem. 4:19–30.

Ladd, J.N., and R.B. Jackson. 1982. Biochemistry of ammonification, p. 173–227. *In* F.J. Stevenson (ed.) Nitrogen in agricultural soil. ASA, Madison, WI.

Landi, L., L. Badalucco, and P. Nannipieri. 1995. Changes in inorganic-N and CO_2 evolution in soil induce by L-methionine-sulfoximine. Soil Biol. Biochem. 27:1345–1351.

Ma, L., and M.J. Shaffer. 2001. A review of carbon and nitrogen cycling in nine U.S. soil nitrogen dynamic models, p. 55–102. *In* M.J. Shaffer (ed.) Modeling carbon and nitrogen dynamics for soil management. Lewis Publishers, Boca Raton, FL.

Manzoni, S., A. Porporato, and J.P. Schimel. 2008. Soil heterogeneity in lumped mineralization–immobilization models. Soil Biol. Biochem. 40:1137–1148.

Marx, M.C., M. Wood, and S.C. Jarvis. 2001. A microplate fluorimetric assay for the study of enzyme diversity in soils. Soil Biol. Biochem. 33:1633–1640.

Marzluf, G.A. 1997. Genetic regulation of nitrogen metabolism in the fungi. Microbiol. Mol. Biol. Rev. 61:17–32.

McCarty, G.W. 1995. The role of glutamine-synthetase in regulation of nitrogen metabolism within the soil microbial community. Plant Soil 170:141–147.

McCarty, G.W., and J.M. Bremner. 1992a. Inhibition of assimilatory nitrate reductase activity in soil by glutamine and ammonium analogs. Proc. Natl. Acad. Sci. USA 89:5834–5836.

McCarty, G.W., and J.M. Bremner. 1992b. Regulation of assimilatory nitrate reductase activity in soil by microbial assimilation of ammonium. Proc. Natl. Acad. Sci. USA 89:453–456.

Melillo, J.M., J.D. Aber, and J.F. Muratore. 1982. Nitrogen and lignin control of hardwood leaf litter decomposition dynamics. Ecology 63:621–626.

Merrick, M.J., and R.A. Edwards. 1995. Nitrogen control in bacteria. Microbiol. Rev. 59:604–622.

Mulvaney, R.L. 2008. Advances in methodology for research on nitrogen transformations in soils, p. 437–504. *In* J.S. Schepers and W.R. Raun (eds.) Nitrogen in agricultural systems. ASA, CSSA, SSSA, Madison, WI.

Murphy, D.V., S. Recous, E.A. Stockdale, I.R.P. Fillery, L.S. Jensen, D.J. Hatch, and K.W.T. Goulding. 2003. Gross nitrogen fluxes in soil: Theory, measurement and application of N-15 pool dilution techniques. Adv. Agron. 79:69–118.

Myrold, D.D., and P.J. Bottomley. 2008. Nitrogen mineralization and immobilization, p. 157–172. *In* W. Raun and J.S. Schepers (eds.) Nitrogen in agricultural soils. ASA, Madison, WI.

Myrold, D.D., and N.R. Posavatz. 2007. Potential importance of bacteria and fungi in nitrate assimilation in soil. Soil Biol. Biochem. 39:1737–1743.

Nannipieri, P., J. Ascher, M.T. Ceccherini, L. Landi, G. Pietramellara, and G. Renella. 2003. Microbial diversity and soil functions. Eur. J. Soil Sci. 54:655–670.

Nannipieri, P., E. Kandeler, and P. Ruggiero. 2002. Enzyme activities and microbiological and biochemical processes in soil, p. 1–33. *In* R.G. Burns and R.P. Dick (eds.) Enzymes in the environment. Marcel Dekker, New York.

Norton, J.M., and M.K. Firestone. 1996. N dynamics in the rhizosphere of *Pinus ponderosa* seedlings. Soil Biol. Biochem. 28:351–362.

Ogilvie, L.A., P.R. Hirsch, and A.W.B. Johnston. 2008. Bacterial diversity of the Broadbalk "classical" winter wheat experiment in relation to long-term fertilizer inputs. Microb. Ecol. 56:525–537.

Parton, W.J., D.S. Schimel, C.V. Cole, and D.S. Ojima. 1987. Analysis of factors controlling soil organic matter levels in Great-Plains grasslands. Soil Sci. Soc. Am. J. 51:1173–1179.

Parton, W.J., J.W.B. Stewart, and C.V. Cole. 1988. Dynamics of C, N, P and S in grassland soils—A model. Biogeochemistry 5:109–131.

Paul, E.A., and F.E. Clark. 1996. Soil microbiology. Academic Press, San Diego, CA.

Paustian, K., W.J. Parton, and J. Persson. 1992. Modeling soil organic matter in organic-amended and nitrogen-fertilized long-term plots. Soil Sci. Soc. Am. J. 56:476–488.

Pietramellara, G., J. Ascher, F. Borgogni, M.T. Ceccherini, G. Guerri, and P. Nannipieri. 2009. Extracellular DNA in soil and sediment: Fate and ecological relevance. Biol. Fertil. Soils 45:219–235.

Powlson, D.S., and D. Barraclough. 1993. Mineralization and assimilation in soil–plant systems, p. 209–242. *In* R. Knowles and T.H. Blackburn (eds.) Nitrogen isotope techniques. Academic Press, San Diego, CA.

Raynaud, X., J.C. Lata, and P.W. Leadley. 2006. Soil microbial loop and nutrient uptake by plants: A test using a coupled C:N model of plant-microbial interactions. Plant Soil 287:95–116.

Recous, S., B. Mary, and G. Faurie. 1990. Microbial immobilization of ammonium and nitrate in cultivated soils. Soil Biol. Biochem. 22:913–922.

Rejsek, K., P. Formanek, and M. Pavelka. 2008. Estimation of protease activity in soils at low temperatures by casein amendment and with substitution of buffer by demineralized water. Amino Acids 35:411–417.

Rice, C.W., and J.M. Tiedje. 1989. Regulation of nitrate assimilation by ammonium in soils and isolated soil microorganisms. Soil Biol. Biochem. 21:597–602.

Saiya-Cork, K.R., R.L. Sinsabaugh, and D.R. Zak. 2002. The effects of long term nitrogen deposition on extracellular enzyme activity in an Acer saccharum forest soil. Soil Biol. Biochem. 34:1309–1315.

Schimel, J.P., and J. Bennett. 2004. Nitrogen mineralization: Challenges of a changing paradigm. Ecology 85:591–602.

Schimel, J.P., and M.K. Firestone. 1989. Inorganic-N incorporation by coniferous forest floor material. Soil Biol. Biochem. 21:41–46.

Schimel, J.P., L.E. Jackson, and M.K. Firestone. 1989. Spatial and temporal effects on plant microbial competition for inorganic nitrogen in a California annual grassland. Soil Biol. Biochem. 21:1059–1066.

Schlesinger, W.H. 2009. On the fate of anthropogenic nitrogen. Proc. Natl. Acad. Sci. USA 106:203–208.

Standford, G., and S.J. Smith. 1972. Nitrogen mineralization potential of soils. Soil Sci. Soc. Am. Proc. 38:103–107.

Stark, J.M. 2000. Nutrient transformations, p. 215–234. *In* O.E. Sala, R.B. Jackson, H.A. Mooney, and R.W. Howarth (eds.) Methods in ecosystem science. Springer-Verlag, New York.

Tabatabai, M.A. 1994. Soil enzymes, p. 778–833. *In* R.W. Weaver, S. Angle, P. Bottomly, D. Bezdicek, S. Smith, A. Tabatabai, and A. Wollum (eds.) Methods of soil analysis. Part 2. Microbiological and biochemical properties. SSSA, Madison, WI.

Vitousek, P.M., J.D. Aber, R.W. Howarth, G.E. Likens, P.A. Matson, D.W. Schindler, W.H. Schlesinger, and G.D. Tilman. 1997. Human alteration of the global nitrogen cycle: Sources and consequences. Ecol. Appl. 7:737–750.

Von Wirén, N., and M.J. Merrick. 2004. Regulation and function of ammonium carriers in bacteria, fungi, and plants. Topics Curr. Genet. 9:95–120.

Wallenstein, M.D., and M.N. Weintraub. 2008. Emerging tools for measuring and modeling the in situ activity of soil extracellular enzymes. Soil Biol. Biochem. 40:2098–2106.

Weintraub, M.N., and J.P. Schimel. 2005. Seasonal protein dynamics in Alaskan arctic tundra soils. Soil Biol. Biochem. 37:1469–1475.

White, D.C. 2007. The physiology and biochemistry of prokaryotes. 3rd edn. Oxford University Press, New York.

Wolf, D.C., J.O. Legg, and T.W. Boutton. 1994. Isotopic methods for the study of soil organic matter dynamics, p. 865–906. *In* R.W. Weaver, S. Angle, P. Bottomly, D. Bezdicek, S. Smith, A. Tabatabai, and A. Wollum (eds.) Methods of soil analysis. Part 2. Microbiological and biochemical properties. SSSA, Madison, WI.

Zak, D.R., C.B. Blackwood, and M.P. Waldrop. 2006. A molecular dawn for biogeochemistry. Trends Ecol. Evol. 21:288–295.

Zhou, J.Z., Z. He, J.D. Van Nostrand, L. Wu, and Y. Deng. 2010. Applying GeoChip analysis to disparate microbial communities. Microbe 5:60–65.

27.3 Nitrification

James I. Prosser

27.3.1 Introduction

Nitrification is the microbial oxidation of reduced forms of nitrogen to nitrate. Most nitrification in soil is carried out by autotrophic ammonia and nitrite oxidizers, sequentially converting ammonia to nitrite and nitrate. Heterotrophic nitrifiers oxidize organic forms of reduced nitrogen, for example, amines to nitrate. The major sources of soil ammonia are mineralization of organic matter, animal excreta, and two anthropogenic sources of increasing importance: nitrogen fertilizers and atmospheric nitrogen deposition, supplying 100 and 25 Tg N year^{-1}, respectively (Gruber and Galloway, 2008). Whereas NH_4^+ is held within the soil, due to the predominance of negatively charged particulate material, NO_3^- is readily leached from the soil. Nitrification is therefore important in controlling the levels of ammonium and nitrate available for plant growth and losses of fertilizer nitrogen, which can lead to polluting levels of nitrate in groundwaters. Within the soil, nitrate is also lost through denitrification to nitrogen and the greenhouse gas nitrous oxide, which is generated by both nitrifiers and denitrifiers.

27.3.2 Nitrifying Microorganisms

Autotrophic ammonia oxidation is carried out by four microbial groups, three within the *bacteria*, the betaproteobacteria, gammaproteobacteria, planctomycetes, and a group within the *archaea*.

27.3.2.1 Autotrophic Ammonia-Oxidizing Bacteria

27.3.2.1.1 Phylogeny

Phylogenetic analysis of 16S rRNA gene sequences places cultivated autotrophic ammonia oxidizers from soil in seven lineages within the betaproteobacteria (Table 27.3). Six are within the genus *Nitrosomonas* and the seventh, *Nitrosospira*, is further subdivided into five "clusters," although this subdivision is less phylogenetically robust and is based, partly, on sequences amplified directly from environmental DNA, rather than pure cultures. The gammaproteobacterial *Nitrosococcus oceanus* appears to be restricted to marine environments. Phylogeny based on the *amoA* gene (encoding ammonia monooxygenase subunit A) is generally congruent with 16S rRNA gene phylogeny. Two groups, *Nitrosomonas*

TABLE 27.3 Phylogenetic Affiliations and Environmental Origin of Isolates and Environmental Sequences of Autotrophic Ammonia- and Nitrite-Oxidizing Bacteria and Ammonia-Oxidizing Archaea

Ammonia Oxidizers	Lineage	Environmental Origin of Isolates (Where Available) and Sequences
Betaproteobacteria	*Nitrosomonas europaea*/*Nitrosococcus mobilis*	Soil, freshwater, marine, wastewater
	Nitrosomonas communis	Soil, freshwater, marine, wastewater
	Nitrosomonas oligotropha	Soil, freshwater, wastewater
	Nitrosomonas marina	Marine, freshwater
	Nitrosomonas cryotolerans	Marine
	Nitrosomonas sp. Nm143	Marine
	Nitrosomonas environmental lineage 5	Soil, freshwater, marine, wastewater (no isolate)
	Nitrosospira "Cluster 0"	Soil, freshwater
	Nitrosospira "Cluster 1"	Soil, marine (no isolate)
	Nitrosospira "Cluster 2"	Soil, freshwater, wastewater
	Nitrosospira "Cluster 3"	Soil, freshwater, wastewater
	Nitrosospira "Cluster 4"	Soil, freshwater
Gammaproteobacteria	*Nitrosococcus*	Marine
Planctomycetes	"Anammox" organisms	Marine, wastewater
Archaea	Crenarchaeal Group 1.1a	Soil and marine (no soil isolate)
	Crenarchaeal Group 1.1b	Soil (no isolate)
	Crenarchaeal Group 1.1a-associated	Soil (no isolate)
Nitrite Oxidizers		
Alphaproteobacteria	*Nitrobacter*	Soil, freshwater
Betaproteobacteria	*Candidatus Nitrotoga arctica*	Soil
Nitrospirae	*Nitrospira*	Soil, freshwater, soil
Gammaproteobacteria	*Nitrococcus*	Marine
Nitrospina	*Nitrospina*	Marine

Sources: Koops, H.P., U. Purkhold, A. Pommerening-Röser, G. Timmermann, and M. Wagner. 2003. The lithoautotrophic ammonia-oxidizing bacteria, p. 778–811. *In* M. Dworkin, S. Falkow, E. Rosenberg, K.H. Schleifer, E. Stackebrandt (eds.) The prokaryotes: An evolving electronic resource for the microbiological community. 3rd edn. Springer-Verlag, New York; Nicol, G.W., and C. Schleper. 2006. Ammonia-oxidising crenarchaeota: Important players in the nitrogen cycle? Trends Microbiol. 14:207–212; Alawi, M., A. Lipski, T. Sanders, E.-M. Pfeiffer, and E. Spieck. 2007. Cultivation of a novel cold-adapted nitrite oxidizing betaproteobacterium from the Siberian arctic. ISME J. 1:256–264.

environmental lineage 5 and *Nitrosospira* "Cluster 1," have no known cultivated representative.

Bacterial ammonia oxidizers are Gram negative. Phenotypically they are distinguished in terms of cell morphology, presence, and location of intracytoplasmic membranes (the site of ammonia oxidation) and inclusion bodies (including carboxysomes, the site of CO_2 fixation) and physiological characteristics, including urea hydrolysis, salt tolerance, and affinity for and inhibition by ammonia. Some links exist between phylogenetic groups and these phenotypic characteristics (Koops and Pommerening-Röser, 2001), but classification and identification are now routinely based on analysis of 16S rRNA or *amoA* gene sequences. Most physiological studies have been carried out on *Nitrosomonas europaea* but complete genomes of several bacterial ammonia oxidizers have been sequenced, allowing analysis of evolutionary relationships and physiological potential (Klotz and Stein, 2008). There is, however, a need for physiological characterization of a wider range of bacterial ammonia oxidizers, and for the use of techniques enabling analysis of in situ physiology, to assess physiological characteristics, and potential ecosystem function of different phylotypes.

27.3.2.1.2 Biochemistry

Soil bacterial ammonia oxidizers are chemolithoautotrophs, gaining energy by oxidation of ammonia (NH_3 rather than NH_4^+), via hydroxylamine, to nitrite. The first step is endergonic, requires oxygen and a source of reducing equivalents and is catalyzed by a membrane-bound ammonia monooxygenase:

$$NH_3 + O_2 + 2H^+ + 2e^- \rightarrow NH_2OH + H_2O. \qquad (27.2)$$

Ureolytic ammonia oxidizers can grow on urea, generating ammonia within the cell by urease activity following urea uptake. The second step, oxidation of hydroxylamine to nitrite, generates reducing equivalents, is catalyzed by hydroxylamine oxidoreductase, and uses oxygen from water:

$$NH_2OH + H_2O \rightarrow NO_2^- + 5H^+ + 4e^-. \qquad (27.3)$$
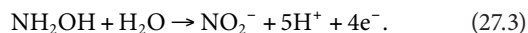
Two electrons are used in the first step, for ammonia oxidation, and the remaining two pass down the electron transport chain. NADH is synthesized by reverse electron flow, as the NO_2^-/NH_3 redox potential is only $-340\,mV$. Generation of H^+ leads to acidification, which eventually limits continued growth in liquid culture and in soil. Arp and Stein (2003) provide further details of the biochemical mechanisms involved in ammonia oxidation.

Ammonia oxidizers are autotrophs, obtaining organic carbon by carbon dioxide fixation and possess Type I Rubisco genes. This creates a demand for reducing equivalents, which are generated by reverse electron flow, consuming ATP. Consequently, the energy generated by oxidation of ammonia to nitrite is barely sufficient to support growth. An ammonia oxidizer must convert approximately 10 times its own weight of ammonia to double in size. Growth yields are therefore low and ammonia oxidizers have low growth rates, with minimum doubling times for

laboratory strains of approximately 8 h. Autotrophic bacterial ammonia oxidizers can assimilate organic carbon, for example, acetate, formate, and pyruvate, potentially increasing growth yield; but heterotrophic growth is not possible. In addition, the ammonia monooxygenase enzyme can oxidize a range of polluting organic compounds, including butane, bromoethane, benzene, ethylbenzene, chlorobenzene, and naphthalene (Arp and Stein, 2003). Ammonia oxidizers may therefore contribute to remediation in oligotrophic soils.

Soil ammonia oxidizers and methanotrophs are not closely related phylogenetically, but ammonia and methane monooxygenases can oxidize both ammonia and methane and are inhibited by similar compounds. Neither functional group can grow on the substrate of the other and ammonia oxidizers have much lower affinity and specific activities for methane oxidation than methanotrophs. There is, however, evidence for competitive inhibition, which may reduce methane oxidation in heavily fertilized soils. Ammonia oxidizers also have greater affinity for carbon monoxide than cultivated carboxydobacteria and may contribute to oxidation of this greenhouse gas.

Ammonia oxidizers produce nitrous oxide through chemical decomposition of intermediates during hydroxylamine conversion to nitrite and through nitrifier denitrification. The latter process involves sequential reduction of nitrite to nitric and nitrous oxides and nitrogen gas (Wrage et al., 2001; Stein and Yung, 2003). Functional genes encoding nitrite, nitrous oxide, and nitric oxide reductases are present in all *Nitrosomonas* and *Nitrosospira* strains investigated, are less well conserved than *amoA* genes, and may have been subject to lateral gene transfer (Cantera and Stein, 2007; Garbeva et al., 2007). Nitrous oxide production increases with decreasing oxygen concentration and is typically equivalent to around 1% of nitrite produced (Jiang and Bakken, 1999b; Garbeva et al., 2007). Technical restrictions in distinguishing nitrous oxide production by denitrifiers and nitrifiers make it difficult to estimate the contribution of ammonia oxidizers, but there is evidence that up to 80% of total soil N_2O is produced by ammonia oxidizers under some conditions (Wrage et al., 2005) and production increases following addition of artificial urine (Wrage et al., 2004a) and ammonium (Avrahami et al., 2002).

Ammonia oxidation can also occur under anaerobic conditions by two processes, although neither has been reported in soil. The first, the anammox process, is common in marine environments and wastewater treatment plants, is carried out by planctomycetes and involves the generation of nitrate and nitrogen using nitrite as an electron acceptor, with hydrazine and hydroxylamine as intermediates:

$$NH_4^+ + 1.32NO_2^- + 0.066HCO_3^- + 0.13H^+$$

$$\rightarrow 0.26NO_3^- + 1.02N_2 + 0.066CH_2O + 0.5N_{0.15} + 2.03H_2O$$

$$(27.4)$$
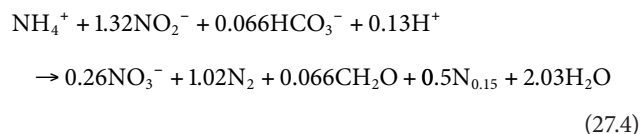
The second, reported in laboratory cultures of *Nitrosomonas eutropha*, involves use of nitrogen dioxide or nitrogen tetroxide as an electron acceptor, generating hydroxylamine and nitric

oxide, which is then converted to nitrite and dinitrogen gas (Schmidt et al., 2002):

$$NH_3 + N_2O_4 + 2H^+ + 2e^- \rightarrow NH_2OH + H_2O + 2NO \quad (27.5)$$

$$NH_2OH + H_2O \rightarrow HNO_2 + 4H^+ + 4e^- \quad (27.6)$$

$$NH_3 + N_2O_4 \rightarrow HNO_2 + 2NO + 2H^+ + 2e^- \quad (27.7)$$

$$HNO_2 + 3H^+ + 3e^- \rightarrow 0.5N_2 + 2H_2O. \quad (27.8)$$
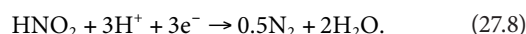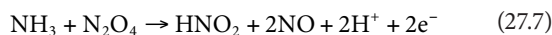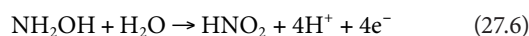
Nitric oxide can stimulate growth, activity, and cell yield and reduce the lag phase, providing potential benefits in soil, where ammonia supply is likely to be intermittent.

Although both processes are anaerobic, they require substrates produced under oxic conditions and may therefore be important at oxic–anoxic interfaces within the soil, in biofilms, or in regions alternating between oxic and anoxic conditions.

27.3.2.2 Autotrophic Ammonia-Oxidizing Archaea

There is increasing evidence that archaea contribute to soil ammonia oxidation, following the discovery of putative *amoA* and *amoB* genes and a 16S rRNA gene representative of the crenarchaeal Group 1.1b sublineage on a soil metagenome fragment (Treusch et al., 2005). The potential for archaeal ammonia oxidation was confirmed by the isolation, from a marine aquarium, of a crenarchaeal, chemolithoautotrophic NH_3 oxidizer, "*Candidatus Nitrosopumilus maritimus*" (Könneke et al., 2005). This strain is mesophilic, belongs to the crenarchaeal Group 1.1a lineage, possesses *amoA* and *B* sequences similar to those from metagenome studies, and has growth characteristics similar to those of bacterial ammonia oxidizers. No soil crenarchaeal ammonia oxidizer has yet been isolated, but crenarchaeal *amoA* genes can be amplified from all soils investigated and are frequently more abundant than bacterial *amoA* genes. Evidence for the relative contributions of bacteria and archaea to soil ammonia oxidation is reviewed in Prosser and Nicol (2008).

27.3.2.3 Autotrophic Nitrite Oxidizers

Nitrite-oxidizing bacteria fall within five phylogenetic groups (Table 27.3), three of which are represented in soil. *Nitrobacter* belongs to the alphaproteobacteria, possesses intracytoplasmic membranes, and is related to photosynthetic bacteria. *Nitrospira* belongs to the deltaproteobacteria, does not contain intracytoplasmic membranes, and is rarely found in soil enrichment cultures. The third group is represented by a single isolate, "*Candidatus Nitrotoga arctica*," recently isolated from an Arctic soil and belonging to the betaproteobacteria (Alawi et al., 2007). Two other groups, *Nitrococcus* (gammaproteobacteria) and *Nitrospina*, have not been detected in soil. Nitrite oxidation is catalyzed by nitrite oxidoreductase, with oxygen supplied by water:

$$NO_2^- + H_2O \rightarrow NO_3^- + 2H^+ + 2e^- \quad (27.9)$$
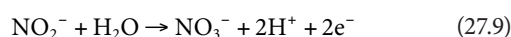
The NO_3^-/NO_2^- redox potential is $-430\,mV$ and, as for ammonia oxidizers, reducing equivalents are generated by reverse electron flow, consuming ATP. Autotrophic growth occurs at rates similar to those of autotrophic ammonia oxidizers. However, nitrite oxidizers range from autotrophs to mixotrophs and heterotrophs, growing on low-molecular-weight organic substrates. Heterotrophic growth is much slower, with doubling times of 30–150 h, but with higher biomass yield, and this may partly explain their greater abundance in soil. Nitrite oxidation to nitrate is effectively a reversible process and, under anaerobic conditions, nitrite oxidoreductase enables use of nitrate as an electron acceptor during anaerobic heterotrophic growth. Nitrite oxidizers can therefore also denitrify and contribute to production of nitrogen oxides. Biochemical studies have focused on *Nitrobacter* because of its ease of isolation and faster growth in laboratory culture. However, growing realization of the environmental importance of nitrospiras, targeted isolation of new strains, and genomic studies are increasing our understanding of the biology of nitrite oxidizers (Starkenburg et al., 2006, 2008; Maixner et al., 2008).

27.3.2.4 Heterotrophic Nitrification

Heterotrophic nitrification is the oxidation of inorganic and organic reduced forms of nitrogen to nitrite or nitrate during heterotrophic growth of fungi or bacteria. Heterotrophic nitrification does not contribute to growth or energy production and specific rates of heterotrophic nitrification are much lower than autotrophic nitrification. Nevertheless, there is evidence for its importance in acidic soils, including forest soil, where autotrophic ammonia oxidizers are inhibited or at low abundance (De Boer and Kowalchuk, 2001). In some organisms, heterotrophic nitrification operates with similar mechanisms to autotrophs, while "fungal nitrification" involves reaction of reduced organic compounds with hydroxyl radicals, produced in the presence of hydrogen peroxide and superoxide, and is linked to lignin degradation.

27.3.2.5 Cultivation and Isolation of Nitrifying Bacteria

Ammonia- and nitrite-oxidizing bacteria are cultivated in liquid mineral salts medium containing ammonium or nitrite, respectively. Growth is assessed by qualitative or quantitative changes in substrate and/or product concentrations. Growth of ammonia oxidizers is also determined by inclusion of a pH indicator in the growth medium (typically phenol red) to detect production of acid. Cultures are susceptible to contamination by heterotrophs, which generally grow faster, making purification difficult. Nitrifiers form only microscopic colonies on solid medium after incubation for many weeks, making purification from isolated colonies difficult, while purification by endpoint dilution methods requires very low levels of inoculation of fresh medium. Isolates are usually maintained as live cultures, in addition to preservation in liquid nitrogen or in a freeze-dried state because of poor survival and induction of very long lag phases during resuscitation.

It is becoming increasingly feasible to determine the physiological in situ activity of specific nitrifier phylotypes using cultivation-independent, molecular methods in association with

stable isotope probing, radiolabeling and associated techniques (Freitag et al., 2006; Hatzenpichler et al., 2008). Nevertheless, analysis of pure cultures remains an important tool in understanding the ecology and functional diversity of soil nitrifiers, for comparative genomics and for validation of molecular techniques. Soil archaeal ammonia oxidizers have yet to be isolated, but nitrifying soil bacteria can be enriched by inoculation of appropriate mineral salts medium with a soil suspension and incubation for several weeks. The use of inorganic medium reduces but does not prevent growth of heterotrophic contaminants, even with successive subculturing. Heterotrophs grow on by-products of nitrifier growth but also on organic contaminants of culture vessels and volatile organic compounds. Contamination may therefore be reduced by oven heating or acid washing of glassware and sealing of flasks, while ensuring adequate supply of oxygen (Aakra et al., 1999). Isolation of pure cultures is much more difficult, typically takes several months, due to slow growth of nitrifiers, and cultures frequently die with continued subculturing. Laboratory enrichment is likely to select for fast growing strains, but soil isolates often exhibit very low growth rates, making maintenance and physiological characterization difficult. Enrichment of strains that are representative of natural soil communities is now made easier by molecular analysis of enrichment cultures to identify dominant community members (Smith et al., 2001). Ammonium concentration in enrichment media is also important, as high ammonium concentration can be inhibitory.

27.3.2.6 Enumeration of Nitrifiers

The microscopic size of nitrifier colonies on solid media and difficulties in eliminating heterotrophs prevent viable cell enumeration by plate counting and necessitate the use of the most probable number (MPN) method. A dilution series (typically 10-fold) of a soil suspension is used to inoculate, in replicate for each dilution, liquid mineral salts medium containing ammonium or nitrite and (for ammonia oxidizers) a pH indicator. Growth is assessed qualitatively after incubation (typically for 4 weeks) and statistical tables or software programs are used to estimate cell abundance. The method is limited by requirement for laboratory growth and selection of different ammonia oxidizers at different ammonium concentrations (Phillips et al., 2000a). In addition, the number of tubes showing growth increases with increasing incubation time (Matulewich et al., 1975) and intrinsic errors result from the requirement for statistical estimation of cell number. The highest MPN dilutions exhibiting growth can, however, be used for enrichment and isolation, to increase the probability of obtaining representatives of dominant soil organisms.

The limitations of MPN enumeration can be avoided by use of PCR-based, molecular methods for quantification of diagnostic genes. Competitive PCR has been used to quantify bacterial ammonia oxidizer 16S rRNA genes (Phillips et al., 2000b; Hermansson et al., 2004) and real-time PCR for quantification of 16S rRNA and *amoA* genes (Okano et al., 2004; Leininger et al., 2006; Nicol et al., 2008). qPCR methods do not distinguish living, dead, and dormant cells and are subject to biases (Smith et al., 2006), including those common to all molecular methods. In addition, they may overestimate abundance of organisms with multiple copies of the target gene within the chromosome, and gene copy number will depend on growth rate and position on the chromosome.

Nevertheless, qPCR is rapidly becoming the standard method for enumeration of nitrifiers in environmental samples because of its ability to detect, and quantify, unculturable organisms. Phillips et al. (2000a) found that MPN counts were typically 1–2 orders of magnitude less than molecular estimates of abundance, based on qPCR of 16S rRNA genes (Figure 27.7). Molecular analysis of positive MPN tubes also indicated selection of nitrosomonads in medium with higher ammonium concentrations.
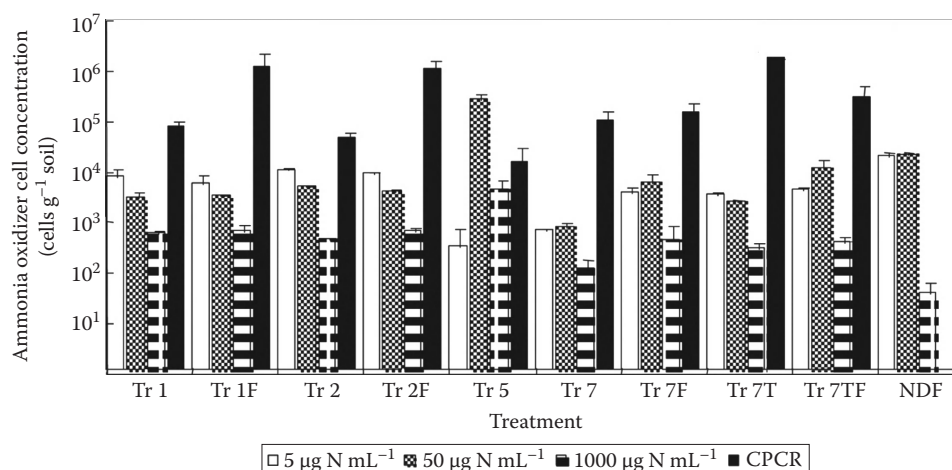
FIGURE 27.7 Abundance of bacterial ammonia oxidizers in long-term ecological research plots subjected to different management regimes. Viable cell concentration was determined using the MPN method with medium containing 5, 50, or 1000 mg NH_4^+–N mL^{-1}. Total bacterial ammonia oxidizer abundance was determined by competitive PCR amplification of 16S rRNA genes. Treatments were Tr 1, conventional tilling; Tr 2, no tilling; Tr 5, *Populus* perennial cover crop; Tr 7, historically tilled. Suffixes T and F indicate tillage and fertilization, respectively. Error bars represent the standard errors. (From Phillips, C.J., D. Harris, S.L. Dollhopf, K.L. Gross, J.I. Prosser, and E.A. Paul. 2000a. Effects of agronomic treatments on structure and function of ammonia-oxidizing communities. Appl. Environ. Microbiol. 66:5410–5418.)
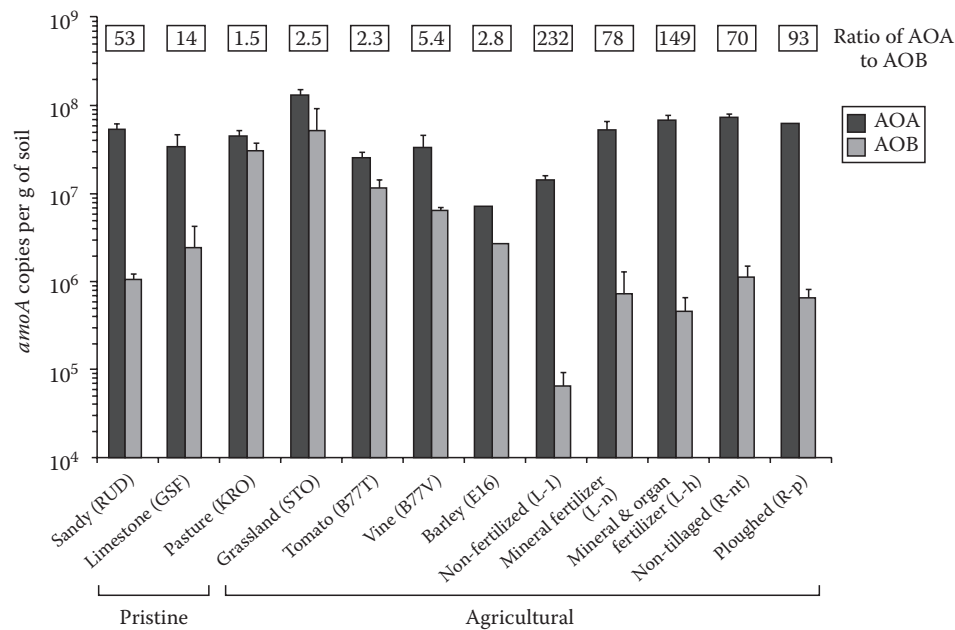
FIGURE 27.8 Abundance and ratios of abundances (in boxes) of bacterial (AOB) and archaeal (AOA) ammonia oxidizers, estimated by qPCR amplification of *amoA* genes, in 12 soils with a range of soil management and fertilization histories. (From Leininger, S., T. Urich, M. Schloter, L. Schwark, J. Qi, G.W. Nicol, J.I. Prosser, S.C. Schuster, and C. Schleper. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442:806–809.)

Quantification of archaeal and bacterial *amoA* genes provided the first evidence for a significant role for archaea in soil nitrification (Leininger et al., 2006). Archaeal *amoA* genes were more abundant in all soils investigated, sometimes by two orders of magnitude (Figure 27.8). qPCR has also been used to compare abundances of *amoA* gene transcripts, which may provide a better indication of relative activities of bacterial and archaea (see below), but gene and gene transcript abundances do not necessarily relate directly to soil activity (Prosser and Nicol, 2008).

27.3.3 Nitrifier Activity and Nitrification Rates

27.3.3.1 Growth and Activity of Laboratory Cultures

Growth parameters and cell activities have been measured for pure cultures of a limited number of ammonia and nitrite oxidizers in cell suspensions, batch culture, and continuous culture. Growth kinetics are typical of those of other bacteria. In batch culture, cell abundance increases exponentially until ammonium or nitrite is fully utilized or, for ammonia oxidizers, pH reduces specific growth rate. Product concentration exhibits quasi-exponential increases and can be used to determine specific growth rate in standard batch cultures. Maximum specific growth rates of cultivated strains fall within the range $0.02-0.09\,h^{-1}$ (Prosser, 1989). However, detailed studies of growth kinetics and parameters have been restricted to strains isolated on high concentrations of substrate and maintained in laboratory culture for extended periods, potentially selecting faster growing strains. Michaelis–Menten and Monod kinetics have been used to describe the influence of substrate or oxygen concentration on cell activity (in cell suspensions) or specific

growth rate (in chemostats), respectively. Saturation constants for growth and activity of bacterial ammonia oxidizers fall within the ranges $0.12-14\,mM\ NH_4^+/NO_2^-$, and $0.05-0.07\,mM$ NH_4^+/NO_2^-, respectively. Total cell counts enable calculation of cell activities and cultivated ammonia oxidizers have activities in the range $0.9-31.3\,fmol\ NO_2^-/NO_3^-\ cell^{-1}\ h^{-1}$ (Prosser, 1989).

27.3.3.2 Growth and Activity in Soil

In the soil, growth parameters will be influenced by a number of factors. Most soil nitrifiers are likely to be attached to surfaces, possibly in biofilms, and NH_4^+ will be adsorbed onto negatively charged minerals and organic matter. Although this can be utilized by ammonia oxidizers, specific growth rate is believed to be controlled by concentrations of ammonia in soil solution, rather than total (extractable) ammonium concentration. Ammonia oxidation rates of cultivated ammonia oxidizers are therefore reduced in soils with high cation exchange capacity due to reduction in soluble ammonia concentration and pH (Stark and Firestone, 1995; Hommes et al., 1998). Ammonia concentration will also vary throughout soil, due to heterogeneity and complexity in soil physicochemical characteristics and spatial and temporal variation in ammonia supply and utilization. These, in turn, contribute to heterogeneity in nitrifier diversity and community structure and kinetic analysis of ammonia oxidation in natural soil therefore averages activities of all members of the community occupying microenvironments with a range of conditions.

Kinetic constants can be determined in soil slurries (Stark and Firestone, 1996) or in systems that maintain some soil structure, in which moisture content can be controlled (Stark and Firestone, 1995). This approach often leads to constants that differ

significantly from those obtained in pure cultures. For example, Stark and Firestone (1996) reported a K_m value of $15\,\mu M\ NH_4^+$ for soil ammonia oxidation, significantly lower than those for laboratory cultures. The complexity of soil communities and the soil environment make it difficult to interpret these values, but they may be more relevant, possibly reflecting our inability to isolate truly representative strains. This may be particularly relevant when considering competition between ammonia oxidizers, plants, and soil heterotrophs for ammonium. Autotrophic ammonia oxidizers are often considered to be poor competitors, because of high saturation constants determined for pure cultures, but the low values observed by Stark and Firestone (1996) suggest much greater competitive ability at low ammonium concentration. Estimates of cell activities for soil nitrifiers traditionally relied on MPN counts but values are now being reassessed using molecular abundance measurements and often do not differ substantially from those obtained in pure cultures of soil ammonia oxidizers.

It is interesting to consider the relevance of these growth parameters for soil nitrifiers. High maximum specific growth rates will provide an advantage where ammonium concentration is high but, in many situations, ammonium concentration will be low and high substrate affinity will be advantageous (Taylor and Bottomley, 2006). However, the ability to respond rapidly to increases in ammonium concentration and survive starvation may be more important than growth parameters traditionally considered to give nitrifiers a competitive advantage.

Nitrate production in soil is often described in terms of logistic growth kinetics, assuming that all cells will experience similar conditions and have similar activities and growth rates. In fact, nitrate production curves can be described equally well by considering ammonia oxidizers as spatially separated clusters, growing on discrete sources of ammonium within microenvironments. Onset of activity in these microaggregates or microenvironments will be determined by the speed of recovery following appearance of ammonia. Cessation of activity will occur when acid production limits growth. Molina (1985) showed that nitrate production kinetics followed the cumulative distribution of these lag periods, suggesting that recovery from starvation may be more important than other growth parameters and that nitrification may be limited, locally, by acidification of microenvironments, rather than low concentrations of ammonia.

27.3.3.3 Nitrification Rates in Soil

Soil nitrification process rates are frequently measured without reference to the abundance or diversity of nitrifier communities present, effectively treating soil as a "black box." Potential nitrification is determined as the rate of nitrification under optimal conditions. It is calculated as the rate of ammonia consumption or nitrite + nitrate production during incubation, at constant temperature, of soil or soil slurry at nonlimiting ammonia concentration. Ammonia may be added if ammonia concentrations in soil are limiting and soils are aerated, usually by rapid shaking of soil slurries, to minimize loss of nitrate through denitrification. Potential nitrite oxidation can be determined by addition of nitrite, rather than ammonia.

In the absence of significant nitrifier growth, zero order kinetics are observed and nitrate production rates are the product of nitrifier specific activity and biomass or cell concentration. In fact, this method has been used to estimate nitrifier biomass. If nitrifier abundances are low and/or incubation is continued, growth will lead to accelerating rates of nitrate production, which can be used to estimate the maximum specific growth rate of soil nitrifiers. Similarly, if ammonia concentration is low in relation to saturation constants for ammonium, the rate of nitrate production will decrease during incubation, as ammonium is utilized.

A major complication in measuring soil nitrification is the effect of other nitrogen cycle processes on ammonia and nitrate concentrations. Mineralization will increase ammonia supply; denitrification, immobilization, and leaching will remove nitrate; and plant uptake can reduce ammonia and/or nitrate. These factors are considered in net nitrification rate, calculated as nitrate production minus nitrate consumption, and gross nitrification rate, calculated as nitrate production plus nitrate consumption. Net nitrification rate is measured in the same way as potential nitrification rate but without attempts to inhibit denitrification, by aeration, or other nitrate consumption processes. Where ammonia concentrations are low, sensitivity can be increased by use of stable isotope methods, which are also necessary for determination of nitrate consumption in measuring gross nitrification. One approach is to add $^{15}NH_4^+$ as a tracer and measure production of $^{15}NO_3^-$. This suffers from the need to add ammonium, potentially stimulating nitrification, dilution of $^{15}NH_4^+$ by ammonium from mineralization, and lack of assessment of heterotrophic nitrification (Hart et al., 1994). In a second approach, the pool dilution method, soil is amended with $^{15}NO_3^-$ and its dilution by unlabeled nitrate produced by nitrifiers from unlabeled ammonium is measured (Barraclough and Puri, 1995). When combined with specific inhibitors of ammonia oxidation, such as acetylene or nitrapyrin, these techniques can distinguish autotrophic and heterotrophic nitrification. More sophisticated stable isotope approaches are used for discrimination of nitrogen cycle processes. These include addition of double-labeled compounds ($^{14}NH_4^{15}NO_3$ or $^{15}NH_4^{15}NO_3$), measurement of variations in $^{15}N/^{14}N$ and $^{18}O/^{16}O$ isotope ratios, and analysis of isotopomers, which discriminate between labeling of nitrogen atoms within the N_2O molecule (Sutka et al., 2006; Ostrom et al., 2007).

Nitrification rates can be determined in laboratory microcosms or mesocosms, or in the field. Microcosms are useful for assessing the influence of environmental factors on nitrification under controlled and well-monitored conditions. Field measurements may be more relevant where effects of naturally changing environmental conditions are being studied, including the influence of plant communities. Rates measured in soil are more variable than in pure culture systems, due to heterogeneity in soil physicochemical conditions, in particular ammonium concentration, pH and oxygen concentration, and heterogeneity in nitrifier biomass and community composition. Variability is greatest in unmanaged grasslands soils, where ammonia

concentration is often determined by grazing patterns of sheep and cattle and concentration of ammonium in urine patches. There is evidence for reduced variability in nitrification rates within agricultural soils, through increased homogeneity of pH, ammonia concentration, and ammonia oxidizer communities (Webster et al., 2002). As a consequence, deviations from expected kinetics should not be ignored, as they may provide important information on the factors controlling nitrification.

27.3.4 Nitrifier Communities and Diversity

Nitrifier communities are now routinely characterized using molecular methods for analysis of 16S rRNA and *amoA* gene sequences amplified from nucleic acids extracted directly from soil. Specific primers are available for betaproteobacterial ammonia oxidizer 16S rRNA (Kowalchuk et al., 1997; Purkhold et al., 2000; Mahmood et al., 2006) and *amoA* genes (Rotthauwe et al., 1997; Stephen et al., 1999; Webster et al., 2002), for archaeal and crenarchaeal (not necessarily ammonia oxidizer) 16S rRNA genes, and for crenarchaeal *amoA* genes. Nitrite oxidizer communities may be investigated using primers for 16S rRNA and nitrite oxidoreductase genes, but separate primers sets are required for *Nitrobacter* and *Nitrospira* (Freitag et al., 2005; Wertz et al., 2008). No specific primers are available for heterotrophic nitrifiers.

Amplicons are analyzed by cloning and sequencing, to determine phylogenetic relationships within community members and to identify phylotypes present, by comparison with sequences in databases. More rapid, but less detailed analysis of communities is achieved using DNA fingerprinting methods, such as denaturing gradient gel electrophoresis (DGGE) (e.g., Nicol et al., 2008), temperature gradient gel electrophoresis (TGGE) (e.g., Gray et al., 2003), or terminal-restriction fragment length polymorphism (T-RFLP) (e.g., Horz et al., 2000; Bottomley et al., 2004) analysis. These allow analysis of changes in nitrifier communities associated with different environmental factors and soil management treatments and identification of phylotypes dominating communities and responding most to change.

Many studies target RNA, rather than DNA, enabling assessment of the activities of different phylotypes. DNA is removed from extracted nucleic acids, RNA is reverse transcribed to DNA and genes are then amplified, using either 16S rRNA or functional gene primers. rRNA-targeted analysis of 16S rRNA genes increases sensitivity, as cells contain more ribosomes than rRNA genes, and also assesses general cellular activity, on the assumption that actively growing cells will contain more ribosomes than dormant cells. Amplification of functional genes from mRNA assesses transcriptional activity. Both approaches allow quantification of active nitrifiers and identification of phylotypes that are active (e.g., Leininger et al., 2006; Nicol et al., 2008). Stable isotope probing has also been used to determine which nitrifiers are actively fixing carbon dioxide. This involves incubation of environmental samples in the presence of ^{13}C-labeled CO_2 and molecular analysis of ^{13}C-labeled and unlabeled nucleic acids,

after separation by density gradient centrifugation (Freitag et al., 2006; Jia and Conrad, 2009). ^{13}C-labeled DNA and RNA will only contain gene sequences of nitrifiers that have assimilated CO_2. Correlation with nitrification activity is strong evidence of which nitrifiers are active in the soil and indicated that bacteria, rather than archaea, may be responsible for nitrification in an agricultural soil (Jia and Conrad, 2009).

These techniques now enable meaningful studies of nitrifier communities in soil. Their major advantage is the avoidance of the requirement to cultivate these slow-growing organisms but sequence analysis also provides much easier and more reliable phylogenetic analysis and identification. Analysis of both 16S rRNA and *amoA* gene clone libraries from soil demonstrates high levels of diversity in bacterial ammonia oxidizers and duplicate sequences are rarely seen (Stephen et al., 1996; Prosser and Embley, 2002). Environmental sequences differ from those of pure cultures but are sometimes identical to those of ammonia oxidizers enriched from the same soil (Smith et al., 2001). Environmental sequences generally confirmed phylogenies based on cultivated ammonia oxidizers but indicate the existence of two groups not represented in cultivated ammonia oxidizers and the possible subdivision of *Nitrosospira* into five clusters (Table 27.3). Cultivation-based studies indicated that nitrosomonads were "typical" soil organisms but molecular studies demonstrated that most soils are dominated by nitrosospiras (Stephen et al., 1996; Prosser and Embley, 2002). This is believed to reflect selection of nitrosomonads in laboratory cultures and *Nitrosomonas* is frequently found in cultures in final dilutions of MPN counts, even in soils dominated by *Nitrosospira*.

Congruence of phylogenies of bacterial 16S rRNA and *amoA* genes (Purkhold et al., 2003) allows cross-referencing of analyses using both approaches, but this is more difficult for crenarchaeal ammonia oxidizers, because of the lack of soil isolates and enrichment cultures. Most soil crenarchaeal 16S rRNA genes fall within the 1.1b group, but a significant number are also found in the "marine" 1.1a and 1.1a-associated groups (Table 27.3). Crenarchaeal *amoA* sequences have been amplified from all soils investigated, show high diversity, and most fall within a single clade. Again, however, a small but significant number of *amoA* sequences fall within a group more commonly amplified from marine samples. Metagenomic studies show an association of *amoA* sequences with Group 1.1b and there appear to be links between clusters within these groups, based on distributions of sequences from soils of different pH (Nicol et al., 2008; Tourna et al., 2008).

Nitrite rarely accumulates in soil and attention has generally focused on ammonia oxidizer communities, which are seen as a more important control point in nitrification. In addition, study of nitrite oxidizers is made more difficult by the requirement for different primer sets for *Nitrobacter* and *Nitrospira*. Phylogenetic analysis of 16S rRNA gene sequences from soil indicates greater diversity within *Nitrospira* than *Nitrobacter* (Freitag et al., 2005), as in other environments (Daims et al., 2000; Maixner et al., 2006). Analysis of the nitrite oxidoreductase functional gene, *nxrA*, suggests some sequence diversity within *Nitrobacter* (Poly et al., 2008; Wertz et al., 2008) but has not yet been attempted with *Nitrospira*.

27.3.5 Links between Nitrifier Communities and Soil Nitrification Activity

Techniques for measurement of soil nitrification rates, and the influence of soil conditions on environmental change on rates, have been available for many years. The major advance in the past decade has been the ability to characterize nitrifier communities, and to determine the links between nitrification rates and nitrifier abundance, community structure, and diversity. These studies have involved analysis of nitrifiers in soils with different physico-chemical characteristics, changes in communities in response to changing conditions, and, more recently, analysis of in situ activity. The major factors investigated are those considered to be most important in controlling soil nitrification: pH and ammonium concentration. Most studies have focused on bacterial ammonia, rather than nitrite oxidizers, and, more recently, the potential role of archaeal ammonia oxidizers in soil nitrification.

27.3.5.1 Soil pH

Cultivated bacterial ammonia oxidizers do not grow in batch culture below pH 7, because of reduced availability of NH_3, due to ionization (Allison and Prosser, 1991; Jiang and Bakken, 1999a). This is also considered to be the major reason for low nitrification rates in acid soils, although a survey of gross nitrification rates in a wide range of soils indicates no significant effect of soil pH (Booth et al., 2005). Acidophilic heterotrophs may be responsible for nitrification in some acid soils, but autotrophic nitrification occurs in soils with pH values as low as 3 (De Boer and Kowalchuk, 2001). In addition, bacterial autotrophic ammonia oxidizers are readily enriched from acid soils and molecular data indicate the presence and abundance of both bacterial and archaeal *amoA* genes.

In agricultural systems, liming of acid soils increases nitrification rates and consequent loss of ammonium-based fertilizer and leaching of nitrate. Knowledge of soil pH is important in predicting rates of nitrification and optimizing the timing of application of fertilizer and nitrification inhibitors. In forest ecosystems, and other nonmanaged acid soils, nitrification rates are low. However, ammonia concentrations may be high and soils may be nitrogen-saturated when nitrogen supply exceeds plant demand. Such high concentrations may themselves be a result of low nitrification rates, because of low pH, or due to atmospheric nitrogen deposition. These soils may have significant nitrification potential if pH is increased, through liming or clear-cutting. Ammonia oxidizers are generally less abundant in acid soils (typically 10^3–10^4 cells g^{-1} soil) and sometimes below detection limits of both MPN and qPCR methods (Schmidt et al., 2007). However, increasing pH, by liming, can increase abundance to detectable levels (Backman et al., 2003) and may lead to high rates of nitrification and increased nitrous oxide production.

Characterization of bacterial ammonia oxidizers in soils from a range of geographical regions provides evidence for pH selection of different *Nitrosospira* phylotypes, with increased relative abundance of *Nitrosospira* cluster 2 sequence-types in acid soils and of *Nitrosospira* cluster 3 at higher pH (Kowalchuk and Stephen, 2001). Selection is more evident in pH gradients
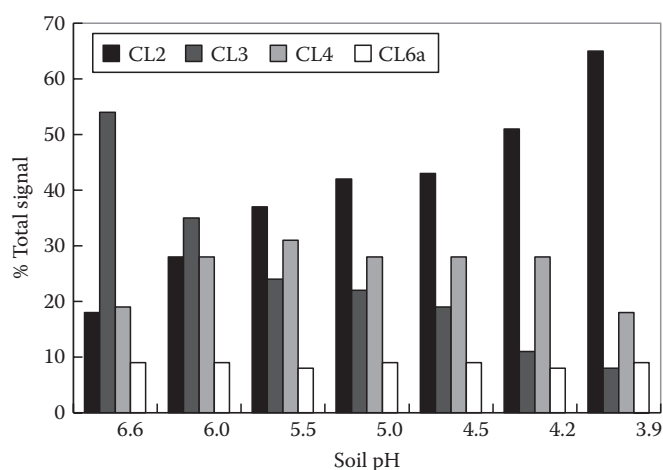
FIGURE 27.9 The influence of soil pH on bacterial ammonia oxidizer community structure. The relative abundances of *Nitrosospira* clusters 2, 3, and 4 and *Nitrosomonas* cluster 6a estimated as the relative intensities of DGGE bands following amplification of ammonia oxidizer 16S rRNA genes from DNA extracted from soil maintained at a range of pH values. (From Stephen, J.R., G.A. Kowalchuk, M.A.V. Bruns, A.E. McCaig, C.J. Phillips, T.M. Embley, and J.I. Prosser. 1998. Analysis of beta-subgroup proteobacterial ammonia oxidizer populations in soil by denaturing gradient gel electrophoresis analysis and hierarchical phylogenetic probing. *Appl. Environ. Microbiol.* 64:2958–2965.)

maintained for many years. Analysis of one such site, with a gradient from pH 3.9–6.6, established for 36 years, showed a change in dominance from *Nitrosospira* cluster 2 to cluster 3 sequences as soil pH increased (Figure 27.9; Stephen et al., 1998). The stability of these long-term community effects was confirmed by analysis of the same soils 9 years later (Nicol et al., 2008). The latter study also found significant pH-associated differences in crenarchaeal *amoA* gene sequence types and determined relative abundances and transcriptional activities of bacterial and archaeal ammonia oxidizers (Figure 27.10). Crenarchaeal *amoA* genes and gene transcripts were always more abundant than bacterial genes, suggesting a greater contribution by crenarchaea to ammonia oxidation in these soils. The most striking effect of pH was, however, on the ratio between *amoA* genes and gene transcripts, which may be a better measure of activity than transcript number (Nicolaisen et al., 2008; Freitag and Prosser, 2009). With increasing soil pH, this ratio fell sharply for crenarchaeal genes, and increased for bacteria, and was greater for bacteria above pH 6.5. The links between *amoA* transcriptional activity and nitrification rate require further study, but the results indicate that the relative contributions of crenarchaea to nitrification may be greater in acid soils. The results were confirmed in short-term studies in which crenarchaeal and bacterial transcriptional activities were compared in microcosms containing mixtures of low and neutral pH soil readjusted to pH 4.5 and 7 (Nicol et al., 2008).

Growth and activity of ammonia oxidizers in low pH soils may result from protection by growth on anionic surfaces (where ammonium will be concentrated), protection in biofilms, activity in microenvironments in which pH is neutral, and
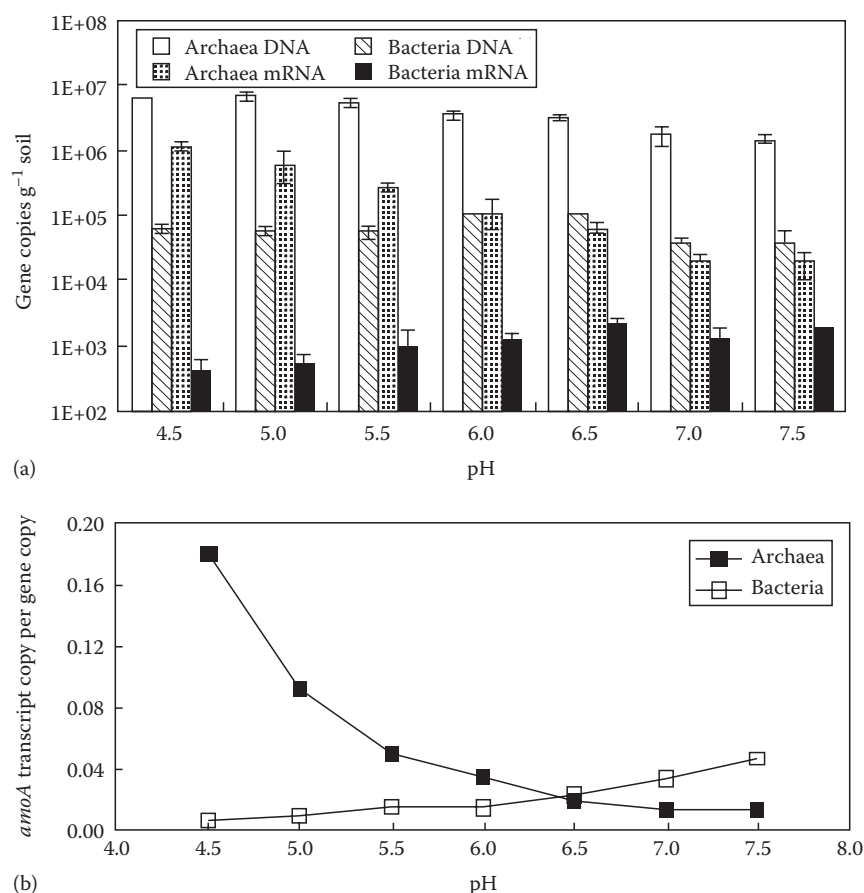
FIGURE 27.10 The influence of soil pH on bacterial and crenarchaeal ammonia oxidizer abundance and transcriptional activity. Abundances were determined by qPCR amplification of *amoA* genes and *amoA* gene transcripts in DNA extracted from soil maintained at a range of pH values. (a) Crenarchaeal and bacterial *amoA* gene and transcript abundance. (b) Ratios of crenarchaeal and bacterial gene transcript:gene abundance. Error bars represent standard errors of replicate field samples at each soil pH. (Based on data in Nicol, G.W., S. Leininger, C. Schleper, and J.I. Prosser. 2008. The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. Environ. Microbiol. 10:2966–2978.)

favorable for growth, or utilization of urea as an ammonium source (De Boer and Kowalchuk, 2001). Ammonium oxidizers with urease activity convert ammonia generated intracellularly, after uptake of urea, and growth on urea is independent of pH in the range 4–8 (Burton and Prosser, 2001). Physiological studies are required to determine whether these characteristics are responsible for selection of different phylotypes in acid soils.

27.3.5.2 Nitrogen Fertilization and Ammonia Concentration

The nitrification potential of a soil will be determined by nitrifier biomass, specific rates of ammonia oxidation, and soil physicochemical conditions, particularly pH and oxygen concentration. If conditions are favorable for nitrification, nitrifier biomass will be determined by the balance between flux of ammonia into the soil and death and removal. Flux, in turn, will be controlled by rates of addition, through mineralization, input of animal waste, fertilizer addition and atmospheric deposition, and removal, through root uptake and conversion to nitrate. In contrast, specific rates of nitrification will vary with ammonia concentration. Quantitative effects of ammonia

concentration on specific growth rate vary between ammonia oxidizer phylotypes, through differences in maximum specific growth rates, substrate affinities and sensitivity to high ammonia concentration. Growth rate will be determined by the concentration of ammonia in the soil solution, rather than bulk ammonia concentrations, because of adsorption to negatively charged soil particles. Measured nitrification rates will also be influenced by heterogeneity in ammonium concentrations, with high and potentially inhibitory concentrations of ammonia near decomposing organic matter, fertilizer particles, or urine patches. Heterogeneity will be reduced by plowing and there is evidence for reduced variability in nitrification rates (White et al., 1987), physicochemical characteristics, and ammonia oxidizer diversity in soils that are not plowed (Webster et al., 2002; Patra et al., 2006; Dell et al., 2008).

Ammonia oxidizer abundance is generally greater in fertilized soils and those subjected to disturbance by fertilizer application and plowing (Figure 27.8) (Bruns et al., 1999; Phillips et al., 2000a, 2000b). Bacterial ammonia communities are influenced by agricultural and land management practices (reviewed in Kowalchuk and Stephen, 2001), with evidence for selection of *Nitrosospira*

cluster 3 in fertilized and managed systems, *Nitrosospira* cluster 4 dominant in unmanaged systems and *Nitrosomonas*-like sequences in heavily fertilized grassland soils (Webster et al., 2002) and compost (Kowalchuk et al., 1999; Avrahami et al., 2002). Ammonia concentration also affects development of communities in soil microcosms (Mahmood and Prosser, 2006), with increases in *Nitrosospira* cluster 2 phylotypes at all applied nitrogen levels and in clusters 3 and 4 at low and high nitrogen, respectively. Nitrite oxidizers also respond to long-term ammonia fertilizer application and plowing (Freitag et al., 2005), with changes in *Nitrosospira* but not *Nitrobacter* communities.

Lack of information on comparative physiology of nitrifiers makes it difficult to explain community changes or their significance for ecosystem function. However, some changes may be due to inhibition by high ammonia concentration, which will occur in urine patches following urea hydrolysis. Webster et al. (2005) found that nitrification of artificial urine added to unmanaged soil microcosms occurred only after lag periods of several weeks in samples dominated by *Nitrosospira* cluster 3a strains, which are sensitive to high ammonia concentration. Microcosms dominated by ammonia-tolerant *Nitrosospira* cluster 3b strains showed only short lag periods. Bacterial ammonia oxidizer community structure consequently impacted significantly on the availability of ammonia for plant growth. These effects of grazing were diminished by long-term application of inorganic nitrogen fertilizer, which will have reduced heterogeneity and also increased abundance of all ammonia oxidizer phylotypes, such that lag phases were independent of urine addition. Differences in nitrification rate and ammonia oxidizer community structure and abundance have also been reported in grazed and ungrazed field sites (Le Roux et al., 2003, 2008; Patra et al., 2005). Communities and nitrification rates reverted after switches from grazed to ungrazed, and vice versa, within several months, returning to values seen in equivalent, unaltered soils within 12 months.

27.3.6 Control and Management of Soil Nitrification

Conversion of ammonia-based fertilizers to nitrate, and subsequent loss and environmental pollution through leaching and denitrification to nitrous oxide, has led to development of commercial application of nitrification inhibitors. Inhibitors also have an important role in distinguishing different nitrogen cycling processes, when measuring nitrification rates in soil and in investigations of nitrifier biochemistry and physiology (Arp and Stein, 2003).

Commercial inhibitors target ammonia oxidation; and Subbarao et al. (2006) and Singh and Verma (2007) review current literature on available compounds, their mechanisms of action, their effectiveness in controlling soil nitrification in agricultural soils, and factors determining choice of inhibitor. Most research has been carried out on Nitrapyrin [2-chloro-6-(trichloromethyl) pyridine, *N*-Serve] but other commonly used inhibitors include dicyandiamide (DCD), allylthiourea,

3,4-dimethylpyrazol-phosphate (DMPP), 2-amino-4-chloro-6-methylpyrimidine (AM), and acetylene. Nitrite oxidation is inhibited by chlorate and has been used to measure soil nitrification as nitrite production, avoiding the need for analysis of both nitrite and nitrate. Traditionally, nitrification inhibitors have been used to reduce fertilizer loss, through leaching and denitrification of nitrate, and associated nitrate pollution. More recently, their value has been reassessed in relation to reducing greenhouse gas production by both nitrifiers and denitrifiers. The value of agricultural use of nitrification inhibitors has been reviewed for a number of studies (Wolt, 2004), and include average 7% increases in crop yield and 51% decreases in nitrous oxide emissions.

Despite conflicting reports, reviewed by Subbarao et al. (2006), there is evidence that allelopathic inhibitors of nitrification are produced, naturally, by plant and tree roots. Inhibitory compounds include phenolics, terpenes, and flavonoids and may function to sustain levels of soil ammonium, particularly in climax ecosystems. Mechanistic studies of allelopathic inhibition are hampered by technical difficulties in determining the amounts and inhibitory activities of complex organic compounds in soil. Recently, a luminescence-marked strain of *Nitrosomonas europaea* was used to screen for ammonia oxidizer inhibitors produced by a range of plants (Subbarao et al., 2007), and this approach may increase our understanding of plant-derived inhibitors and more efficient and sustainable measures to reduce fertilizer loss.

Both ammonia and nitrite oxidizers are protected from the effects of inhibitors by attachment to surfaces, including soil minerals and organic matter, and approximately 10-fold higher concentrations are required for inhibition in soil. Despite the availability of molecular techniques for analysis of nitrifier community structure, there is little information on variations in sensitivity of different ammonia oxidizer phylotypes. Wrage et al. (2004b) found that acetylene inhibited nitrous oxide production by *Nitrosomonas europaea* but not *Nitrosospira briensis*, despite inhibition of ammonia oxidation by both strains. Variation in sensitivity of different ammonia oxidizers, and differences in sensitivity of bacterial and archaeal ammonia oxidizers, may have implications for the effectiveness of nitrification inhibitors and environmental consequences.

27.3.7 Conclusions

Nitrification is central to the soil nitrogen cycle and results in considerable economic losses of nitrogen fertilizer and environmental pollution, through increased levels of atmospheric nitrous oxide and groundwater nitrate. The past decade has seen major advances in our knowledge of the diversity and community structure of soil nitrifiers. This complements existing, extensive knowledge on soil nitrification processes, which has also been advanced, in particular, through developments in use of ^{15}N-based techniques for unraveling different nitrogen cycling processes. Ammonia- and nitrite-oxidizing bacteria are considerably more diverse than indicated by cultivation studies alone;

and crenarchaea, in addition to bacteria, can oxidize ammonia and are abundant in soil. Molecular techniques for analysis of microbial communities continue to advance rapidly and are likely to increase our understanding of the links between nitrifier community structure, rates of soil nitrification, and the impact of environmental change. Integration of process and community studies has the potential to increase our ability to control nitrification, to reduce fertilizer demand, and to reduce greenhouse gas production.

References

Aakra, A., J.B. Utåker, I.F. Nes, and L.R. Bakken. 1999. An evaluated improvement of the extinction dilution method for isolation of ammonia-oxidizing bacteria. J. Microbiol. Methods 39:23–31.

Alawi, M., A. Lipski, T. Sanders, E.-M. Pfeiffer, and E. Spieck. 2007. Cultivation of a novel cold-adapted nitrite oxidizing betaproteobacterium from the Siberian arctic. ISME J. 1:256–264.

Allison, S.M., and J.I. Prosser. 1991. Urease activity in neutrophilic autotrophic ammonia-oxidizing bacteria isolated from acid soils. Soil Biol. Biochem. 23:45–51.

Arp, D.J., and L.Y. Stein. 2003. Metabolism of inorganic nitrogen compounds by ammonia-oxidizing bacteria. Crit. Rev. Biochem. Mol. Biol. 38:471–495.

Avrahami, S., R. Conrad, and G. Braker. 2002. Effect of soil ammonium concentration on N_2O release and on the community structure of ammonia oxidizers and denitrifiers. Appl. Environ. Microbiol. 68:5685–5692.

Backman, J.S.K., A. Hermansson, C.C. Tebbe, and P.E. Lindgren. 2003. Liming induces growth of a diverse flora of ammonia-oxidising bacteria in acid spruce forest soil as determined by SSCP and DGGE. Soil Biol. Biochem. 35:1337–1347.

Barraclough, D., and G. Puri. 1995. The use of ^{15}N pool dilution and enrichment to separate the heterotrophic and autotrophic pathways of nitrification. Soil Biol. Biochem. 27:17–22.

Booth, M.S., J.M. Stark, and E. Rastetter. 2005. Controls on nitrogen cycling in terrestrial ecosystems: A synthetic analysis of literature data. Ecol. Monogr. 75:139–157.

Bottomley, P.J., A.E. Taylor, S.A. Boyle, S.K. McMahon, J.J. Rich, K. Cromack, and D.D. Myrold. 2004. Responses of nitrification and ammonia-oxidizing bacteria to reciprocal transfers of soil between adjacent coniferous forest and meadow vegetation in the Cascade Mountains of Oregon. Microb. Ecol. 48:500–508.

Bruns, M.A., J.R. Stephen, G.A. Kowalchuk, J.I. Prosser, and E.A. Paul. 1999. Comparative diversity of ammonia oxidizer 16S rRNA gene sequences in native, tilled, and successional soils. Appl. Environ. Microbiol. 65:2994–3000.

Burton, S.A.Q., and J.I. Prosser. 2001. Autotrophic ammonia oxidation at low pH through urea hydrolysis. Appl. Environ. Microbiol. 67:2952–2957.

Cantera, J.J.L., and L.Y. Stein. 2007. Molecular diversity of nitrite reductase genes (*nirK*) in nitrifying bacteria. Environ. Microbiol. 9:765–776.

Daims, H., P.H. Nielsen, J.L. Nielsen, S. Juretschko, and M. Wagner. 2000. Novel *Nitrospira*-like bacteria as dominant nitrite-oxidizers in biofilms from wastewater treatment plants: Diversity and in situ physiology. Water Sci. Technol. 41:85–90.

De Boer, W., and G.A. Kowalchuk. 2001. Nitrification in acid soils: Micro-organisms and mechanisms. Soil Biol. Biochem. 33:853–866.

Dell, E.A., D. Bowman, T. Rufty, and W. Shi. 2008. Intensive management affects composition of betaproteobacterial ammonia oxidizers in turfgrass systems. Microb. Ecol. 56: 178–190.

Freitag, T.E., L. Chang, C.D. Clegg, and J.I. Prosser. 2005. Influence of inorganic nitrogen-management regime on the diversity of nitrite oxidizing bacteria in agricultural grassland soils. Appl. Environ. Microbiol. 71:8323–8334.

Freitag, T.E., L. Chang, and J.I. Prosser. 2006. Changes in the community structure and activity of betaproteobacterial ammonia-oxidizing sediment bacteria along a freshwater-marine gradient. Environ. Microbiol. 8:684–696.

Freitag, T.E., and J.I. Prosser. 2009. Correlation of methane production and functional gene transcriptional activity in a peat soil. Appl. Environ. Microbiol. 75:6679–6687.

Garbeva, P., E.M. Baggs, and J.I. Prosser. 2007. Phylogeny of nitrite reductase (*nirK*) and nitric oxide reductase (*norB*) genes from *Nitrosospira* species isolated from soil. FEMS Microbiol. Lett. 266: 83–89.

Gray, N.D., R.C. Hastings, S.K. Sheppard, P. Loughnane, D. Lloyd, A.J. McCarthy, and I.M. Head. 2003. Effects of soil improvement treatments on bacterial community structure and soil processes in an upland grassland soil. FEMS Microbiol. Ecol. 46: 11–22.

Gruber, N., and J.N. Galloway. 2008. An earth-system perspective of the global nitrogen cycle. Nature 451:293–296.

Hart, S.C., J.M. Stark, E.A. Davidson, and M.K. Firestone. 1994. Nitrogen mineralization, immobilization and nitrification, p. 985–1018. *In* R.W. Weaver (ed.) Methods of soil analysis: Microbiological and biochemical properties. 3rd edn. SSSA, Madison, WI.

Hatzenpichler, R., E.V. Lebedeva, E. Spieck, K. Stoecker, A. Richter, H. Daims, and M. Wagner. 2008. A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. Proc. Natl. Acad. Sci. USA 105:2134–2139.

Hermansson, A., J.S.K. Bäckman, B.H. Svensson, and P.E. Lindgren. 2004. Quantification of ammonia-oxidising bacteria in limed and non-limed acidic coniferous forest soil using real-time PCR. Soil Biol. Biochem. 36:1935–1941.

Hommes, N.G., S.A. Russell, P.J. Bottomley, and D.J. Arp. 1998. Effects of soil on ammonia, ethylene, chloroethane, and 1,1,1-trichloroethane oxidation by *Nitrosomonas europaea*. Appl. Environ. Microbiol. 64:1372–1378.

Horz, H.P., J.H. Rotthauwe, T. Lukow, and W. Liesack. 2000. Identification of major subgroups of ammonia-oxidizing bacteria in environmental samples by T-RFLP analysis of *amoA* PCR products. J. Microbiol. Methods 39:197–204.

Jia, Z., and R. Conrad. 2009. Bacteria rather than archaea dominate microbial ammonia oxidation in an agricultural soil. Environ. Microbiol. 11:1658–1671.

Jiang, Q.Q., and L.R. Bakken. 1999a. Comparison of *Nitrosospira* strains isolated from terrestrial environments. FEMS Microbiol. Ecol. 30:171–186.

Jiang, Q.Q., and L.R. Bakken. 1999b. Nitrous oxide production and methane oxidation by different ammonia-oxidizing bacteria. Appl. Environ. Microbiol. 65: 2679–2684.

Klotz, M.G., and L.Y. Stein. 2008. Nitrifier genomics and evolution of the nitrogen cycle. FEMS Microbiol. Lett. 278:146–156.

Könneke, M., A.E. Bernhard, J.R. De La Torre, C.B. Walker, J.B. Waterbury, and D.A. Stahl. 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437: 543–546.

Koops, H.P., and A. Pommerening-Röser. 2001. Distribution and ecophysiology of the nitrifying bacteria emphasizing cultured species. FEMS Microbiol. Ecol. 37:1–9.

Koops, H.P., U. Purkhold, A. Pommerening-Röser, G. Timmermann, and M. Wagner. 2003. The lithoautotrophic ammonia-oxidizing bacteria, p. 778–811. *In* M. Dworkin, S. Falkow, E. Rosenberg, K.H. Schleifer, E. Stackebrandt (eds.) The prokaryotes: An evolving electronic resource for the microbiological community. 3rd edn. Springer-Verlag, New York.

Kowalchuk, G.A., Z.S. Naoumenko, P.J.L. Derikx, A. Felske, J.R. Stephen, and I.A. Arkhipchenko. 1999. Molecular analysis of ammonia-oxidizing bacteria of the beta subdivision of the class proteobacteria in compost and composted materials. Appl. Environ. Microbiol. 65:396–403.

Kowalchuk, G.A., and J.R. Stephen. 2001. Ammonia-oxidizing bacteria: A model for molecular microbial ecology. Annu. Rev. Microbiol. 55: 485–529.

Kowalchuk, G.A., J.R. Stephen, W. De Boer, J.I. Prosser, T.M. Embley, and J.W. Woldendorp. 1997. Analysis of ammonia-oxidizing bacteria of the β subdivision of the class proteobacteria in coastal sand dunes by denaturing gradient gel electrophoresis and sequencing of PCR-amplified 16S ribosomal DNA fragments. Appl. Environ. Microbiol. 63:1489–1497.

Le Roux, X., M. Bardy, P. Loiseau, and F. Louault. 2003. Stimulation of soil nitrification and denitrification by grazing in grasslands: Do changes in plant species composition matter? Oecologia 137:417–425.

Le Roux, X., F. Poly, P. Currey, C. Commeaux, B. Hai, G.W. Nicol, J.I. Prosser, M. Schloter, E. Attard, and K. Klumpp. 2008. Effects of aboveground grazing on coupling among nitrifier activity, abundance and community structure. ISME J. 2:221–232.

Leininger, S., T. Urich, M. Schloter, L. Schwark, J. Qi, G.W. Nicol, J.I. Prosser, S.C. Schuster, and C. Schleper. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442:806–809.

Mahmood, S., T.E. Freitag, and J.I. Prosser. 2006. Comparison of PCR primer-based strategies for characterization of ammonia oxidizer communities in environmental samples. FEMS Microbiol. Ecol. 56:482–493.

Mahmood, S., and J.I. Prosser. 2006. The influence of synthetic sheep urine on ammonia oxidizing bacterial communities in grassland soil. FEMS Microbiol. Ecol. 56:444–454.

Maixner, F., D.R. Noguera, B. Anneser, K. Stoecker, G. Wegl, M. Wagner, and H. Daims. 2006. Nitrite concentration influences the population structure of *Nitrospira*-like bacteria. Environ. Microbiol. 8:1487–1495.

Maixner, F., M. Wagner, S. Lücker, E. Pelletier, S. Schmitz-Esser, K. Hace, E. Spieck, R. Konrat, D. Le Paslier, and H. Daims. 2008. Environmental genomics reveals a functional chlorite dismutase in the nitrite-oxidizing bacterium 'Candidatus *Nitrospira defluvii*.' Environ. Microbiol. 10:3043–3056.

Matulewich, V.A., P.F. Strom, and M.S. Finstein. 1975. Length of incubation for enumerating nitrifying bacteria present in various environments. Appl. Microbiol. 29:265–268.

Molina, J.A.E. 1985. Components of rates of ammonium oxidation in soil. Soil Sci. Soc. Am. J. 49:603–609.

Nicol, G.W., S. Leininger, C. Schleper, and J.I. Prosser. 2008. The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. Environ. Microbiol. 10:2966–2978.

Nicol, G.W., and C. Schleper. 2006. Ammonia-oxidising crenarchaeota: Important players in the nitrogen cycle? Trends Microbiol. 14:207–212.

Nicolaisen, M.H., J. Bælum, C.S. Jacobsen, and J. Sørensen. 2008. Transcription dynamics of the functional *tfdA* gene during MCPA herbicide degradation by *Cupriavidus necator* AEO106 (pRO101) in agricultural soil. Environ. Microbiol. 10:571–579.

Okano, Y., K.R. Hristova, C.M. Leutenegger, L.E. Jackson, R.F. Denison, B. Gebreyesus, D. Lebauer, and K.M. Scow. 2004. Application of real-time PCR to study effects of ammonium on population size of ammonia-oxidizing bacteria in soil. Appl. Environ. Microbiol. 70:1008–1016.

Ostrom, N.E., A. Piit, R. Sutka, P.H. Ostrom, A.S. Grandy, K.M. Huizinga, and G.P. Robertson. 2007. Isotopologue effects during N_2O reduction in soils and in pure cultures of denitrifiers. J. Geophys. Res. G: Biogeosci. 112:G02005.

Patra, A.K., L. Abbadie, A. Clays-Josserand, V. Degrange, S.J. Grayston, N. Guillaumaud, P. Loiseau et al. 2006. Effects of management regime and plant species on the enzyme activity and genetic structure of N-fixing, denitrifying and nitrifying bacterial communities in grassland soils. Environ. Microbiol. 8:1005–1016.

Patra, A.K., L. Abbadie, A. Clays-Josserand, V. Degrange, S.J. Grayston, P. Loiseau, F. Louault et al. 2005. Effects of grazing on microbial functional groups involved in soil N dynamics. Ecol. Monogr. 75:65–80.

Phillips, C.J., D. Harris, S.L. Dollhopf, K.L. Gross, J.I. Prosser, and E.A. Paul. 2000a. Effects of agronomic treatments on structure and function of ammonia-oxidizing communities. Appl. Environ. Microbiol. 66:5410–5418.

Phillips, C.J., E.A. Paul, and J.I. Prosser. 2000b. Quantitative analysis of ammonia oxidising bacteria using competitive PCR. FEMS Microbiol. Ecol. 32:167–175.

Poly, F., S. Wertz, E. Brothier, and V. Degrange. 2008. First exploration of *Nitrobacter* diversity in soils by a PCR cloning-sequencing approach targeting functional gene *nxrA*. FEMS Microbiol. Ecol. 63:132–140.

Prosser, J.I. 1989. Autotrophic nitrification in bacteria. Adv. Microb. Physiol. 30:125–181.

Prosser, J.I., and T.M. Embley. 2002. Cultivation-based and molecular approaches to characterisation of terrestrial and aquatic nitrifiers. Antonie Van Leeuwenhoek Int. J. Gen. Mol. Microbiol. 81:165–179.

Prosser, J.I., and G.W. Nicol. 2008. Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. Environ. Microbiol. 10:2931–2941.

Purkhold, U., A. Pommerening-Röser, S. Juretschko, M.C. Schmid, H.P. Koops, and M. Wagner. 2000. Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and *amoA* sequence analysis: Implications for molecular diversity surveys. Appl. Environ. Microbiol. 66:5368–5382.

Purkhold, U., M. Wagner, G. Timmermann, A. Pommerening-Röser, and H.P. Koops. 2003. 16S rRNA and *amoA*-based phylogeny of 12 novel betaproteobacterial ammonia-oxidizing isolates: Extension of the dataset and proposal of a new lineage within the nitrosomonads. Int. J. Syst. Evol. Microbiol. 53:1485–1494.

Rotthauwe, J.H., K.P. Witzel, and W. Liesack. 1997. The ammonia monooxygenase structural gene *amoA* as a functional marker: Molecular fine-scale analysis of natural ammonia-oxidizing populations. Appl. Environ. Microbiol. 63:4704–4712.

Schmidt, C.S., K.A. Hultman, D. Robinson, K. Killham, and J.I. Prosser. 2007. PCR profiling of ammonia-oxidizer communities in acidic soils subjected to nitrogen and sulphur deposition. FEMS Microbiol. Ecol. 61:305–316.

Schmidt, I., O. Sliekers, M. Schmid, I. Cirpus, M. Strous, E. Bock, J.G. Kuenen, and M.S.M. Jetten. 2002. Aerobic and anaerobic ammonia oxidizing bacteria competitors or natural partners? FEMS Microbiol. Ecol. 39:175–181.

Singh, S.N., and A. Verma. 2007. The potential of nitrification inhibitors to manage the pollution effect of nitrogen fertilizers in agricultural and other soils: A review. Environ. Pract. 9:266–279.

Smith, Z., A.E. McCaig, J.R. Stephen, T.M. Embley, and J.I. Prosser. 2001. Species diversity of uncultured and cultured populations of soil and marine ammonia oxidizing bacteria. Microb. Ecol. 42:228–237.

Smith, C.J., D.B. Nedwell, L.F. Dong, and A.M. Osborn. 2006. Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples. Environ. Microbiol. 8:804–815.

Stark, J.M., and M.K. Firestone. 1995. Mechanisms for soil moisture effects on activity of nitrifying bacteria. Appl. Environ. Microbiol. 61:218–221.

Stark, J.M., and M.K. Firestone. 1996. Kinetic characteristics of ammonium-oxidizer communities in a California oak woodland-annual grassland. Soil Biol. Biochem. 28:1307–1317.

Starkenburg, S.R., P.S.G. Chain, L.A. Sayavedra-Soto, L. Hauser, M.L. Land, F.W. Larimer, S.A. Malfatti et al. 2006. Genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium *Nitrobacter winogradskyi* nb-255. Appl. Environ. Microbiol. 72:2050–2063.

Starkenburg, S.R., F.W. Larimer, L.Y. Stein, M.G. Klotz, P.S.G. Chain, L.A. Sayavedra-Soto, A.T. Poret-Peterson et al. 2008. Complete genome sequence of *Nitrobacter hamburgensis* X14 and comparative genomic analysis of species within the genus *Nitrobacter*. Appl. Environ. Microbiol. 74:2852–2863.

Stein, L.Y., and Y.L. Yung. 2003. Production, isotopic composition, and atmospheric fate of biologically produced nitrous oxide. Annu. Rev. Earth Planet. Sci. 31:329–356.

Stephen, J.R., Y.J. Chang, S.J. Macnaughton, G.A. Kowalchuk, K.T. Leung, C.A. Flemming, and D.C. White. 1999. Effect of toxic metals on indigenous soil β-subgroup proteobacterium ammonia oxidizer community structure and protection against toxicity by inoculated metal-resistant bacteria. Appl. Environ. Microbiol. 65:95–101.

Stephen, J.R., G.A. Kowalchuk, M.A.V. Bruns, A.E. McCaig, C.J. Phillips, T.M. Embley, and J.I. Prosser. 1998. Analysis of beta-subgroup proteobacterial ammonia oxidizer populations in soil by denaturing gradient gel electrophoresis analysis and hierarchical phylogenetic probing. Appl. Environ. Microbiol. 64:2958–2965.

Stephen, J.R., A.E. McCaig, Z. Smith, J.I. Prosser, and T.M. Embley. 1996. Molecular diversity of soil and marine 16S rRNA gene sequences related to beta-subgroup ammonia-oxidizing bacteria. Appl. Environ. Microbiol. 62:4147–4154.

Subbarao, G., O. Ito, K. Sahrawat, W. Berry, K. Nakahara, T. Ishikawa, T. Watanabe, K. Suenaga, M. Rondon, and I. Rao. 2006. Scope and strategies for regulation of nitrification in agricultural systems—Challenges and opportunities. Crit. Rev. Plant Sci. 25:303–335.

Subbarao, G.V., M. Rondon, O. Ito, T. Ishikawa, I.M. Rao, K. Nakahara, C. Lascano, and W.L. Berry. 2007. Biological nitrification inhibition (BNI)—Is it a widespread phenomenon? Plant Soil 294:5–18.

Sutka, R.L., N.E. Ostrom, P.H. Ostrom, J.A. Breznak, H. Gandhi, A.J. Pitt, and F. Li. 2006. Distinguishing nitrous oxide production from nitrification and denitrification on the basis of isotopomer abundances. Appl. Environ. Microbiol. 72:638–644.

Taylor, A.E., and P.J. Bottomley. 2006. Nitrite production by *Nitrosomonas europaea* and *Nitrosospira* sp. AV in soils at different solution concentrations of ammonium. Soil Biol. Biochem. 38:828–836.

Tourna, M., T.E. Freitag, G.W. Nicol, and J.I. Prosser. 2008. Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. Environ. Microbiol. 10:1357–1364.

Treusch, A.H., S. Leininger, A. Kietzin, S.C. Schuster, H.P. Klenk, and C. Schleper. 2005. Novel genes for nitrite reductase and AMO-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. Environ. Microbiol. 7:1985–1995.

Webster, G., T.M. Embley, T.E. Freitag, Z. Smith, and J.I. Prosser. 2005. Links between ammonia oxidiser species composition, functional diversity and nitrification kinetics in grassland soils. Environ. Microbiol. 7:676–684.

Webster, G., T.M. Embley, and J.I. Prosser. 2002. Grassland management regimens reduce small-scale heterogeneity and species diversity of beta-proteobacterial ammonia oxidizer populations. Appl. Environ. Microbiol. 68:20–30.

Wertz, S., F. Poly, X. Le Roux, and V. Degrange. 2008. Development and application of a PCR-denaturing gradient gel electrophoresis tool to study the diversity of *Nitrobacter*-like *nxrA* sequences in soil. FEMS Microbiol. Ecol. 63:261–271.

White, R.E., R.A. Haigh, and J. Macduff. 1987. Frequency distributions and spatially dependent variability of ammonium and nitrate concentrations in soil under grazed and ungrazed grassland. Fertil. Res. 11:193–208.

Wolt, J.D. 2004. A meta-evaluation of nitrapyrin agronomic and environmental effectiveness with emphasis on corn production in the Midwestern USA. Nutr. Cycling Agroecosyst. 69:23–41.

Wrage, N., J.W. Van Groenigen, O. Oenema, and E.M. Baggs. 2005. A novel dual-isotope labelling method for distinguishing between soil sources of N_2O. Rapid Commun. Mass Spectrom. 19:3298–3306.

Wrage, N., G.L. Velthof, H.J. Laanbroek, and O. Oenema. 2004a. Nitrous oxide production in grassland soils: Assessing the contribution of nitrifier denitrification. Soil Biol. Biochem. 36:229–236.

Wrage, N., G.L. Velthof, O. Oenema, and H.J. Laanbroek. 2004b. Acetylene and oxygen as inhibitors of nitrous oxide production in *Nitrosomonas europaea* and *Nitrosospira briensis*: A cautionary tale. FEMS Microbiol. Ecol. 47:13–18.

Wrage, N., G.L. Velthof, M.L. van Beusichem, and O. Oenema. 2001. Role of nitrifier denitrification in the production of nitrous oxide. Soil Biol. Biochem. 33:1723–1732.

27.4 Denitrification

Jean-Claude Germon

Laurent Philippot

27.4.1 Introduction

The considerable increase in nitrogen fertilization in the second half of the twentieth century and the damage caused by the acceleration of the nitrogen cycle in the disturbed environment raised interest in the fate of nitrogen in soils, in particular gaseous losses and denitrification (Galloway et al., 2008). Denitrification, which was first described by Gayon and Dupetit (1882), is a microbial respiratory process consisting in the sequential reduction of nitrate (NO_3^-) or nitrite (NO_2^-) to nitric oxide (NO), nitrous oxide (N_2O), or nitrogen gas (N_2). These nitrogen oxides are used as alternative electron acceptors

when oxygen for generating a transmembrane proton electrochemical potential across the cytoplasmic membrane is limited. As a consequence, denitrification occurs in soils when nitrate or nitrite is available as a substrate and when there is little or no oxygen. Denitrification also tends to concentrate around hot spots of organic matter (Parkin, 1987), since the most common energy sources for denitrifying bacteria are organic compounds. Denitrification exhibits both a high temporal variability, generally associated with rain events, and a high spatial variability with variations by orders of magnitude even within small areas (Folorunso and Rolston, 1984).

Denitrification is of interest for several reasons. Firstly, from an agronomical point of view, denitrification is responsible for the loss of an essential plant nutrient as the reduction of NO_3^- to gaseous nitrogen depletes the soil of NO_3^-. It has been reported that between 0% and 25% of the applied nitrogen can end up as nitrogen gas or N_2O, thus limiting crop production (Aulakh et al., 1992; DeKlein and van Logtestijn, 1994; Mogge et al., 1999). From an environmental point of view, denitrification contributes to global warming by emitting N_2O, which is one of the major greenhouse gases. Nitrous oxide has a global warming potential 300 times higher than that of CO_2 as estimated by the Intergovernmental Panel on Climate Change (IPCC) and contributes up to 6% of the anthropogenic greenhouse effect (Cicerone, 1989). Both N_2O and NO are also involved in the destruction of the earth's ozone layer (Waibel et al., 1999). From a human health point of view, denitrification is of interest as it removes NO_3^- from water. There is a potential health problem primarily for young children if the nitrate concentration rises above the levels acceptable for drinking water quality (World Health Organization standard of $10\,mg\,L^{-1}\,NO_3^-$–N). The reduction of NO_3^- to NO_2^- by bacteria in the digestive tract can lead to a disorder called methemoglobinemia, as, after entering the blood stream, NO_2^- reacts with hemoglobin, reducing the oxygen-carrying capacity and causing the asphyxiation of new born babies (Shearer et al., 1972). The ingestion of high levels of NO_3^- may also result in the formation of carcinogenic nitrosamines by the reaction of ingested secondary aliphatic and aromatic amines with NO_2^- in the human stomach.

Despite these effects of denitrification on agriculture and the environment, which are vital for human welfare, many aspects of this process are still poorly understood. There is a lack of interdisciplinary collaboration between scientists, which is an obstacle to providing a comprehensive understanding of denitrification. Significant progress could be achieved through an integrative approach linking field measurements of denitrification N flows to the genetics, biochemistry, physiology, and ecology of the denitrifiers.

27.4.2 Denitrification Pathway

The complete denitrification process consists of a sequential process of nitrate respiration, nitrite respiration, nitric oxide reduction, and nitrous oxide respiration (Figure 27.11). The first step is the reduction of nitrate to nitrite, which is catalyzed by
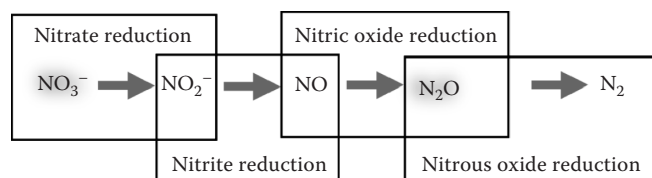
FIGURE 27.11 The four steps of the denitrification pathway.

the membrane-bound nitrate reductase (Nar) or the periplasmic nitrate reductase (Nap). The two types of nitrate reductases can be present in the same strain of bacteria (Carter et al., 1995; Roussel-Delif et al., 2005). The second step of the denitrifying pathway is a key step since it consists in the reduction of soluble nitrite to gaseous nitric oxide. This reaction distinguishes denitrification from the dissimilatory reduction of nitrate to ammonium and can be catalyzed either by a copper nitrite reductase (NirK) or a cytochrome cd_1 nitrite reductase (NirS) (Zumft, 1997). Although the structure and prosthetic metal of these enzymes are different, transfer of the gene encoding copper nitrite reductase to a mutant of *Pseudomonas stutzeri* lacking cytochrome cd_1 nitrite reductase due to a transposon Tn5 insertion in its structural gene can restore the denitrification pathway lost in the mutation (Glockner et al., 1993). All known denitrifiers possess only one of these two types of nitrite reductase. The next step, reduction of the toxic intermediate nitric oxide into nitrous oxide, is also performed by two types of enzyme: cNor or qNor. The first type of nitric oxide reductase (cNor) receives the electrons from cytochrome *c* or pseudoazurin and the second (qNor) receives the electrons from a quinol pool. The last step of the denitrification pathway, the reduction of the greenhouse gas nitrous oxide to dinitrogen, is catalyzed by the multicopper homodimeric nitrous oxide reductase (Nos) (Zumft, 1997).

The denitrification pathway is best described as the modular organization of the different steps. Denitrifiers may have a truncated denitrification pathway as was revealed by the complete genome sequencing of *Agrobacterium tumefaciens* C58, which lacks the *nosZ* gene encoding the catalytic subunit of the nitrous oxide reductase and is, therefore, genetically unable to reduce nitrous oxide. In addition, a few microorganisms cannot reduce nitrate and use nitrite as the first electron acceptor in the denitrification process. The distribution of the various denitrification reductases also varies between denitrifiers (Philippot, 2002). For example, the first two steps of the denitrification pathway are catalyzed by Nap and NirK in many denitrifying rhizobia while it is catalyzed by both Nar and Nap and NirS in *P. aeruginosa* strains.

In most studies, the expression of the denitrification reductases could be induced by a low O_2 partial pressure and presence of nitrogen oxides. In pure cultures, synthesis of the denitrification reductases after a shift to conditions favorable to denitrification can be achieved within a few hours (Baumann et al., 1996). Other factors such as soil pH can affect the expression or the activity of these enzymes. Environmental conditions can also have different effects on the reductases catalyzing the various

steps of the denitrification pathway, resulting in the accumulation of intermediates (Firestone et al., 1980; Thomsen et al., 1994). This is highly significant for the environment as it can result in changes in the N_2/N_2O ratio with increased emission of the greenhouse gas N_2O (see Section 27.4.5).

27.4.3 Denitrifying Organisms

27.4.3.1 Abundance and Diversity of Denitrifiers

When studying the respiratory capacities of bacteria isolated from various environments, many strains have been wrongly identified as denitrifiers. Mahne and Tiedje (1995), therefore, proposed two main criteria for identifying "true" denitrifiers: (1) nitrous oxide and/or nitrogen gas must be the main end product of nitrate or nitrite reduction and (2) this reduction must be coupled with a greater increase in growth than would occur if the nitrate or nitrite were simply acting as electron sinks. These criteria can be used to distinguish denitrifiers from bacteria reducing nitrate into ammonium (Cole, 1996) and bacteria able to reduce nitric oxide as a protection against nitrosative stress (Philippot, 2005).

Many soil prokaryotes can denitrify and the proportion of denitrifiers within the soil microbial community was estimated, using both culture based and direct molecular approaches, to be up to a few percents, giving a range of 10^5–10^8 denitrifiers per gram of soil (Tiedje, 1988; Vinther et al., 1999; Henry et al., 2004, 2006; Dandie et al., 2008; Hallin et al., 2009). The diversity of microorganisms able to denitrify is also very large, with more than 60 bacterial genera containing denitrifying strains as well as some archaea and fungi (Philippot et al., 2007). In one of the first comprehensive studies exploring denitrifying communities in soil, Gamble et al. (1977) found that the dominant isolated denitrifiers belonged to the *Pseudomonas* and *Alcaligenes* groups. In addition to these genera, other cultivation-based studies showed that the genera *Achromobacter*, *Ralstonia*, *Burkholderia*, *Bacillus*, and *Streptomyces* were numerically significant denitrifiers in soils (Garcia, 1977; Weier and Mac Rae, 1992; Chèneby et al., 2000, 2004; Dandie et al., 2007). Furthermore, the number of taxa known to include denitrifiers is likely continuing to increase as yet unknown soil microorganisms may also contribute to the overall denitrification process. As a consequence of this wide taxonomical diversity of denitrifiers, the distribution of the denitrification trait among microorganisms cannot be predicted using their taxonomical affiliation. While distantly related microorganisms can denitrify, closely related strains can exhibit very different respiratory pathways. For example, analysis of the ability to use nitrate as alternative electron acceptor among a collection of fluorescent pseudomonads showed that strains were denitrifiers, nitrate reducers, or neither (Clays-Josserand et al., 1995). Denitrifiers can also be involved in the other steps of the nitrogen cycle, such as nitrification or nitrogen fixation. For example, ammonia-oxidizing *Nitrosospira* and *Nitrosomonas* strains and several nitrogen-fixing rhizobia have been shown to be denitrifiers (Daniel et al., 1982; Bock et al., 1995; Shaw et al., 2006).

27.4.3.2 Ecology of Denitrifiers

There were very few studies of the ecology of denitrifiers in soil and other functional communities until the end of the 1990s owing to the lack of appropriate tools for targeting them in this complex environment. Molecular tools are now being developed or refined to assess both diversity and size of the denitrifying community in soil and other ecosystems (Taroncher-Oldenburg et al., 2003; Throbäck et al., 2004; Henry et al., 2006). The most common way of analyzing the soil denitrifier community is to extract the DNA from the soil and amplify it by PCR (Hallin et al., 2007). Since the ability to denitrify is not associated with any specific taxonomic group (Section 27.4.3.1), existing primers for studying denitrifiers target the genes encoding the various denitrification reductases, which are used as molecular markers for this community. Several techniques are available for analyzing the sequence polymorphism of PCR-amplified denitrification genes. The most informative technique is to clone and sequence the PCR amplicons, but a more rapid analysis is achieved using fingerprinting techniques (denaturing gradient gel electrophoresis [DGGE], terminal restriction fragment length polymorphism [T-RFLP], etc.). The polymorphism of the denitrification genes is used as a proxy for the level of diversity and composition of the denitrifier community. The size of the denitrifier community is now usually evaluated as the gene copy numbers per gram of soil using real-time quantitative PCR of the denitrification genes (Philippot and Hallin, 2005). Since the number of the denitrification gene copies per genome is between 1 and 3 but usually only 1, the gene copy numbers can be then converted into cell numbers. However, the major limitation of these methods is that they target all bacteria capable of denitrification and therefore do not distinguish those that are actively denitrifying at the time of sampling from the nondenitrifying fraction of the denitrifier community. The fact that these methods target potential denitrifiers is also a major drawback for the understanding of the relationships between the diversity and activity of the denitrifying community. Over the last 15 years, these molecular methods have mainly been applied to study how the denitrifying community in soil is affected by agricultural practices or environmental conditions, or more generally, to identify the factors controlling the size and the composition of this functional community.

The composition of the denitrifier community varied significantly between soil habitats (Cavigelli and Robertson, 2001; Priemé et al., 2002; Rich et al., 2003; Rich and Myrold, 2004). For agricultural soils, the main focus has been on the effect of fertilization on denitrifiers. It has been shown that the composition of the denitrifier community may be modified after a few weeks by high ammonium concentrations varying from 60 to 400 mg NH_4^+–N g^{-1} of dry soil (Avrahami et al., 2002). Organic fertilization can also modify the denitrifier community composition depending not only on the fertilization regime but also on the duration for which it is applied (Wolsing and Prieme, 2004; Dambreville et al., 2006; Schauss, 2006). In a long-term study on the impact of the fertilization regime on denitrifiers in an experimental field amended with calcium nitrate, ammonium sulfate, cattle manure, and sewage sludge, fingerprint analyses showed the highest differences in the denitrifier community structure in plots treated with ammonium sulfate and sewage sludge, which exhibited the lowest pH (Enwall et al., 2005). When analyzing the size of the denitrifier community by quantitative PCR, the numbers of denitrification genes encoding the nitrite and nitrous oxide reductases were one to two orders of magnitude lower in the ammonium sulfate treatment compared with the other treatments, thus confirming the strong impact of soil pH on this functional community (Hallin et al., 2009).

The presence of plants can also influence the denitrifying community since the availability of all the major factors regulating denitrification—organic compounds, oxygen and nitrate—are directly and indirectly modified by the plant rhizosphere. Changes in the composition of the denitrifier community in the plant rhizosphere compared to the bulk soil have been reported in several studies (Chèneby et al., 2004). Plant species can also influence the composition of the denitrifier community in soil (Sharma et al., 2005; Bremer et al., 2007), probably through specific rhizodeposition (Mounier et al., 2004; Henry et al., 2008).

Recent studies showed that, similar to denitrification activity, the size and composition of the denitrifier community had a high spatial variability at the field scale with macroscale spatial dependence (several meters) (Philippot et al., 2009; Enwall et al., 2010). Characterization of denitrifier distribution patterns at field scale is an important step in modeling distribution of functional bacterial communities at a scale compatible with land management strategies and can help in the future bridge the gap between studies analyzing denitrification N flows and denitrifier community ecology.

27.4.4 Denitrification Activity in Soils

27.4.4.1 Methods of Measuring Denitrification Activity in Soil

Denitrification in soil was first measured by Dehérain and Maquenne (1882) who observed that nitrate reduction producing an N_2O and N_2 mixture occurred when the soil was enriched with organic matter and incubated in an oxygen-free atmosphere. Dehérain (1897) noted that this mechanism constitutes a nitrogen loss for growing plants but that its quantitative evaluation in soils is hindered by three main problems: (1) the simultaneous formation and reduction of nitrates, (2) the nitrogen gas produced is naturally diluted in the atmospheric nitrogen, and (3) the level of nitrogen loss is generally of the same order as the uncertainty of the total nitrogen measurements in soils. Denitrification is, therefore, still considered as a difficult process to measure (Groffman et al., 2006) and requires different methodological approaches for an accurate evaluation of the rates at different scales.

The first quantitative denitrification estimates in soil were performed using various nitrogen budgets and balances. Allison (1955) reviewed the N balances from several long-term lysimeter

field experiments and concluded that gaseous nitrogen losses from normal, well-aerated soils are not negligible and can represent 20% of the available mineral nitrogen during the year in cultivated, fertilized soils and 12% in uncultivated soils. From the previous synthesis, Broadbent and Clark (1965) drew the conclusion that it is difficult to reduce nitrogen losses by denitrification in cultivated fields to less than 15% of the mineral nitrogen available during the cultivation cycle. The main advantage of the N balance methodology is that it can be applied over a large range of scales, from plots to basins (Billen and Garnier, 1999), regions, countries (David et al., 2001), and open marine systems. This approach can be applied to evaluating the global N cycle to estimate the denitrification rate required to balance the cycle (Galloway et al., 2004). However, such balancing must take account of the various, major uncertainties in estimating the various N sources and sinks and changes in N storage pools (Kroeze et al., 2003). Nitrogen losses by denitrification estimated using mass balances must, therefore, be verified by direct measurements (Groffman et al., 2006).

Different nitrogen compartments can be distinguished using ^{15}N-labeled compounds. This improves the N balance sensitivity and makes it possible to measure denitrification directly in laboratory conditions by accumulation of labeled ^{15}N gaseous products (Wijler and Delwiche, 1954; Nömmik, 1956; Guiraud and Berliet, 1970). This methodology requires the added labeled nitrate to be homogeneously and rapidly mixed with the endogenous nitrate in the soil and an exact knowledge of the $^{15}N/^{14}N$ ratio in the denitrified nitrate to assay denitrified N using ^{15}N-labeled gases. Hauck et al. (1958) demonstrated that the ratio of the enriched species $^{15}N^{15}N$ and $^{14}N^{15}N$ can be used to estimate the $^{15}N/^{14}N$ ratio in the denitrified source. More direct measurements of the isotope mixing ratio can be obtained by an accurate isotope analysis of N_2O produced as an intermediate denitrification product (Stevens and Laughlin, 2001). In situ ^{15}N-labeled fertilization was used in many field experiments to estimate the N balance deficit at the end of the vegetation cycle, based on inorganic and organic nitrogen measurements. This was preferentially assigned to denitrification rather than to other gaseous loss mechanisms (Broadbent and Carlson, 1978; Dowdell and Webster, 1984; Recous et al., 1988; Addiscott and Powlson, 1992). However, this method is affected by measurement uncertainties on the different N pools analyzed (Myrold, 1991) and by the possible N losses by emission mechanisms other than denitrification. In situ denitrification can also be estimated by the aid of highly ^{15}N-labeled fertilizer applications and labeled gas emission measurements: This method requires high-performance mass spectrometry to measure the various diluted isotopic compounds produced and is not very sensitive: It was, however, sufficiently reliable to be considered as a standard way of comparing different methods (Mosier et al., 1986; Aulack et al., 1991). ^{15}N isotopes ($^{15}NO_3^-$ and $^{15}NH_4^+$) can be used as tracers to evaluate the contribution of nitrification and denitrification in soils to gaseous nitrogen oxide emission (Stevens et al., 1997; Russow et al., 2000; Wolf and Russow, 2000).

Mathieu et al. (2006) observed that, under unsaturated and saturated conditions, both processes were significantly involved in N_2O production, with 60% of $N–N_2O$ coming from nitrification under unsaturated conditions and 85%–90% of $N–N_2O$ coming from denitrification under saturated conditions. The proportion of nitrified N emitted as N_2O changed from 0.13% to 2.32% in unsaturated and saturated conditions, respectively.

The natural isotopic abundance variations can be used as an indicator of denitrification by assessing nitrogen isotopic fractionation and enrichment factors during this N transformation (Mariotti et al., 1982; Shearer and Kohl, 1988): They can allow estimating denitrification intensity in soils and aquifers such as basins, riparian, or benthic zones and rivers (Lehmann et al., 2003; Sebilo et al., 2003, 2006). However, isotopic fractionation can be affected by other N transformations such as nitrification or by diffusional constraints conditioning NO_3^- availability. Consequently, the magnitude of the isotope effect in a physically heterogeneous medium such as soil may not reflect the amount of NO_3^- reduced (Groffman et al., 2006).

The discovery of the inhibiting properties of acetylene (C_2H_2) on N_2O reduction by denitrifying bacteria (Federova et al., 1973; Balderston et al., 1976; Yoshinari and Knowles, 1976) paved the way for a significant development of techniques for direct measurement of soil denitrification in situ (Ryden et al., 1979a, 1979b; Mosier et al., 1986; Aulack et al., 1991) or in laboratory conditions (Klemedtsson et al., 1977; Yoshinari et al., 1977; Smith et al., 1978). Various acetylene-based methods have been used to measure in situ denitrification in agriculture or forest soils. The first was based on the use of in situ chambers and acetylene diffusion in soil (Ryden et al., 1979b; Rolston et al., 1982; Colbourn et al., 1984). This method was time-consuming, and it was difficult to ensure that the acetylene was distributed evenly throughout the pores in the soil. It has therefore been little used. The majority of measurements of denitrification in field conditions using acetylene inhibition are based on the use of undisturbed soil cores sampled in the field and incubated in laboratory conditions as close as possible to field conditions. Using either recirculating or static soil cores, this method provides a better control of acetylene distribution and other parameters regulating the denitrifying activity (Parkin et al., 1984; Ryden and Skinner, 1987; Hénault and Germon, 2000; Ullah et al., 2005; Van der Salm et al., 2007).

This method, however, suffers from several problems making its use controversial in various situations. The most important are (1) The inhibitory effect of C_2H_2 on ammonia oxidation by autotrophic microflora (Hynes and Knowles, 1978), which can decrease NH_4^+ conversion to NO_3^- and limit the denitrification rate in soils with a low nitrate content (Mosier, 1980); (2) possible microflora adaptation in some soils, biodegrading C_2H_2 after several days in a C_2H_2-enriched atmosphere leading to increased soil respiration and denitrification (Germon, 1980; Terry and Duxburry, 1985; Topp and Germon, 1986); and (3) an overestimated denitrification rate due to the lower electron flow required to reduce nitrate to N_2O rather than to N_2 (Myrold, 1991).

However, a comparison of N loss data obtained by C_2H_2 methods and by isotopic tracers indicated similar results with differences generally within the measurement variability (Parkin et al., 1985a; Myrold, 1991). In the absence of a viable alternative method and despite these limitations (Groffman et al., 2006), this approach continues to be a standard tool for evaluating denitrification in many situations (Ullah et al., 2005; Van der Salm et al., 2007; Woodward et al., 2009).

27.4.4.2 Regulation of Denitrification Activity in Soil

The denitrification rate in soil results from the activity of the denitrifying community, which is regulated by various environmental factors. Most works have indicated that the denitrifying activity is generally not limited by the lack of denitrifiers (see Section 27.4.3.1). This functional community is generally abundant in the upper soil horizons (Gamble et al., 1977; Chéneby et al., 1998; Hallin et al., 2009) and is significant in deeper horizons (Weier and Mac Rae, 1992), allowing denitrification in subsoils and aquifers (Mariotti, 1994; Hashimoto et al., 2005). In field conditions, denitrification rates exhibit considerable spatial and temporal variability (Folorunso and Rolston, 1984), which is the consequence of simultaneous regulation by a range of environmental factors, so that the activity has hot spots and hot moments. This variability must be taken into account in denitrification estimates, particularly in denitrification models (Groffman et al., 2009). Groffman et al. (1987) and Tiedje (1988) designed a conceptual schema with hierarchical control by environmental parameters, distinguishing between proximal controls (carbon, nitrate, oxygen) and distal controls (humidity, temperature, soil respiration, plants, etc.).

The first condition for denitrifying activity in soils is the presence of nitrogen oxidized forms, mainly nitrate or nitrite. The nitrate availability for denitrifiers depends on the rate of nitrate production and nitrogen fertilizer application, nitrate transport from solution in the soil to denitrifying microsites, and the rate of nitrate consumption by plants or microbes. The denitrification rate appears to be seriously limited by nitrate availability in unfertilized or slightly fertilized soils. However, in the majority of fertilized agricultural soils, denitrification is not or only slightly limited by nitrate availability (Barton et al., 1999). The regulatory effect of nitrate concentration on denitrification rate can be described by Michaelis–Menten kinetics with an apparent K_m dependent on the soil properties and generally of several orders of magnitude higher than the K_m measured on microorganisms. While a theoretical K_m in soil could be estimated from microbial cultures on a possible range of 0.01–0.1 mg N kg^{-1} dry soil (Myrold and Tiedje, 1985), K_m values higher than 20 mg N kg^{-1} dry soil have been observed in undisturbed soil cores (Hénault and Germon, 2000; Heinen, 2006a; Van der Salm et al., 2007).

Oxygen availability is generally considered to be the most important parameter regulating denitrification as oxygen inhibits the activity and neosynthesis of most denitrifying enzymes. Denitrification activity in apparently well-aerated soils is the consequence of the soil structure heterogeneity and the variable transfer of oxygen through the soil pores. Parkin and Tiedje (1984) observed a dramatic decrease in the denitrification rate in soil when the O_2 content in macropores is higher than 0.5%, while Sexstone et al. (1985) described a steep O_2 gradient in soil aggregates incubated in air and demonstrated that anaerobic centers were present in all aggregates that showed denitrification while not all aggregates that had anaerobic zones showed denitrification. The effect of aeration and oxygen transfer on soil denitrification has been described by various authors (Smith, 1980; Arah, 1990; Sierra et al., 1995). However, aeration in soil is strongly dependent on water content and, in particular, on water-filled pore space (WFPS). Linn and Doran (1984) described soil respiration, nitrification, and denitrification in relation to WFPS while Grundman and Rolston (1987) defined an exponential relationship between field denitrification rates and WFPS. This type of relationship has been confirmed by many authors and has been used in several models of soil denitrification (Johnsson et al., 1991; Hénault and Germon, 2000; Heinen, 2006a).

Denitrifiers are mainly heterotrophic organisms and the organic matter content was the first factor identified as limiting denitrifying activity in soils (Dehérain and Maquenne, 1882; Bremner and Shaw, 1958). Available organic carbon increases heterotrophic respiration in aerobic soils, decreasing soil oxygen concentrations and creating anaerobic microsites for denitrifiers. Significant relationships were reported between denitrifying activity and various forms of available carbon (Burford and Bremner, 1975; Standford et al., 1975; Reddy et al., 1982). Among the denitrification regulating factors, organic matter content is still considered to be the main control, suggesting that it may be possible to estimate the potential denitrifying activity directly from the soil carbon content or from the kinetics of carbon biodegradation (Hénault and Germon, 2000; Heinen, 2006a).

The pH is frequently reported as being another factor affecting denitrifying activity (Parkin et al., 1985b; Simek and Hopkins, 1999). Soil denitrifying activity is greatest at neutral pH and is significantly reduced at a pH lower than 5 or higher than 8.5. A limiting function of the pH on the denitrifying activity that can be used in denitrification models was proposed by Heinen (2006a).

27.4.4.3 Potential and In Situ Denitrification Activity

The effect of regulating parameters and the great variability of soil denitrification activity means that the soil potential activity measured in standard laboratory conditions must be distinguished from the real or in situ activity determined in natural or field conditions. A potential denitrifying activity measurement, or denitrifying enzyme activity (DEA), was proposed by Smith and Tiedje (1979) as the rate of N_2O production in the presence of acetylene from soil samples placed under anaerobic conditions and supplied with excess carbon and nitrate sources. These authors suggested that chloramphenicol should be used in DEA assays to inhibit synthesis of new denitrifying enzymes. This method in a standard form (Tiedje et al., 1989) is used with variations. Glucose can be replaced by a more efficient carbon source, such as succinate, for the denitrifying community (Heylen et al., 2006).

Chloramphenicol can disrupt the existing denitrifying enzymes and result in an underestimation of their activities (Brooks et al., 1992). Pell et al. (1996), therefore, proposed a method with different chloramphenicol concentrations to extrapolate the rate in the absence of this inhibitor while Murray and Knowles (1999) concluded that DEA essays should be carried out with just one, sufficiently low concentration. Other works demonstrate that DEA is not significantly affected by chloramphenicol (Simek et al., 2002; Yu et al., 2008) and confirm that DEA can be estimated using a short incubation without this inhibitor (Parry et al., 1999). Measured on soil slurries, DEA gives generally high values varying from 1 to 100 mg N kg^{-1} dry soil day^{-1} corresponding to several kg or hundreds of kg N ha^{-1} day^{-1} with a low variability for a given soil, while the denitrification rates in field conditions or on undisturbed soil cores are two to three orders of magnitude lower with a much larger variability (Parkin et al., 1987; Ullah et al., 2005).

Denitrifying activity can be usefully evaluated on undisturbed soil samples with water-saturated pore space and no limiting nitrate content in the absence of an added source of carbon. This can be considered to be the potential denitrifying rate of the microbial community in a soil in an undisturbed or slightly disturbed structural state, depending on the endogenous carbon compounds and organic matter mineralization in the soil (Heinen, 2006a). This type of activity used in denitrification models is generally significantly lower than the DEA activity measured on soil slurries. Burford and Bremner (1975) and Van der Salm et al. (2007) measured denitrification rates varying from 2 to 55 mg N kg^{-1} dry soil day^{-1} and from 10 to 70 mg N kg^{-1} dry soil day^{-1} on wet and homogenized soils using protocols similar to DEA, while Hénault and Germon (2000) estimated values on undisturbed and water-saturated soil cores with the nitrate not a limiting factor, varying from 1 to 30 kg N ha^{-1} day^{-1} corresponding to 0.4 to 12 mg N kg^{-1} dry soil day^{-1}.

27.4.4.4 Evaluation of Denitrification N Losses in Soils: Measurements and Models

An accurate evaluation of N losses due to denitrification is essential for understanding N cycling in the environment and for optimizing the use of nitrogen fertilizers in agriculture. There are two main approaches for evaluating N losses by denitrification: (1) Local and repeated gaseous loss measurements with integration over space and time. In order to take into account the temporal variability, denitrification losses are usually integrated over 1 year. (2) Estimation of the denitrification using models taking account of the regulating factors.

In a comprehensive synthesis of in situ denitrification measurements using acetylene, Barton et al. (1999) reported that denitrification in uplands with agricultural and forest soils varied between 0 and 239 kg N ha^{-1} $year^{-1}$ with a clear distinction between the two types of soils. Annual denitrification rates in forest soils ranged from 0 to 40 kg N ha^{-1} $year^{-1}$ with over 50% of the rates being lower than 1 kg N ha^{-1} $year^{-1}$, 80% lower than 10 kg N ha^{-1} $year^{-1}$, and a geometrical mean rate of 1.9 kg N ha^{-1} $year^{-1}$. In agricultural soils, denitrification losses varied from

0 to 239 kg N ha^{-1} $year^{-1}$ with only 10% of the rates being lower than 1 kg N ha^{-1} $year^{-1}$, 20% higher than 50 kg N ha^{-1} $year^{-1}$, and a geometrical mean rate of 13 kg N ha^{-1} $year^{-1}$. These orders of magnitude are confirmed by recent data on Dutch soils (Van der Salm et al., 2007) and by Hofstra and Bouwman's (2005) review using measurements based on chamber, soil core, and N balance techniques, with denitrification rates in agricultural soils from 0 to 341 kg N ha^{-1} $year^{-1}$, and higher values from the N balance techniques than from soil core measurements using acetylene.

In order to avoid meticulous and time-consuming denitrification measurements, a number of predictive models have been developed to evaluate denitrification losses in a range of situations. Some mechanistic models are useful for understanding denitrification mechanisms at a small scale (Smith, 1980; Sierra et al., 1995, Standing et al., 2007) but are inappropriate for evaluating denitrification in varied ecological situations. Many attempts have been made to evaluate denitrification at field scale using a statistical approach (Myrold, 1988; Jarvis et al., 1991; Weier et al., 1991), but these models are inadequate when used in situations other than those for which they were developed.

Simplified process models appear to be the most effective models for predicting denitrification in soils and more than 50 models of this type have been described (Heinen, 2006a). Rolston et al. (1984) proposed a simplified process model, which has been applied by several groups with some modifications (Johnsson et al., 1991; Hénault and Germon, 2000; Van der Salm et al., 2007). The NEMIS model, developed by Hénault and Germon (2000), has the general form: $D_A = D_P F_N F_W F_T$, where D_A is the actual denitrification rate (kg N ha^{-1} day^{-1}), D_P is the potential denitrification rate on undisturbed soil cores, and F_N, F_W, and F_T are dimensionless functions taking into account the effects of soil nitrate content, the WFPS, and the soil temperature. Values of this potential denitrification rate D_P include the availability of C, avoiding a specific function for available C. In this approach, D_P is defined by experimental measurements on soil cores and is assumed to be constant over the year or for part of the year during which organic matter availability does not change. The nitrate function F_N is defined as an apparent Michaelis–Menten function, F_W is an exponential function of WFPS as defined in accordance with Grundman and Rolston (1987) with a threshold allowing denitrification, and F_T is a classic van't Hoff temperature function. The NEMIS model can be applied subject to two main limiting conditions: (1) It requires the determination of a potential denitrification rate on undisturbed soil cores, which could be avoided if a clear relationship could be established between the potential denitrification rate and organic matter content and (2) NEMIS is highly sensitive to the F_W function: The main point of dispute over the validity of this water function in varying situations concerns the WFPS threshold allowing denitrification activity. Such models need to be calibrated before use, but they can be a powerful tool for evaluating denitrification in different situations (Heinen, 2006b) taking account of the variations due to the regulating factors over the year (Van der Salm et al., 2007). However, this type of model does not take into account the spatial variability characterized by hot spots of denitrifying

activity. Considering such a variability remains a key challenge for modeling denitrification, even if analysis methods for variations and spatial structures can be developed for denitrification activity (Groffman et al., 2009).

27.4.5 Greenhouse Gas Emissions by Denitrification

Denitrification is involved in nitrous oxide (N_2O) and nitric oxide (NO) emissions, two gases contributing to the greenhouse effect. Nitrous oxide is a potent and long-lived greenhouse gas (Prather et al., 1995). Its contribution to radiative forcing is estimated to $0.16 \pm 0.02\,W\,m^{-2}$ (IPCC, 2007), and its concentration in the atmosphere continues to increase at a rate of 0.2%–0.3% year^{-1} with an atmospheric concentration in 2005 of 319 ± 0.12 ppbv (IPCC, 2007). Nitric oxide and its derived product NO_2 are tropospheric precursors of ozone, the third largest contributor to positive radiative forcing. In the troposphere, NO and NO_2 concentrations are very variable, ranging from natural concentrations of a few ppb to 2 orders of magnitude higher in highly polluted environments (IPCC, 2001; Wang and Lu, 2006).

In soils, N_2O and NO are mainly produced by denitrification and by nitrification (Firestone and Davidson, 1989), while denitrification is also a sink for atmospheric N_2O. The same environmental factors affect the production of both nitrogen oxides (Davidson et al., 2000), but the respective contribution of denitrification and nitrification varies with these factors, mainly with aeration. In aerobic conditions nitrification is the main process producing both gases formed as by-products proportionally to the nitrification rate (Li, 2000) by a poorly understood pathway. In anaerobic conditions, NO and N_2O are produced by denitrification, which generally occurs in soil when the soil WFPS is higher than 60% (Davidson, 1991) and accounts for all the N_2O produced when the soil WFPS is 70% or higher (Bateman and Baggs, 2005). The ratio between these two gases can vary depending on the soils and the environmental conditions (Li, 2000; Garrido et al., 2002) with a significant relationship between the NO/N_2O ratio and WFPS (Davidson et al., 2000). Well-aerated soils are favorable to NO emissions, while wetter soils preferentially emit N_2O, which can be reduced into N_2 by denitrifying bacteria in anaerobic conditions. Among the environmental factors regulating denitrification, low pH and high nitrate concentrations are known to increase N_2O emissions while the relative availability of C and NO_3^- influence both the denitrification rate and the N_2O/N_2 ratio (Miller et al., 2008).

Emissions of N_2O and NO from soils can be directly evaluated by micrometeorological methods (Jambert et al., 1997; Laville et al., 1997; Magiotto et al., 2000) or by chamber methods, measuring the concentration of N_2O by gas chromatography with electron capture detection and NO by chemiluminescence (Garrido et al., 2002). Several models have been built using these data to forecast N_2O and NO emissions at different scales. While a statistical approach makes it possible to evaluate emissions using emission factors (Bouwman, 1996; Mosier et al., 1998; Boeckx and Van Cleemput, 2001; Bouwman et al., 2002a) as

in the IPCC recommendations for assessing national N_2O emissions (IPCC, 2006), N_2O and NO emissions can be also evaluated using other models based on a more mechanistic approach taking into account nitrification and denitrification kinetics and regulatory factors (Potter et al., 1997; Davidson et al., 2000). These models should provide better tracking of processes affecting gas production/consumption for drawing up mitigation methods at field or regional scales (Chen et al., 2008).

Among the most frequently used models for N_2O emissions, the NGAS model proposed by Mosier and Parton (1985) for predicting daily N_2O losses was initially a simple mechanistic model accounting for the effect of water, nitrate, and ammonium content in the soil on these emissions. It was developed to become NGAS-DAYCENT, the daily time step version of the CENTURY ecosystem model (Parton et al., 1996, 1998), and can simulate NO, N_2O, and N_2 trace gas fluxes.

The DNDC (DeNitrification–DeComposition) model proposed by Li et al. (1992) was initially developed for N_2O fluxes and CO_2 production and was extended to NO, CH_4, and NH_3 emissions (Li, 2000). It is mainly based on nitrification and denitrification kinetics and consists of two components taking into account the impact of regulating factors: (1) The first component includes climate, crop growth, and decomposition submodels predicting the changes in physical parameters and substrate concentration profiles based on ecological drivers; (2) the second component consists of nitrification, denitrification, and fermentation submodels and predicts trace gas fluxes based on soil environmental variables.

The N_2O emission (NOE) model proposed by Hénault et al. (2005) is a semiempirical algorithm that simulates the production and reduction of N_2O in agricultural soils through both the denitrification and the nitrification pathways. The advantage of this model is that it takes into account measurable biological N transformation rates in soils and only requires a limited number of equations and site-specific parameters. This model was validated using different data sets and integrated into a CERES crop model for regional estimation of N_2O emission (Gabrielle et al., 2006a, 2006b).

Davidson et al. (2000) proposed a model of soil emissions of N_2O and NO based on the "hole-in-the-pipe" (HIP) model (Firestone and Davidson, 1989) using two functions based on soil nitrogen availability and soil water content. Using a relationship between N_2O/NO ratio in emitted gas and soil WFPS they characterize a large part of the observed variation of these emissions.

Global N_2O emissions are presently evaluated at 17.7 Tg N–N_2O year^{-1}, with 11 Tg from natural sources and 6.7 Tg from anthropogenic sources (IPCC, 2007). These estimates with large uncertainties are likely to be revised just as previous estimates were (IPCC, 2001). However, they underline the important contribution of agricultural systems, which amounts to 4.1 Tg N–N_2O year^{-1}, with an emission factor estimated at 0.91% of the applied nitrogen (Bouwman et al., 2002b). The emissions estimated using this factor account for only a part of total emissions (Crutzen et al., 2007), and the respective contributions of denitrification and nitrification to the total emissions are still uncertain.

A similar global inventory of nitric oxide emissions from soils was established by Davidson and Kingerlee (1997) by dividing the surface of the earth into biomes. They concluded that the best current estimate of the global soil source was 21 Tg N year^{-1}, confirming the significant contribution of this soil source to global emissions and the largest emissions for tropical regions at 15.7 Tg N year^{-1}. A possible adsorption of NOx onto plant canopy surfaces might reduce these emissions to the atmosphere to 13 Tg N year^{-1}. The agricultural system contribution estimated at 5 Tg year^{-1} is much higher than the 1.8 Tg established by Stehfest and Bouwman (2006). The contribution of denitrification to these emissions is considered lower than nitrification but needs to be more precisely estimated.

27.4.6 Managing Denitrification in Soils

There are three main reasons for denitrification management in soils: (1) inhibiting or limiting denitrification as it contributes to N loss for agriculture and emission of greenhouse gases, (2) increasing denitrification as this helps by removing nitrates, which can contaminate surface and groundwater, and (3) reducing the N$_2$O/N$_2$ ratio of the denitrification end products to lower the emissions of N$_2$O.

The inhibitory effect of various compounds on soil denitrification has been tested (Heinemeyer, 1979; Yeomans and Bremner, 1985) without any significant applications. The consistency between denitrification estimates by direct measurement and N balance (Van der Salm et al., 2007) implies that the optimization of N fertilization and land management are the best ways to limit denitrification losses by (1) adjusting available N in soils to that required by the plants, (2) maintaining the unused nitrogen in the soil, (3) keeping the mineral nitrogen content in soils as low as possible during fallow periods, (4) managing the soil to avoid the conditions favorable to denitrification. Various methods have been developed to optimize N fertilization and limit N losses, mainly based on the N balance sheet and N mineral measurements (Meynard et al., 2002; Alvarez et al., 2004; Burns, 2006). The implementation of these methods can significantly reduce excessive N fertilization and N losses (Ju et al., 2009). However, in many situations, even with well-adjusted N fertilization, it is impossible in practice to keep the nitrate content in the leached water from cultivated soils below 50 mg L^{-1} (Beaudoin et al., 2005).

Denitrification can be managed on a larger scale by using the denitrifying properties of wetlands and riparian zones to reduce nitrate concentrations in surface water. Such areas can intercept water flow from agricultural runoff and are a favorable environment for denitrification prior to discharge into water bodies (Mc Clain et al., 2003; Ullah et al., 2005; Pinay et al., 2007). However, riparian zones can also provide conditions favorable to dissimilatory NO$_3^-$ reduction to NH$_4^+$ (DNRA), and the balance of nitrate removal between denitrification and DNRA has to be considered (Davis et al., 2008).

The denitrifying capacity of soil can be used for wastewater purification (Lance et al., 1976), in particular as a tertiary treatment method by infiltrating secondary wastewater for nitrogen elimination by coupling nitrification and denitrification (Idelovitch et al., 2003). Enhanced in situ denitrification is presented as an economically attractive process to decrease nitrate concentration in drinking water in rural areas by injecting a carbon substrate such as ethanol or acetate into the contaminated aquifer (Khan and Spalding, 2004). However, such processes have to be managed with great care to avoid clogging the aquifer pores (Mariotti, 1994).

Denitrification management may be a way of mitigating the emissions of N$_2$O. Limiting nitrogen fertilization to the levels required for the plants and avoiding an excess of reactive nitrogen in soils appears to be the first mitigation method to be implemented in order to limit denitrification intensity and N$_2$O emissions. Moreover, some soils appear particularly prone to N$_2$O emissions with a high N$_2$O/N$_2$ molecular ratio from denitrification (Hénault et al., 2001): Specific management of these areas, such as low or no fertilization and pH adjustment, could be a way of reducing the incidence of these emission hot spots. Another way could be a better use of legumes in plant rotation. Rochette and Janzen (2005) have reported low N$_2$O soil emissions when legumes are cultivated. Many N-fixing bacteria are denitrifiers and can reduce N$_2$O into N$_2$. Better management of symbiotic nitrogen fixation by legumes could be a way of managing denitrification and reducing N$_2$O emissions and therefore needs further study.

27.4.7 Conclusions

The global nitrogen cycle is subject to strong anthropogenic disturbance owing to increasing inputs of reactive nitrogen from fertilizer application, atmospheric deposition, enhanced biological nitrogen fixation (BNF) by cultivated legumes and intensive agricultural practices, which stimulate nitrogen transformation rates. Denitrification should be studied within the general framework of the N cycle, taking these disturbances into account. The close correlation between atmospheric CO$_2$ levels, temperature variations, and atmospheric N$_2$O concentrations over the past 60,000 years (Gruber and Galloway, 2008) raises questions about the contribution of available nitrogen in the global carbon cycle and highlights the need for complementary studies on the connection between biological N fixation, denitrification, and N immobilization in soil and biomass.

The N balance between nitrogen fixation and denitrification, which are both processes that control the reactive nitrogen content in soils and water, is a major issue, particularly for the terrestrial ecosystem. Even though the N balance, taking into account N inputs and losses and storage of reactive N, is considered to be globally neutral, there are considerable uncertainties in the various components of this balance (Gruber and Galloway, 2008); while data on N inputs in terrestrial ecosystems is well-defined, estimates of terrestrial N losses by denitrification are still too inaccurate because of the large variability of denitrification in soils. Methods to understand and manage this important step in the N cycle must be developed to improve its estimation in

the environment. There are two main objectives for the further development of denitrification models: Better integration of spatial and temporal variability in these models and the development of methods for spatial integration to evaluate denitrification on a large scale such as basins or regions. Moreover, while major research has been carried out on the regulation of denitrification intensity in soils, more recent data underline the importance of denitrification N losses in aquatic environments, in particular in riparian areas and rivers. This demonstrates a real need for a better evaluation of the components contributing to denitrification in such ecosystems.

Concerning the microbial denitrifying community, the major issue is the role of denitrifier diversity in ecosystem functioning, especially for N flows and N_2O emissions. Denitrifying bacteria in various environments seem to be extremely diverse but basic questions, such as the redundancy of populations and the consequences of population shifts within the denitrifying community and their effect on denitrifying activity, remain unanswered. In particular, our understanding of the role of the active population and its response to environmental factors is still limited by current methods and constitutes a real challenge for the next 10 years.

References

Addiscott, T.M., and D.S. Powlson. 1992. Partitioning losses of nitrogen fertilizer between leaching and denitrification. J. Agric. Sci. 118:101–107.

Allison, F.E. 1955. The enigma of soil nitrogen balance sheets. Adv. Agron. 7:213–250.

Alvarez, R., H.S. Steinbach, S.M. Grigera, E. Cartier, G. Obregon, S. Torri, and R. Garcia. 2004. The balance sheet method as a conceptual framework for nitrogen fertilization of wheat in a pampean agroecosystem. Agron. J. 96:1050–1057.

Arah, J.R.M. 1990. Diffusion-reaction models of denitrification in soil microsites, p. 245–258. *In* N.P. Revsbech and J. Sorensen (eds.) Denitrification in soil and sediment. Plenum Press, New York.

Aulack, M.S., J.W. Doran, and A.R. Mosier. 1991. Field evaluation of 4 methods for measuring denitrification. Soil Sci. Soc. Am. J. 55:1332–1338.

Aulakh, M.S., J.W. Doran, and A.R. Mosier. 1992. Soil denitrification—Significance, measurement and effect of management. Adv. Soil Sci. 18:1–57.

Avrahami, S., R. Conrad, and G. Braker. 2002. Effect of soil ammonium concentration on N_2O release and on the community structure of ammonia oxidisers and denitrifiers. Appl. Environ. Microbiol. 68:5685–5692.

Balderston, W.L., B. Sherr, and W.J. Payne. 1976. Blockage by acetylene of nitrous oxide reduction in *Pseudomonas perfectomarinus*. Appl. Environ. Microbiol. 31:504–508.

Barton, L., C.D.A. McLay, L.A. Schipper, and C.T. Smith. 1999. Annual denitrification rates in agricultural and forest soils: A review. Aust. J. Soil Res. 37:1073–1093.

Bateman, E.J., and E.M. Baggs. 2005. Contributions of nitrification and denitrification to N_2O emissions from soils at different water-filled pore space. Biol. Fertil. Soils 41:379–388.

Baumann, B., M. Snozzi, A.J.B. Zehnder, and J.R. Van Der Meer. 1996. Dynamics of denitrification activity of *Paracoccus denitrificans* in continuous culture during aerobic–anaerobic changes. J. Bacteriol. 178:4367–4374.

Beaudoin, N., J.K. Saad, C. Van Laethem, J.M. Machet, J. Maucorps, and B. Mary. 2005. Nitrate leaching in intensive agriculture in Northern France: Effect of farming practices, soils and crop rotations. Agric. Ecosyst. Environ. 111:292–310.

Billen, G., and J. Garnier. 1999. Nitrogen transfers through the Seine drainage network: A budget bases on the application of the Riverstrahler model. Hydrobiologia 410:111–122.

Bock, E., I. Schmidt, R. Stüven, and D. Zart. 1995. Nitrogen loss caused by denitrifying *Nitrosomonas* cells using ammonium or hydrogen as electron donors and nitrite as electron acceptor. Arch. Microbiol. 163:16–20.

Boeckx, P., and O. Van Cleemput. 2001. Estimates of N_2O and CH_4 fluxes from agricultural lands in various regions in Europe. Nutr. Cycl. Agroecosyst. 60:35–47.

Bouwman, A.F., 1996. Direct emission of nitrous oxide from agricultural soils. Nutr. Cycl. Agroecosyst. 46:53–70.

Bouwman, A.F., L.J.M. Boumans, and N.H. Batjes. 2002a. Emissions of N_2O and NO from fertilized fields. Summary of available data. Global Biogeochem. Cycles 16:1058.

Bouwman, A.F., L.J.M. Boumans, and N.H. Batjes. 2002b. Modeling global annual N_2O and NO emissions from fertilized fields. Global Biogeochem. Cycles 16:1080.

Bremer, C., G. Braker, D. Matthies, A. Reuter, C. Engels, and R. Conrad. 2007. Impact of plant functional group, plant species and sampling time on the composition of *nirK*-type denitrifier communities in soil. Appl. Environ. Microbiol. 21:6876–6884.

Bremner, J.M., and K. Shaw. 1958. Denitrification in soil. I. Methods of investigation. J. Agric. Sci. 51:22–39.

Broadbent, F.E., and A.B. Carlson. 1978. Field trials with isotopically labelled nitrogen fertilizer, p. 1–41. *In* D.R. Nielsen and J.G. MacDonald (eds.) Nitrogen in the environment. Academic Press, New York.

Broadbent, F.E., and F.E. Clark. 1965. Denitrification, p. 344–359. *In* W.V. Bartholomew and F.E. Clark (eds.) Soil nitrogen. ASA, Madison, WI.

Brooks, M.H., R.L. Smith, and D.L. Macalady. 1992. Inhibition of existing denitrification enzyme activity by chloramphenicol. Appl. Environ. Microbiol. 58:1746–1753.

Burford, J.R., and J.M. Bremner. 1975. Relationships between the denitrification capacities of soils and total, water-soluble and readily decomposable soil organic matter. Soil Biol. Biochem. 7:389–394.

Burns, I.G. 2006. Assessing N fertiliser requirements and the reliability of different recommendation systems, p. 35–48. *In* F. Tei, P. Benincasa, and M. Guiducci (eds.) Towards ecologically sound fertilisation strategies for field vegetable production. Proc. Intern. Symp, Book Series Acta Horticulturae, Leuven, Belgium.

Carter, J.P., Y.H. Hsaio, S. Spiro, and D.J. Richardson. 1995. Soil and sediment bacteria capable of aerobic nitrate respiration. Appl. Environ. Microbiol. 61:2852–2858.

Cavigelli, M.A., and G.P. Robertson. 2001. Role of denitrifier diversity in rates of nitrous oxide consumption in a terrestrial ecosystem. Soil Biol. Biochem. 33:297–310.

Chen, D., Y. Li, P. Grace, and A.R. Mosier. 2008. N_2O emissions from agricultural lands: A synthesis of simulation approaches. Plant Soil 309:169–189.

Chèneby, D., A. Hartmann, C. Hénault, E. Topp, and J.C. Germon. 1998. Diversity of denitrifying microflora and ability to reduce N_2O in two soils. Biol. Fertil. Soil 28:19–26.

Chèneby, D., S. Perrez, C. Devroe, S. Hallet, Y. Couton, F. Bizouard, G. Iuretig, J.C. Germon, and L. Philippot. 2004. Denitrifying bacteria in bulk and maize-rhizospheric soil: Diversity and N_2O-reducing abilities. Can. J. Microbiol. 50:469–474.

Chèneby, D.L., H.A. Philippot, C. Hénault, and J.C. Germon. 2000. 16S rDNA analysis for characterization of denitrifying bacteria isolated from three agricultural soils. FEMS Microbiol. Ecol. 34:121–128.

Cicerone, R.J. 1989. Analysis of sources and sinks of atmospheric nitrous oxide (N_2O). J. Geophys. Res. 94:18265–18271.

Clays-Josserand, A., P. Lemanceau, L. Philippot, and R. Lensi. 1995. Influence of two plant species (flax and tomato) on the distribution of nitrogen dissimilative abilities within fluorescent *Pseudomonas* spp. Appl. Environ. Microbiol. 61:1745–1749.

Colbourn, P., M.M. Iqbal, and I.W. Harper. 1984. Estimation of the total gaseous nitrogen losses from clay soils under laboratory and field conditions. J. Soil Sci. 35:11–22.

Cole, J. 1996. Nitrate reduction to ammonia by enteric bacteria: Redundancy, or a strategy for survival during oxygen starvation? FEMS Microbiol. Lett. 136:1–11.

Crutzen, P.J., A.R. Mosier, K.A. Smith, and W. Winiwarter. 2007. N_2O release from agro-biofuel production negates global warning reduction by replacing fossil fuels. Atmos. Chem. Phys. Discuss. 7:11191–11205.

Dambreville, C., S. Hallet, C. Nguyen, T. Morvan, J.C. Germon, and L. Philippot. 2006. Structure and activity of the denitrifying community in a maize-cropped field fertilized with composted pig manure or ammonium nitrate. FEMS Microbiol. Ecol. 56:119–131.

Dandie, C.E., D.L. Burton, B.J. Zebarth, S.L. Henderson, J.T. Trevors, and C. Goyer. 2008. Changes in bacterial denitrifier community abundance over time in an agricultural field and their relationship with denitrification activity. Appl. Environ. Microbiol. 74:5997–6005.

Dandie, C., D. Burton, B. Zebarth, J. Trevors, and C. Goyer. 2007. Analysis of denitrification genes and comparison of *nosZ*, *cnorB* and 16S rDNA from culturable denitrifying bacteria in potato cropping systems. Syst. Appl. Microbiol. 30:128–138.

Daniel, R.M., A.W. Limmer, K.W. Steele, and I.M. Smith. 1982. Anaerobic growth, nitrate reduction and denitrification in 46 *Rhizobium* strains. J. Gen. Microbiol. 128:1811–1815.

David, M.B., G.F. McIsaac, T.V. Royer, R.G. Darmody, and L.E. Genty. 2001. Estimated historical and current nitrogen balances for Illinois. Sci. World 1:597–604.

Davidson, E.A. 1991. Fluxes of nitrous oxide and nitrite oxide from terrestrial ecosystems, p. 219–235. *In* J.E. Rogers and W.B. Whitman (eds.) Microbial production and consumption of greenhouse gases: Methane, nitrogen oxides, and halomethanes. American Society for Microbiology, Washington, DC.

Davidson, E.A., M. Keller, H.E. Erickson, L.E. Verchot, and E. Veldkamp. 2000. Testing a conceptual model of soil emissions of nitrous and nitric oxides. BioScience 50:667–680.

Davidson, E.A., and W. Kingerlee. 1997. A global inventory of nitric oxide emissions from soils. Nutr. Cycl. Agroecosyst. 48:37–50.

Davis, J.H., W.R. Horwath, J.J. Steiner, and D.D. Myrold. 2008. Denitrification and nitrate consumption in an herbaceous riparian area and perennial rye grass seed cropping system. Soil Sci. Soc. Am. J. 72:1299–1310.

Dehérain, P.P. 1897. La réduction des nitrates dans la terre arable. C.R. Acad. Sci. Paris 124:269–273.

Dehérain, P.P., and L. Maquenne. 1882. Sur la réduction des nitrates dans la terre arable. C.R. Acad. Sci. Paris 95:691–693.

DeKlein, C.A.M., and R.S.P. van Logtestijn. 1994. Denitrification in the top soil of managed grasslands in the Netherlands in relation to the soil type and fertilizer level. Plant Soil 162:33–44.

Dowdell, R.J., and C.P. Webster. 1984. A lysimeter study of the fate of fertilizer nitrogen in spring barley crops grown on a shallow soils overlying chalk: Denitrification losses and the nitrogen balance. J. Soil Sci. 35:183–190.

Enwall, K., L. Philippot, and S. Hallin. 2005. Activity and composition of the denitrifying bacterial community respond differently to long-term fertilization. Appl. Environ. Microbiol. 71:8335–8343.

Enwall, K., I. Throbäck, M. Stenberg, M. Söderström, and S. Hallin. 2010. Soil resources influence spatial patterns of denitrifying communities at scales compatible with land management. Appl. Environ. Microbiol. 76:2243–2250.

Federova, R.I., E.I. Milekhina, and N.I. Il'Yukhina. 1973. Evaluation of the method of gas metabolism for detecting extraterrestrial life. Identification of nitrogen fixing microorganisms. Izv. Akad. Nauk. SSSR. Ser. Biol. 6:797–806.

Firestone, M.K., and E.A. Davidson. 1989. Microbiological basis of NO and N_2O production and consumption in soil, p. 7–21. *In* M.O. Andreae and D.S. Schimel (eds.) Exchange of trace gases between terrestrial ecosystems and their atmosphere. John Wiley & Sons, New York.

Firestone, M.K., R.B. Firestone, and J.M. Tiedje. 1980. Nitrous oxide from soil denitrification: Factors controlling its biological production. Science 208:749–751.

Folorunso, O.A., and D. Rolston. 1984. Spatial variability of field-measured denitrification gas fluxes. Soil Sci. Soc. Am. J. 48:1214–1219.

Gabrielle, B., P. Laville, O. Duval, B. Nicoullaud, J.C. Germon, and C. Henault. 2006a. Process-based modeling of nitrous oxide emissions from wheat-cropped soils at the subregional scale. Global Biogeochem. Cycles 20:GB4018.

Gabrielle, B., P. Laville, C. Henault, B. Nicoullaud, and J.C. Germon. 2006b. Simulation of nitrous oxide emissions from wheat-cropped soils using CERES. Nutr. Cycl. Agroecosyst. 74:133–146.

Galloway, J.N., F.J. Dentener, D.G. Capone, E.W. Boyer, R.W. Howarth, S.P. Seitzinger, G.P. Asner et al. 2004. Nitrogen cycles: Past, present, and future. Biogeochemistry 70:153–226.

Galloway, J.N., A.R. Townsend, J.W. Erisman, M. Bekunda, Z. Cai, J.R. Freney, L.A. Martinelli, S.P. Seitzinger, and M.A. Sutton. 2008. Transformation of the nitrogen cycle: Recent trends, questions, and potential solutions. Science 320:889–892.

Gamble, T.N., H.R. Betlach, and J.M. Tiedje. 1977. Numerically dominant denitrifying bacteria from world soils. Appl. Environ. Microbiol. 33:926–939.

Garcia, J. 1977. Analyse de différents groupes composant la microflore dénitrifiante de sols de rizière du Sénégal. Ann. Microbiol. Inst. Pasteur 128:433–446.

Garrido, F., C. Hénault, H. Gaillard, S. Pérez, and J.C. Germon. 2002. N_2O and NO emissions by agricultural soils with low hydraulic potentials. Soil Biol. Biochem. 34:559–575.

Gayon, U., and G. Dupetit. 1882. Sur la fermentation des nitrates. C.R. Acad. Sci. 95:644–646.

Germon, J.C. 1980. Etude quantitative de la dénitrification biologique dans le sol à l'aide de l'acétylène. II. Evolution de l'effet inhibiteur de l'acétylène sur la N_2O réductase. Ann. Microbiol. Inst. Pasteur 131B:81–90.

Glockner, A.B., A. Jungst, and W.G. Zumft. 1993. Copper-containing nitrite reductase from *Pseudomonas aureofaciens* is functional in a mutationally cytochrome cd_1 free background (*NirS⁻*) of *Pseudomonas stutzeri*. Arch. Microbiol. 160:18–26.

Groffman, P.M., M.A. Altabet, J.K. Böhlke, K. Butterbach-Bahl, M.B. David, M.K. Firestone, A.E. Giblin, T.M. Kana, L.P. Nielsen, and M.A. Voytek. 2006. Methods for measuring denitrification: Diverse approaches to a difficult problem. Ecol. Appl. 16:2091–2122.

Groffman, P.M., K. Butterbach-Bahl, R.W. Fulweiler, A.J. Gold, J.L. Morse, E.K. Stander, C. Tague, C. Tonitto, and P. Vidon. 2009. Challenges to incorporating spatially and temporally explicit phenomena (hotspots and hot moments) in denitrification models. Biogeochemistry 93:49–77.

Groffman, P.M., J.M. Tiedje, G.P. Robertson, and S. Christensen. 1987. Denitrification at different temporal and geographical scales: Proximal and distal controls, p. 174–192. *In* J.R. Wilson (ed.) Advances in nitrogen cycling in agricultural ecosystems. CAB International, Oxon, U.K.

Gruber, N., and J.N. Galloway. 2008. An earth-system perspective of the global nitrogen cycle. Nature 451:293–296.

Grundman, G.L., and D.D. Rolston. 1987. A water function approximation to degree of anaerobiosis associated with denitrification. Soil Sci. 144:437–441.

Guiraud, G., and Y. Berliet. 1970. Détermination quantitative et isotopique par spectrométrie de masse de composés gazeux produits dans la denitrification. Chim. Anal. 52:53–56.

Hallin, S., G. Braker, and L. Philippot. 2007. Molecular tools to assess diversity and density of denitrifiers in their habitats, p. 313–330. *In* H. Bothe, S.J. Ferguson, and W.E. Newton (eds.) Biology of the nitrogen cycle. Elsevier, Amsterdam, the Netherlands.

Hallin, S., C. Jones, M. Schloter, and L. Philippot. 2009. Relationship between N-cycling communities and ecosystem functioning in a 50-year-old fertilization experiment. ISME J. 3:597–605.

Hashimoto, T., K. Furue, and T. Kubota. 2005. Accumulation of denitrifying bacteria and enhancement of denitrification activity in deep subsurface soil in relation to the geological profile. Soil Sci. Plant Nutr. 51:899–903.

Hauck, R.D., S.W. Melsted, and P.E. Yankwich. 1958. Use of N-isotopes distribution in nitrogen gas in the study of denitrification. Soil Sci. 86:287–291.

Heinemeyer, O. 1979. Effect of commercial pesticides on microbial denitrification in solution and in soils. Mitt. Deutsch. Bodenkund. Gesell. 29:331–337.

Heinen, M. 2006a. Simplified denitrification models: Overview and properties. Geoderma 133:444–463.

Heinen, M. 2006b. Application of a widely used denitrification model to Dutch data sets. Geoderma 133:464–473.

Hénault, C., F. Bizouard, P. Laville, B. Gabrielle, B. Nicoullaud, J.C. Germon, and P. Cellier. 2005. Predicting in situ soil N_2O emissions using NOE algorithm and soil database. Global Change Biol. 11:115–127.

Hénault, C., D. Chèneby, K. Heurlier, F. Garrido, S. Perez, and J.C. Germon. 2001. Laboratory kinetics of soil denitrification are useful to discriminate soils with potentially high levels of N_2O emission on the field scale. Agronomy 21:713–723.

Hénault, C., and J.C. Germon. 2000. NEMIS, a predictive model of denitrification on the field scale. Eur. J. Soil Sci. 51:257–270.

Henry, S., E. Baudoin, J.C. Lopez-Gutierrez, F. Martin-Laurent, A. Brauman, and L. Philippot. 2004. Quantification of denitrifying bacteria in soils by *nirK* gene targeted real-time PCR. J. Microbiol. Methods 59:327–335.

Henry, S., D. Bru, B. Stres, S. Hallet, and L. Philippot. 2006. Quantitative detection of the *nosZ* gene, encoding nitrous oxide reductase, and comparison of the abundance of 16S rRNA, *narG*, *nirK*, and *nosZ* genes in soils. Appl. Environ. Microbiol. 72:5181–5189.

Henry, S., S. Texier, S. Hallet, D. Bru, C. Dambreville, D. Chèneby, F. Bizouard, J.C. Germon, and L. Philippot. 2008. Disentangling the rhizosphere effect on nitrate reducers and denitrifiers: Insight into the role of root exudates. Environ. Microbiol. 10:3082–3092.

Heylen, K., B. Vanparys, L. Vittebolle, W. Verstraete, N. Boon, and P. De Vos. 2006. Cultivation of denitrifying bacteria: Optimization of isolation conditions and diversity study. Appl. Environ. Microbiol. 72:2637–2643.

Hofstra, N., and A.F. Bouwman. 2005. Denitrification in agricultural soils: Summarizing published data and estimating global annual rates. Nutr. Cycl. Agroecosyst. 72:267–278.

Hynes, R., and R. Knowles. 1978. Effect of acetylene on autotrophic and heterotrophic nitrification. Can. J. Microbiol. 28:334–340.

Idelovitch, E., N. Icekson-Tal, O. Avraham, and M. Michall. 2003. The long-term performance of soil aquifer treatment (SAT) for effluent reuse. Water Sci. Technol. 4:239–246.

IPCC. 2001. Climate change 2001. Scientific basis. Cambridge University Press, Cambridge, U.K.

IPCC. 2006. Guidelines for national greenhouse gas inventories. IGES, Hayama, Japan.

IPCC. 2007. Climate change 2007. The physical science basis. Cambridge University Press, Cambridge, U.K.

Jambert, C., D. Serça, and R. Delmas. 1997. Quantification of N losses as NH_3, NO and N_2O and N_2 from fertilized maize fields in southwestern France. Nutr. Cycl. Agroecosyst. 48:91–104.

Jarvis, S.C., D. Barraclough, J. Williams, and A.J. Rook. 1991. Patterns of denitrification loss from grazed grassland: Effects of N fertilizer inputs at different sites. Plant Soil 131:77–88.

Johnsson, H., L. Klemedtsson, A. Nilsson, and B.H. Svensson. 1991. Simulation of field scale denitrification losses from soils under grass ley and barley. Plant Soil 138:287–302.

Ju, X.T., G.X. Xing, X.P. Chen, S.L. Zhang, L.J. Zhang, X.J. Liu, Z.L. Cui et al. 2009. Reducing environmental risk by improving N management in intensive Chinese agricultural systems. Proc. Natl. Acad. Sci. USA 106:3041–3046.

Khan, I.A., and R.F. Spalding. 2004. Enhanced in situ denitrification for a municipal well. Water Res. 38:3382–3388.

Klemedtsson, L., B.H. Svensson, T. Lindberg, and T. Rosswall. 1977. The use of acetylene inhibition of nitrous-oxide reductase in quantifying denitrification in soils. Swed. J. Agric. Res. 7:179–185.

Kroeze, C., R. Aerts, N. van Breemen, D. van Dam, K. van der Hoek, P. Hofschreuder, M. Hoosbeek et al. 2003. Uncertainties in the fate of nitrogen. I: An overview of sources of uncertainty illustrated with a Dutch case study. Nutr. Cycl. Agroecosyst. 66:43–69.

Lance, J.C., F.D. Whisler, and R.C. Rice. 1976. Maximizing denitrification during soil filtration of sewage water. J. Environ. Qual. 5:102–107.

Laville, P., C. Hénault, P. Renault, P. Cellier, X. Devis, A. Oriol, and J.C. Germon. 1997. Field comparison of nitrous oxide emission measurement using micrometeorological and chambers methods. Agronomy 17:375–388.

Lehmann, M.F., P. Reichert, S.M. Bernasconi, A. Barbieri, and A. McKenzie. 2003. Modelling nitrogen and oxygen isotope fractionation during denitrification in a lacustrine redox-transition zone. Geochim. Cosmochim. Acta 67:2529–2542.

Li, C.S. 2000. Modeling trace gas emissions from agricultural ecosystems. Nutr. Cycl. Agroecosyst. 58:259–276.

Li, C.S., S. Frolking, and T.A. Frolking. 1992. A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity. J. Geophys. Res. 97:9759–9776.

Linn, D.M., and J.W. Doran. 1984. Effect of water-filled pore space on carbon dioxide and nitrous oxide production in tilled and non tilled soils. Soil Sci. Soc. Am. J. 48:1267–1272.

Magiotto, S.R., J.A. Webb, C. Wagner-Riddle, and G.W. Thurtell. 2000. Nitrous and nitrogen oxide emissions from turfgrass receiving different forms of, nitrogen fertilizer. J. Environ. Qual. 29:621–630.

Mahne, I., and J.M. Tiedje. 1995. Criteria and methodology for identifying respiratory denitrifiers. Appl. Environ. Microbiol. 61:1110–1115.

Mariotti, A. 1994. Dénitrification in situ dans les eaux souterraines: Processus naturels ou provoqués: Une revue. Hydrogéologie 3:43–68.

Mariotti, A., J.C. Germon, and A. Leclerc. 1982. Nitrogen isotope fractionation associated with NO_2^-/N_2O step of denitrification in soils. Can. J. Soil Sci. 62:227–241.

Mathieu, O., C. Hénault, J. Lévêque, E. Baujard, M.J. Milloux, and F. Andreux. 2006. Quantifying the contribution of nitrification and denitrification to the nitrous oxide flux using ^{15}N tracers. Environ. Pollut. 144:933–940.

Mc Clain, M.E., E.W. Boyer, C.L. Dent, S.E. Gergel, N.B. Grimm, P.M. Groffman, S.C. Hart et al. 2003. Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. Ecosystems 6:301–312.

Meynard, J.M., M. Cerf, L. Guichard, M.H. Jeuffroy, and D. Makowski. 2002. Which decision support tools for the environmental management of nitrogen. Agronomy 22:817–829.

Miller, M.N., B.J. Zebarth, C.E. Dandie, D.L. Burton, C. Goyer, and J.T. Trevors. 2008. Crop residue influence on denitrification, N_2O emissions and denitrifier community abundance in soil. Soil Biol. Biochem. 40:2553–2562.

Mogge, B., E.A. Kaiser, and J.C. Munch. 1999. Nitrous oxide emissions and denitrification N-losses from agricultural soils in the Bornhöved lake region: Influence of organic fertilizers and land-use. Soil Biol. Biochem. 31:1245–1252.

Mosier, A.R. 1980. Acetylene inhibition of ammonium oxidation in soil. Soil Biol. Biochem. 12:443–444.

Mosier, A.R., W.D. Guenzi, and E.E. Schweizer. 1986. Field denitrification estimation by nitrogen-15 and acetylene techniques. Soil Sci. Soc. Am. J. 50:831–833.

Mosier, A., C. Kroeze, C. Nevison, O. Oenema, S. Seitzinger, and O. van Cleemput. 1998. Closing the global N_2O budget: Nitrous oxide emissions through the agricultural nitrogen cycle. OECD/IPCC/IEA phase II development of IPCC guidelines for national greenhouse gas inventory methodology. Nutr. Cycl. Agroecosyst. 52:225–248.

Mosier, A.R., and W.J. Parton. 1985. Denitrification in a shortgrass prairie: A modelling approach, p. 441–460. *In* D.E. Caldwell, J.A. Brierley, and C.L. Brierley (eds.) Planetary ecology. Van Nostrand Reinhold, Princeton, NJ.

Mounier, E., S. Hallet, D. Chèneby, E. Benizri, Y. Gruet, C. Nguyen, S. Piutti et al. 2004. Influence of maize mucilage on the diversity and activity of the denitrifying community. Environ. Microbiol. 6:301–312.

Murray, R.E., and R. Knowles. 1999. Chloramphenicol inhibition of denitrifying enzyme activity in two agricultural soils. Appl. Environ. Microbiol. 65:3487–3492.

Myrold, D.D. 1988. Denitrification in ryegrass and winter wheat cropping systems of western Oregon. Soil Sci. Soc. Am. J. 52:412–416.

Myrold, D.D. 1991. Measuring denitrification in soils using ^{15}N techniques, p. 181–198. *In* N.P. Revsbech and J. Sorensen (eds.) Denitrification in soil and sediment. Plenum Press, New York.

Myrold, D.D., and J.M. Tiedje. 1985. Diffusional constraints on denitrification in soil. Soil Sci. Soc. Am. J. 49:651–657.

Nömmik, H. 1956. Investigations on denitrification in soil. Acta Agric. Scand. 6:195–228.

Parkin, T.B. 1987. Soil microsites as a source of denitrification variability. Soil Sci. Soc. Am. J. 51:1194–1199.

Parkin, T.B., H.F. Kaspar, A.J. Sexstone, and J.M. Tiedje. 1984. A gas-flow soil core method to measure field denitrification rates. Soil Biol. Biochem. 16:323–330.

Parkin, T.B., A.J. Sexstone, and J.M. Tiedje. 1985a. Comparison of field denitrification rates determined by acetylene-based soil core and nitrogen-15 methods. Soil Sci. Soc. Am. J. 49:94–99.

Parkin, T.B., A.J. Sexstone, and J.M. Tiedje. 1985b. Adaptation of denitrifying populations to low soil pH. Appl. Environ. Microbiol. 49:1053–1056.

Parkin, T.B., J.L. Starr, and J.J. Meisinger. 1987. Influence of sample size on measurement of soil denitrification. Soil Sci. Soc. Am. J. 51:1492–1501.

Parkin, T.B., and J.M. Tiedje. 1984. Application of a soil core method to investigate the effect of oxygen concentration on denitrification. Soil Biol. Biochem. 16:331–334.

Parry, S., P. Renault, C. Chenu, and R. Lensi. 1999. Denitrification in pasture and cropped soil clods as affected by pore space structure. Soil Biol. Biochem. 31:493–501.

Parton, W.J., M. Hartman, D. Ojima, and D.W. Schimel. 1998. DAYCENT and its land surface submodel: Description and testing. Global Planet. Change 19:35–48.

Parton, W.J., A.R. Mosier, D.S. Ojima, D.W. Valentine, D.W. Schimel, K. Weier, and A.E. Kulmala. 1996. Generalized model for N_2 and N_2O production from nitrification and denitrification. Global Biogeochem. Cycles 10:401–412.

Pell, M., B. Stenberg, J. Strenström, and L. Torstensson. 1996. Potential denitrification activity assay in soil—With and without chloramphenicol? Soil Biol. Biochem. 28:393–398.

Philippot, L. 2002. Denitrifying genes in bacterial and archeal genomes. Biochim. Biophys. Acta 1577:355–376.

Philippot, L. 2005. Denitrification in pathogenic bacteria: For better or worst? Trends Microbiol. 13:191–192.

Philippot, L., J. Čuhel, N. Saby, D. Chèneby, A. Chroňáková, D. Bru, D. Arrouays, F. Martin-Laurent, and M. Šimek. 2009. Mapping field-scale spatial distribution patterns of size and activity of the denitrifier community. Environ. Microbiol. 11:1518–1526.

Philippot, L., and S. Hallin. 2005. Finding the missing link between diversity and activity using denitrifying bacteria as a model functional community. Curr. Opin. Microbiol. 8:234–239.

Philippot, L., S. Hallin, and M. Schloter. 2007. Ecology of denitrifiers in agricultural soil. Adv. Agron. 96:135–190.

Pinay, G., B. Gumiero, E. Tabacchi, O. Gimenez, A.M. Tabacchi-Planty, M.M. Hefting, T.P. Burt et al. 2007. Pattern of denitrification rates in European alluvial soils under various hydrological regimes. Freshwater Biol. 52:252–266.

Potter, C.S., R.H. Riley, and S.A. Klooster. 1997. Simulation modeling of nitrogen gas emissions along an age gradient of tropical forest soils. Ecol. Model. 97:176–196.

Prather, M.R., D. Derwent, P. Erhalt, P. Fraser, E. Sanhueza, and X. Zhou. 1995. Other trace gases and atmospheric chemistry, p. 73–126. *In* J.T. Houghton et al. (eds.) Climate change 1994: Radiative forcing of climate change and an evaluation of the IPCC IS92 scenarios. Cambridge University Press, Cambridge, U.K.

Priemé, A., G. Braker, and J. Tiedje. 2002. Diversity of nitrite reductase (*nirK* and *nirS*) gene fragments in forested upland and wetland soils. Appl. Environ. Microbiol. 68:1893–1900.

Recous, S., C. Fresneau, G. Faurie, and B. Mary. 1988. The fate of ^{15}N-labelled urea and ammonium nitrate applied to a winter wheat crop. II. Plant uptake and N efficiency. Plant Soil 112:215–224.

Reddy, K.R., P.S.C. Rao, and R.E. Jessup. 1982. The effect of carbon mineralisation on kinetics in mineral and organic soils. Soil Sci. Soc. Am. J. 46:62–68.

Rich, J., R. Heichen, P. Bottomley, K. Cromack, Jr., and D. Myrold. 2003. Community composition and functioning of denitrifying bacteria from adjacent meadow and forest soils. Appl. Environ. Microbiol. 69:5974–5982.

Rich, J., and D. Myrold. 2004. Community composition and activities of denitrifying bacteria from adjacent agricultural soil, riparian soil, and creek sediment in Oregon, USA. Soil Biol. Biochem. 36:1431–1441.

Rochette, P., and H.H. Janzen. 2005. Towards a revised coefficient for estimating N_2O emissions from legumes. Nutr. Cycl. Agroecosyst. 73:171–179.

Rolston, D.E., P.S.C. Rao, J.M. Davidson, and R.E. Jessup. 1984. Simulation of denitrification losses of nitrate fertilizer applied to uncropped, cropped, and manure amended field plots. Soil Sci. 137:270–279.

Rolston, D.E., A.N. Sharpley, D.W. Toy, and F.E. Broadbent. 1982. Field measurement of denitrification. III Rates during irrigation cycles. Soil Sci. Soc. Am. J. 44:289–296.

Roussel-Delif, L., S. Tarnawski, J. Hamelin, L. Philippot, M. Aragno, and N. Fromin. 2005. Frequency and diversity of nitrate reductase genes among nitrate-dissimilating *Pseudomonas* in the rhizosphere of perennial grasses grown in field conditions. Microb. Ecol. 49:63–72.

Russow, R., I. Sich, and H.U. Neue. 2000. The formation of the trace gases NO and N_2O in soils by the coupled processes of nitrification and denitrification: Results of kinetic ^{15}N tracer investigations. Chemosphere Global Change Sci. 2:359–366.

Ryden, J.C., L.J. Lund, and D.D. Focht. 1979a. Direct measurement of denitrification loss from soils. I. Laboratory evaluation of acetylene inhibition of nitrous oxide reduction. Soil Sci. Soc. Am. J. 43:104–110.

Ryden, J.C., L.J. Lund, J. Letey, and D.D. Focht. 1979b. Direct measurement of denitrification loss from soils. II. Development and application of field methods. Soil Sci. Soc. Am. J. 43:110–118.

Ryden, J.C., and J.H. Skinner. 1987. Soil core incubation system for the field measurement of denitrification using acetylene inhibition. Soil Biol. Biochem. 19:753–757.

Schauss, K. 2006. Impact of fermented organic fertilizer on in-situ trace gas fluxes and on soil bacterial denitrifying communities in organic agriculture. Ph.D. Thesis. Justus-Liebig-University. Giessen, Germany.

Sebilo, M., G. Billen, M. Grably, and A. Mariotti. 2003. Isotopic composition of nitrate–nitrogen as a marker of riparian and benthic denitrification at the scale of the whole Seine River system. Biogeochemistry 63:35–51.

Sebilo, M., G. Billen, B. Mayer, D. Billiou, M. Grably, J. Garnier, and A. Mariotti. 2006. Assessing nitrification and denitrification in the Seine river and estuary using chemical and isotopic techniques. Ecosystems 9:564–577.

Sexstone, A.J., N.P. Revsbech, T.B. Parkin, and J.M. Tiedje. 1985. Direct measurement of oxygen profiles and denitrification rates in soil aggregates. Soil Sci. Soc. Am. J. 49:645–651.

Sharma, S., M.K. Aneja, J. Mayer, J.C. Munch, and M. Schloter. 2005. Diversity of transcripts of nitrite reductase genes (*nirK* and *nirS*) in rhizospheres of grain legumes. Appl. Environ. Microbiol. 71:2001–2007.

Shaw, L.J., G.W. Nicol, Z. Smith, J. Fear, J.I. Prosser, and E.M. Baggs. 2006. *Nitrosospira* spp. can produce nitrous oxide via a nitrifier denitrification pathway. Environ. Microbiol. 8:214–222.

Shearer, L.A., J.R. Goldsmith, C. Young, O.A. Kearns, and B.R. Tamplin. 1972. Methemoglobin levels in infants, in an area with high nitrate water supply. Am. J. Public Health 62:1174–1180.

Shearer, G., and D.H. Kohl. 1988. Nitrogen isotopic fractionation and O^{-18} exchange in relation to the mechanism of denitrification by *Pseudomonas stutzeri*. J. Biol. Chem. 263:13231–13245.

Sierra, J., P. Renault, and V. Valles. 1995. Anaerobiosis in saturated soil aggregates: Modelling and experiment. Eur. J. Soil Sci. 46:519–531.

Simek, M., and D.W. Hopkins. 1999. Regulation of potential denitrification by soil pH in long-term fertilized arable soils. Biol. Fertil. Soils 30:41–47.

Simek, M., L. Jisova, and D.W. Hopkins. 2002. What is the so-called optimum pH for denitrification in soil? Soil Biol. Biochem. 34:1227–1234.

Smith, K.A. 1980. A model of the extent of anaerobic zones in aggregated soils and its potential application to estimates denitrification. J. Soil Sci. 31:263–277.

Smith, M.S., M.K. Firestone, and J.M. Tiedje. 1978. The acetylene inhibition method for short-term measurement of soil denitrification and its evaluation using nitrogen-13. Soil Sci. Soc. Am. J. 42:611–615.

Smith, M.S., and J.M. Tiedje. 1979. Phases of denitrification following oxygen depletion in soil. Soil Biol. Biochem. 11:261–267.

Standford, G., R.A. Vander Pol, and S. Dziena. 1975. Denitrification rates in relation to total and extractable soil carbon. Soil Sci. Soc. Am. Proc. 39:284–289.

Standing. D., E.M. Baggs, M. Wattenbach, P. Smith, and K. Killham. 2007. Meeting the challenge of scaling up processes in the plant–soil–microbe system. Biol. Fertil. Soils 44:245–257.

Stehfest, E., and A.F. Bouwman. 2006. N$_2$O and NO emissions from agricultural fields and soils under natural vegetation: Summarizing available measurement data and modeling of global annual emissions. Nutr. Cycl. Agroecosyst. 74:207–228.

Stevens, R.J., and R.J. Laughlin. 2001. Lowering the detection limit for dinitrogen using the enrichment of nitrous oxide. Soil Biol. Biochem. 33:1287–1289.

Stevens, R.J., R.J. Laughlin, L.C. Burns, J.R.M. Arah, and R.C. Hood. 1997. Measuring the contributions of nitrification and denitrification to the flux of nitrous oxide from soil. Soil Biol. Biochem. 29:139–151.

Taroncher-Oldenburg, G., E.M. Griner, C.A. Francis, and B.B. Ward. 2003. Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. Appl. Environ. Microbiol. 69:1159–1171.

Terry, R.E., and J.M. Duxbury. 1985. Acetylene decomposition in soils. Soil Sci. Soc. Am. J. 49:90–94.

Thomsen, J.K., T. Geest, and R.P. Cox. 1994. Mass spectrometric studies of the effect of pH on the accumulation of intermediates in denitrification by *Paracoccus denitrificans*. Appl. Environ. Microbiol. 60:536–541.

Throbäck, I.N., K. Enwall, A. Jarvis, and S. Hallin. 2004. Reassessing PCR primers targeting *nirS*, *nirK* and *nosZ* genes for community surveys of denitrifying bacteria with DGGE. FEMS Microbiol. Ecol. 49:401–417.

Tiedje, J.M. 1988. Ecology of denitrification and dissimilatory nitrate reduction to ammonium, p. 179–244. *In* A.J.B. Zehnder (ed.) Biology of anaerobic microorganisms. John Wiley & Sons, New York.

Tiedje, J.M., S. Simkins, and P.M. Grofmann. 1989. Perspective on measurement of denitrification in the field including recommended protocols for acetylene based method. Plant Soil 115:261–284.

Topp, E., and J.C. Germon. 1986. Acetylene metabolism and stimulation of denitrification in an agricultural soil. Appl. Environ. Microbiol. 52:802–806.

Ullah, S., G.A. Breitenbeck, and S.P. Faulkner. 2005. Denitrification and N$_2$O emission from forested and cultivated alluvial clay soil. Biogeochemistry 73:499–513.

Van der Salm, C., J. Dolfing, M. Heinen, and G.L. Velthof. 2007. Estimation of nitrogen losses via denitrification from a heavy clay soil under grass. Agric. Ecosyst. Environ. 119:311–319.

Vinther, F., F. Eiland, A.-M. Lind, and L. Elsgaard. 1999. Microbial biomass and numbers of denitrifiers related to macropore channels in agricultural and forest soils. Soil Biol. Biochem. 31:603–611.

Waibel, A.E., T. Peter, K.S. Carslaw, H. Oelhaf, G. Wetzel, P.J. Crutzen, U. Pöschl, A. Tsias, E. Reimer, and H. Fisher. 1999. Arctic ozone loss due to denitrification. Science 283:2064–2068.

Wang, X.K., and W.Z. Lu. 2006. Seasonal variation of air pollution index: Hong Kong case study. Chemosphere 63:1261–1272.

Weier, K.L., and I.C. Mac Rae. 1992. Denitrifying bacteria in the profile of a brigalow clay soil beneath a permanent pasture and a cultivated crop. Soil Biol. Biochem. 24:919–923.

Weier, K.L., I.C. Mac Rae, and R.J.K. Myers. 1991. Seasonal variation in denitrification in a clay soil under a cultivated crop and a permanent pasture. Soil Biol. Biochem. 23:355–360.

Wijler, J., and C.C. Delwiche. 1954. Investigations on the denitrifying process in soil. Plant Soil 5:155–169.

Wolf, I., and R. Russow. 2000. Different pathways of formation of N_2O, N_2 and NO in black earth soil. Soil Biol. Biochem. 32:229–239.

Wolsing, M., and A. Prieme. 2004. Observation of high seasonal variation in community structure of denitrifying bacteria in arable soil receiving artificial fertilizer and cattle manure by determining T-RFLP of *nir* gene fragments. FEMS Microbiol. Ecol. 48:261–271.

Woodward, K.B., C.S. Fellows, C.L. Conway, and H.H. Hunter. 2009. Nitrate removal, denitrification and nitrous oxide production in the riparian zone of an ephemeral stream. Soil Biol. Biochem. 41:671–680.

Yeomans, J.C., and J.M. Bremner. 1985. Denitrification in soil: Effects of herbicides. Soil Biol. Biochem. 4:447–452.

Yoshinari, T., R. Hynes, and R. Knowles. 1977. Acetylene inhibition of nitrous oxide reduction and measurement of denitrification and nitrogen fixation in soil. Soil Biol. Biochem. 9:177–183.

Yoshinari, T., and R. Knowles. 1976. Acetylene inhibition of nitrous oxide reduction by denitrifying bacteria. Biochem. Biophys. Res. Commun. 69:705–710.

Yu, K., S. Struwe, A. Kjoller, and G. Chen. 2008. Denitrification rate determined by nitrite disappearance is higher than determined by nitrous oxide production with acetylene blockage. Ecol. Eng. 32:90–96.

Zumft, W.G. 1997. Cell biology and molecular basis of denitrification. Microbiol. Mol. Biol. Rev. 61:533–536.

27.5 Nitrogen in the Environment

Peter M. Groffman

27.5.1 Introduction

While N is an extremely valuable and important nutrient within the soils of nearly all ecosystems on the globe (Vitousek and Howarth, 1991), it becomes a highly problematic pollutant once it leaves one ecosystem and moves into another. Given that human alteration of the global N cycle is extreme and dynamic (Galloway et al., 2004, 2008; Schlesinger, 2009), that there are multiple forms of N that can be considered as pollutants, and that there are multiple vectors of transport for these forms of N, it is not surprising that the topic of "N in the environment" is one of the most well studied and controversial in soil and environmental science.

There are multiple forms of N that are of concern in the environment. Galloway et al. (2003) present a useful framework for thinking about how "reactive (all forms of N other than N_2) N" is added to the environment as a by-product of human agricultural, waste disposal, and energy generation activity and then "cascades" through the environment causing air and water quality problems. Soils play a critical role in the N cascade. As described elsewhere in this volume, soils contain a huge reservoir of nonreactive, but potentially mobilizable organic N and several different pools of reactive N. Nitrogen dynamics in soils contrast sharply with the atmosphere, which is dominated by a vast pool of unreactive N_2, and with aquatic ecosystems, which are dominated by pools of dissolved organic and inorganic N and by interactions between sediments and the water column.

In this section, the author provides a brief review of how each of the key species of reactive N are produced, transformed, and transported in the environment and then provides a framework for site or ecosystem-scale evaluations of (1) is this site (ecosystem) a source of N to the environment (surrounding ecosystems) and (2) does this site (ecosystem) have the potential to absorb N, that is, function as a sink. This framework is based on the idea that we can best address N pollution problems in a "landscape-scale context," where a landscape consists of a range of ecosystems (e.g., different cropping systems, different forest types on different soils, wetlands, pastures) that can be viewed as either N sources or sinks and that are connected via hydrologic and atmospheric transport pathways (National Research Council, 1993; Jones et al., 2001; King et al., 2005). A landscape approach allows for reconciliation of the within-system benefits versus external pollution problems that are inherent in analysis of N pollution problems and provides a strong basis for evaluation and management of these problems in the field. My goal is to provide the reader with a series of questions, and an approach to answering these questions, that allow for evaluation of key N source, sink, and transport processes at any point, in any landscape.

27.5.2 Forms of N That Are of Concern in the Environment

Detailed descriptions of the forms and transformations of N can be found elsewhere in this volume and earlier in this chapter. Here, the author provides a brief discussion of why certain forms of N are of concern in the environment, how they are produced and/or consumed within ecosystems, and how they are transported between ecosystems (Table 27.4).

TABLE 27.4 Forms of N of Concern in the Environment

N Form	Sources	Dominant Transport Vectors	Environmental Effects
Nitrate (NO_3^-)	Fertilizer disturbance combustion (acid rain)	Groundwater	Drinking water pollutant Eutrophication
Ammonia (NH_3, NH_4^+)	Fertilizer Animal waste	Surface runoff Atmosphere	Drinking water pollutant Eutrophication
Nitrous oxide (N_2O)	By-product of multiple transformations	Atmosphere Groundwater	Greenhouse gas Ozone destruction
Nitric oxide (NO)	By-product of multiple transformations	Atmosphere	Ozone precursor
Dissolved organic N	By-product of natural decomposition	Surface runoff Groundwater	Possible Eutrophication

27.5.2.1 Nitrate

Nitrate is regulated as a drinking water pollutant (U.S. EPA, 1990) and is considered to be a prime agent of eutrophication in estuarine and coastal ecosystems (Ryther and Dunstan, 1971; Vitousek et al., 1997; Boesch et al., 2001; Rabalais et al., 2002). This ion is added directly to ecosystems in some fertilizer types and is produced via mineralization (conversion of organic N to NH_4^+) and nitrification (conversion of NH_4^+ to NO_3^-) of other fertilizer types and organic materials (e.g., plant detritus, manures, organic wastes). Nitrate is highly soluble and does not have a strong affinity for soil particles and is therefore readily transported in surface runoff and percolation. It is frequently the dominant form of N in percolating water and in aquifers.

Ecosystems tend to be sources of NO_3^- to the environment if they are heavily fertilized and/or highly disturbed. In many ecosystems, NO_3^- losses are prevented by competition between plants and microbes, primarily because nitrifying organisms do not compete well with plants and other microbes for NH_4^+ (Chapman et al., 2006). Therefore, NO_3^- loss occurs only where N inputs from mineralization of soil organic matter, fertilization, atmospheric, or hydrologic sources are high enough to satisfy plant and microbial demands for N or where some type of disturbance disrupts plant and microbial competition, allowing nitrification to occur (Aber et al., 1989). There are many types of disturbance that can cause this disruption. Disturbances to the plant community that induce NO_3^- losses include harvest of crops (Gold et al., 1990; Randall and Mulla, 2001), clear-cutting of forest vegetation (Bormann and Likens, 1979), insect outbreaks (Eshleman et al., 1998; Lovett et al., 2002; Lewis and Likens, 2007), and hurricanes (Schaefer et al., 2000; Bruland et al., 2008). The dominant soil disturbance that can stimulate NO_3^- losses is tillage but more subtle disturbances, for example, earthworm activity (Bohlen et al., 2004) or cultivation can also be important.

In recent years, considerable attention has been paid to atmospheric transport of NO_3^- (Holland et al., 2005). This NO_3^- is formed in the atmosphere from N oxides emitted by combustion processes and returns to earth in rainfall (wet deposition) or by surface deposition as particles and/or vapor (dry deposition). Rates of atmospheric NO_3^- deposition range from 5 to more than 20 kg N ha^{-1} year^{-1} worldwide (National Atmospheric Deposition Program, 2007). Recognition of the importance of NO_3^- deposition has led to concern about N "saturation" of natural (i.e., unfertilized) ecosystems, which has been linked to a variety of deleterious chemical and biological changes in temperate ecosystems (Aber et al., 1989).

Many ecosystems function as "sinks" for NO_3^- in the landscape. Ecosystems where the growth of plants and/or microbes is strongly N limited will take up NO_3^- that is added in fertilizer or that enters via atmospheric or hydrologic transport (Vitousek and Howarth, 1991; Chapman et al., 2006). Ecosystems with high actual or potential rates of denitrification (i.e., with significant volumes of anaerobic soil) function as very effective sinks for NO_3^- (Seitzinger et al., 2006). The location of "sink ecosystems" relative to sources (e.g., wet riparian zones located between intensively farmed uplands and streams) is often a critical determinant of the nature and extent of NO_3^- pollution problems in a landscape (Lowrance et al., 1984; Vidon and Hill, 2006). The capacity of groundwater to support denitrification is another critical (and highly variable) determinant of landscape-scale NO_3^- dynamics (Korom, 1992; Nolan and Stoner, 2000; Rivett et al., 2008). There also appear to be abiotic sinks for N in soil, likely caused by chemical and physical condensation reactions between organic matter, iron, and NO_3^- (Davidson et al., 1991; Fitzhugh et al., 2003; Colman et al., 2008).

The challenge in evaluations of NO_3^- pollution problems is to determine the location, strengths, and hydrologic and atmospheric connections between sources and sinks in the landscape. Procedures for making these determinations are described in Section 27.5.3.

27.5.2.2 Ammonia

Ammonia is likely the most common inorganic form of N in soils. Most N fertilizers are NH_3 (or urea, which rapidly degrades to NH_3) based and NH_3 is the first inorganic N form released in decomposition (mineralization). However, free NH_3 is rarely present in soils of low and moderate pH (i.e., <8.0). Under these conditions, NH_3 exists in a protonated form, NH_4^+, which has a strong affinity for negatively charged clay and organic matter particles and is therefore much less mobile than NO_3^-. Free NH_3 is quite toxic however, and both NH_3 and NH_4^+ can readily cause eutrophication in N-limited aquatic ecosystems. There is a vast flux of NH_4^+ in most soils via mineralization–immobilization turnover (Booth et al., 2005), but most of this NH_4^+ never leaves the surface horizons of the soil. In contrast to NO_3^-, NH_4^+ is seldom found in groundwater.

The dominant vector of NH_3 transport in the environment is atmospheric. In areas of intensive animal production, rates of NH_3 volatilization are high enough to cause significant loading (e.g., 40–50 kg N ha^{-1} year^{-1}) of adjacent ecosystems by NH_4^+ deposition (Asman et al., 1998; Pinder et al., 2008). Some of the

most dramatic concerns about N saturation of natural ecosystems have arisen in areas of intensive animal production (e.g., the Netherlands). Ammonium is also frequently transported in surface runoff, most commonly in soils that have been treated with organic wastes (Sims et al., 2005).

Similar to NO_3^-, many ecosystems have a high capacity to function as sinks for NH_4^+. In most cases, NH_4^+ is more conservative than NO_3^- because it is less mobile and because there can be strong physicochemical sinks for NH_4^+ on cation exchange sites and in the interlayer spaces of ammonium fixing clays.

27.5.2.3 Nitrous Oxide

Nitrous oxide is produced and/or consumed by several N transformation processes including nitrification, denitrification, and dissimilatory nitrate reduction to ammonia (DNRA). There is environmental concern about N_2O because it is a "greenhouse" gas, which can influence the earth's radiative budget and it plays a role in stratospheric ozone destruction (Mooney et al., 1987; Prather et al., 1995).

Identifying and quantifying sources of N_2O in the landscape is hindered by the high spatial and temporal variability of the dynamics of this gas (Folorunso and Rolston, 1984; Robertson et al., 1988; Parkin, 1993). This variability is caused by the fact that there are multiple processes that affect this gas and there are multiple factors influencing each of these processes.

A variety of approaches have been taken to conceptualize the factors influencing N_2O emissions from ecosystems. Brumme et al. (1999) presented a hierarchy of controls in which long-term change in state variables such as soil and vegetation type generates site to site variation, while short-term changes in physical and chemical conditions such as drying and rewetting and freezing and thawing events produce temporal variation in flux. Mosier et al. (1998) presented a "cradle to the grave" approach that accounts for emission of N_2O as N fertilizer added to agricultural fields moves across the landscape from crop fields to groundwater to streams to rivers to estuaries to the ocean, with emission of N_2O at each step. Tropical soils tend to have higher N_2O fluxes than temperate soils (Stehfest and Bouwman, 2006).

Although N_2O is readily consumed by several processes in soils, most notably by denitrification in wet soils, it is quite stable once it gets into the atmosphere. This stability contributes greatly to the environmental concern about this gas. High concentrations of N_2O are also common in groundwater (Hiscock et al., 2003; Clough et al., 2005). Similar to NO_3^-, wet soils with a high potential for denitrification should be considered to be potential sinks for N_2O (Blicher-Mathiesen and Hoffmann, 1999). In areas where N_2O-rich groundwater interacts with such soils, this potential could be important to landscape N_2O fluxes.

27.5.2.4 Nitric Oxide

Similar to N_2O, NO is a gas that is produced as a by-product of several soil N cycle processes, especially nitrification and denitrification. In contrast to N_2O however, NO is quite reactive and short-lived in the atmosphere. There is environmental concern about NO because it is a precursor to ozone in the lower

atmosphere and contributes to the formation of atmospheric NO_3^- deposition (Guenther et al., 2000). The highly reactive nature of NO makes it even more difficult to predict ecosystems that will be strong NO sources. However, similar to N_2O, ecosystems that are N rich are likely to produce relatively large amounts of NO (Davidson and Verchot, 2000). Wet ecosystems are likely to have a high potential to consume NO.

27.5.2.5 Dissolved Organic N

Dissolved organic N (DON) has generally not been considered to be of environmental concern. However, recent studies have shown that DON is a significant pathway of N loss from many natural (Michalzik et al., 2001) and agricultural (van Kessel et al., 2009) ecosystems that increases in response to N inputs (Pellerin et al., 2006; Brookshire et al., 2007; Fang et al., 2009). Moreover, much of the DON that leaves ecosystems is highly labile and can decompose to inorganic N and cause eutrophication problems in N-limited waters (Seitzinger et al., 2002; Wiegner et al., 2006). It is likely that DON will receive increased attention as an environmental concern in the near future.

27.5.3 Evaluating N Sources and Sinks and Hydrologic Connections in the Landscape

In this section, the author presents a series of questions that allow for evaluation of key N source, sink, and transport processes at any point, in any landscape (Table 27.5). Posing these questions serves two purposes. First, they are a convenient vehicle for summarizing the diverse sources of information on N dynamics that are relevant to evaluating "N in the environment" and second, they provide a practical basis for managing N pollution problems. Consider the landscape depicted in Figure 27.12. To evaluate N in this landscape we need to be able to identify which

TABLE 27.5 Criteria for Determining if a Site Is a Source or Sink of N in the Landscape

Criteria	Determinants
Is the site N rich?	Fertilized
	Fine texture
	Legumes
	Wet tropics
Is the site highly disturbed?	Disturbance of plant uptake (e.g., harvest)
	Stimulation of mineralization (e.g., tillage)
	Disturbance of links between plant and microbial processes (e.g., tillage)
Does the site have a high potential for denitrification?	Wet soil
	High organic matter (especially important in groundwater)
	Fine soil texture
	Well-structured soil (large aggregates with anaerobic centers)
Does the site have a high potential for NH_3 volatilization?	High pH (>8.0)

A—Upland source area
B—Riparian zone

1—Surface runoff
2—Perched water
3—Groundwater flow
4—Underflow or deep recharge
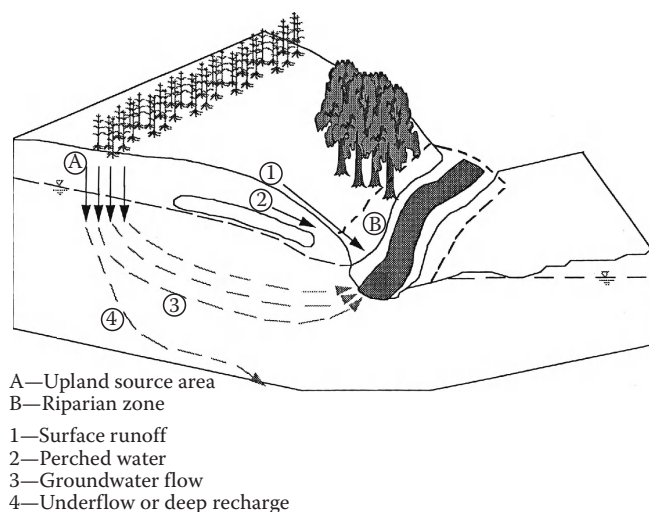
FIGURE 27.12 Conceptual diagram of a landscape showing different hydrologic flowpaths from uplands to riparian zones to streams. (Reprinted from Schnabel, R.R., W.J. Gburek, and W.L. Stout. 1994. Evaluating riparian zone control on nitrogen entry into northeast streams, p. 423–435. *In* Riparian ecosystems in the humid U.S.: Functions, values and management. Proc. Conf. March 15–18, 1993. Atlanta, GA. National Association of Conservation Districts, Washington, DC. Used with permission of National Association of Conservation Districts.)

ecosystems are sources of N, how different forms of N move from one ecosystem to the other and which ecosystems function as sinks for N. Note that Figure 27.12 does not show atmospheric transport. The following questions and text are designed to provide a basis for "thinking" about N in a landscape context and to point to key landscape-scale factors that should be considered in evaluating N in the environment.

1. Is this site (ecosystem) potentially a sink or source of N in the landscape?

There are several relatively simple factors (summarized in Table 27.5) that can be used to determine if a particular site is a "source" of N in the landscape. The primary factor is to determine if "this is an inherently N-rich site" where N supply via mineralization exceeds the capacity for plant uptake. In natural ecosystems, inherent N richness is strongly linked to soil texture and to a lesser extent to parent material. Ecosystems that develop on coarse-textured soils tend to have vegetation that exhibits adaptations to frequent water stress. These adaptations (e.g., thick cuticle) result in the production of low-quality plant litter that results in low rates of N availability (Pastor et al., 1984). Coarse-textured soils also hold less organic matter and support lower microbial biomass than fine-textured soils (Paul, 2007), which also reduces the inherent N richness of a site.

All fertilized ecosystems can be considered to be N rich, but there are marked differences in the N source strength of different fertilized ecosystems, for example, fertilized forests will tend to be much weaker sources of N in the landscape than fertilized agricultural fields. Much of this difference is related to the nature and extent of fertilization and to the disturbance regime

of the site, for example, fertilized agricultural sites lose more N than fertilized forests because they are more highly disturbed. In addition to fertilizer, sites subject to N inputs via symbiotic fixation by legumes, high rates of N deposition, or high rates of groundwater or surface runoff N loading should also be considered to be potential sources of N in the landscape. Tropical soils tend to be more N rich than temperate zone soils (Vitousek and Howarth, 1991).

Next to N richness, the dominant factor controlling if a site is an N source in the landscape is disturbance regime. There is a large body of literature describing the effects of disturbances ranging from clear-cutting to tillage to grazing to acid rain on N losses from natural and agricultural ecosystems (Foster et al., 1997; Randall and Mulla, 2001). It is important to note that not all disturbances foster N losses from all ecosystems. The effect of disturbance on N cycling is dependent both on the nature and extent of the disturbance and on properties of the ecosystem being disturbed, for example, clear-cutting of forests is more likely to produce N losses from inherently N-rich forests than from inherently N-poor forests (Vitousek et al., 1982). Disturbances that alter key components of the N cycle are most likely to cause N loss. Key disturbances of this type include anything that reduces plant uptake (e.g., clear-cutting, harvest, herbiciding, grazing, fire), stimulates microbial mineralization activity (e.g., tillage), or disrupts linkages between microbial mineralization and plant uptake (e.g., tillage, harvest, earthworm activity, fire). It is important to stress that the effects of disturbance on N loss are often transient and/or difficult to predict. For example, many disturbances that reduce plant uptake in the short term (e.g., clear-cutting reduces plant uptake for several months, prairie fires decrease plant uptake for several weeks), increase uptake in the mid and long term (years for clear-cutting, months for prairie fire).

In addition to criteria for predicting if a site is a source of N in the landscape (N richness, disturbance regime), there are several factors that can be evaluated to determine if a site is a potential "sink" for N. Obviously, if a site is inherently N poor (e.g., N-poor soils, young actively growing vegetation), it will have a high potential to absorb inorganic N and transform it into organic N forms that are of less environmental concern. In addition to considering N richness, it is important to consider if a site is particularly supportive of key processes, for example, denitrification, DNRA, that consume reactive N forms of environmental concern. There is particular interest in sites with high volumes of anaerobic soil, for example, wetlands, that have a high potential to support denitrification and/or DNRA (Burgin and Hamilton, 2007). Fine-textured soils tend to have higher potential to support anaerobic conditions and denitrification than coarse-textured soils. Well-structured soils, with large aggregates capable of supporting anaerobic centers also support denitrification.

Evaluation of site sink potential is complicated in several ways however. First, not all inherently N-poor sites will be effective sinks for N, for example, sites on coarse-textured soils transmit water so fast that N is not taken up by plants or microbes very effectively (Vidon and Hill, 2004). Second, it is important

to recognize that consumption of one N form of environmental concern can lead to production of another form of concern, for example, denitrification can convert NO_3^- into N_2O. Finally, there can be a big difference between the actual and potential sink strength of an ecosystem within a landscape. If an ecosystem with a high potential to act as an N sink is not situated in a position to receive N from landscape N sources, it will not function as a sink. This final factor is discussed below.

While, it is clear that there are several readily observable criteria (N richness, disturbance regime, wet soil) that can be used to predict if a site is an N source or sink in the landscape, it is also clear that there are numerous complicating factors that influence the nature and extent of source and sink strengths that motivate more sophisticated evaluation techniques. The most common techniques involve simulation modeling that synthesizes the diverse factors affecting N dynamics to produce an integrated depiction of the dynamics of N at a site. There are numerous site or ecosystem-scale models of N dynamics for both agricultural (Tonitto et al., 2007) and natural (Band et al., 1993) ecosystems. A review of these models, which vary widely in their approach, emphasis, complexity, and ease of application, is well beyond the scope of this chapter. The key challenge for landscape-scale assessments of "N in the environment" is the ease with which a particular model can be used to identify sources and sinks within the landscape and then to link these sources and sinks via hydrologic or atmospheric flows (Tague, 2009). Identifying multiple sources and sinks in a landscape is challenging to site or ecosystem-scale models because soil, plant, and management conditions can vary widely across the landscape (e.g., only a few models are capable of simulating N dynamics in both agricultural and natural components of the landscape). Linking sources and sinks via hydrologic or atmospheric flows with site or ecosystem-scale models is difficult because these models seldom consider transport processes at scales relevant to internal N dynamics. Some efforts at true landscape-scale modeling of N dynamics are described below.

2. How does N enter and/or leave this site via hydrologic or atmospheric vectors?

Once we have determined if a site has an inherent potential to function as a sink or source of N in the landscape, we need to examine the nature and extent of the connections between this site and adjacent landscape components. The first part of this examination relates to the way that water moves into, through, and out of a site and to the potential of a site to absorb or transmit N to the atmosphere.

Hydrologic transport into, through, and out of a site is influenced primarily by soil texture, structure, and drainage and by slope, properties that are discussed extensively elsewhere in this volume. For the purposes of this chapter, our objective is only to highlight key properties that predict the hydrologic nature of a site as the basis for more detailed investigations (Table 27.6).

Sites with coarse soil texture have a high capacity for infiltration relative to sites with fine texture. As a result, coarse-textured soils are likely to produce leaching and movement of

TABLE 27.6 Critical Variables Affecting the Movement of N into, through, and out of a Site

Variable	Status	Results
Texture	Coarse	Infiltration—movement to groundwater
	Fine	Surface runoff
Structure	Well	Infiltration—movement to groundwater
	Poor	Surface runoff
Drainage	Well	Infiltration—movement to groundwater
	Poor	Surface runoff
Soil cover	High	Infiltration—movement to groundwater
	Low	Erosion by wind and water
pH	High	NH_3 volatilization
	Low	NH_3 stabilization as NH_4^+
		Potential for high yield of N_2O and NO

N to groundwater, while fine-textured soils are more likely to produce movement of N in surface runoff. Texture effects on infiltration are strongly modified by soil structure and cover, both of which increase the capacity for infiltration. Soil drainage class, which is controlled by the height of the water table in the soil profile, is a good predictor of water movement into and out of a site to the extent that saturated (more poorly drained) soils are less able to absorb incoming water. Slope has obvious effects on the transmission of incoming water.

The soil variables that most strongly influence the potential of a site to transmit N via atmospheric vectors are soil cover and pH. Soil cover strongly influences the potential of a site to have wind erosion. Sites with high soil pH (>8.0) have a high potential to produce NH_3 gas via volatilization. Sites with low soil pH (<5.0) have a relatively high potential to produce N_2O and NO.

3. Where does this site sit in the landscape?

The juxtaposition of N sources and sinks in the landscape is clearly a major determinant of overall landscape N flux. Figure 27.12 illustrates why riparian forests have been a major focus in research on controlling N outputs from agricultural watersheds (Lowrance et al., 1984; Vidon and Hill, 2006), that is, the presence of an ecosystem with a high potential to act as an N sink (an unfertilized forest, with wet soil) in a position to absorb N moving in both surface runoff and groundwater can have a big impact on watershed N output.

Figure 27.12 also illustrates the difficulties in carrying out quantitative landscape-scale evaluations of N flux. These evaluations require models capable of depicting internal and input/output N dynamics in diverse ecosystems as well as landscape scale hydrologic processes (Tague, 2009). These requirements are a challenge to modelers who need to balance the need for detail in their models with the need to be comprehensive. These requirements also demand large amounts of high-resolution input data on soil, hydrologic, and ecological conditions (Baker et al., 2007). Linking simulation models with geographic information systems to carry out landscape-scale assessments of N flux is rapidly developing into a major approach for managing N in the environment (Groffman et al., 2009).

4. What are subsurface conditions in this site?

Landscape-scale assessments of N flux must consider subsurface processes. The most common concern about N in the environment is NO_3^- pollution of groundwater. Figure 27.12 shows the importance of considering the physical, chemical, and biological conditions of the subsurface material that NO_3^--laden groundwater passes through as it moves through the landscape. While some studies have found high potential and/or actual denitrification in subsurface material others have found little or no activity (Korom, 1992; Rivett et al., 2008).

Most of the uncertainty around subsurface processes concerns the presence of energy sources to support microbial processes. A major challenge for the next few years is to develop a basis for predicting N dynamics in the subsurface based on factors such as soil type, geologic setting, or regional hydrologic conditions. There is also an important need to determine if there are functional links between groundwater and surface soil processes.

27.5.4 Conclusions

It is clear that we have much of the information needed to assess the dynamics of "N in the environment." Our detailed knowledge of N cycle processes (described in other sections of this chapter) allows us to predict and manage internal N dynamics in ecosystems. The challenge, however, is to apply this knowledge at the landscape and watershed scales, which is where we need to assess and manage N pollution problems. In a conceptual sense, we need to develop the ability to "think" at larger scales; to consider landscape units such as those depicted in Figure 27.12 as our objects of study rather than the single crop fields or forest stands that have been the traditional focus of much soil research. In a practical sense, we need to modify existing models and/or develop new models that are capable of dealing with the wide range of sites and ecosystems within a landscape and that are capable of incorporating true landscape-scale hydrologic and atmospheric transport phenomena. These conceptual and practical challenges will likely be quite important in many areas of soil science over the next couple of decades.

References

Aber, J.D., C.L. Goodale, S.V. Ollinger, M.L. Smith, A.H. Magill, M.E. Martin, R.A. Hallett, and J.L. Stoddard. 2003. Is nitrogen deposition altering the nitrogen status of northeastern forests? Bioscience 53:375–389.

Aber, J.D., K.J. Nadelhoffer, P. Steudler, and J.M. Melillo. 1989. Nitrogen saturation in northern forest ecosystems. Bioscience 39:378–386.

Asman, W.A.H., M.A. Sutton, and J.K. Schjorring. 1998. Ammonia: Emission, atmospheric transport and deposition. New Phytol. 139:27–48.

Baker, M.E., D.E. Weller, and T.E. Jordan. 2007. Effects of stream map resolution on measures of riparian buffer distribution and nutrient retention potential. Landsc. Ecol. 22:973–992.

Band, L.E., P. Patterson, R. Nemani, and S.W. Running. 1993. Forest ecosystem processes at the watershed scale: Incorporating hillslope hydrology. Agric. For. Meteorol. 63:93–126.

Blicher-Mathiesen, G., and C.C. Hoffmann. 1999. Denitrification as a sink for dissolved nitrous oxide in a freshwater riparian fen. J. Environ. Qual. 28:257–262.

Boesch, D.F., R.B. Brinsfield, and R.E. Magnien. 2001. Chesapeake Bay eutrophication: Scientific understanding, ecosystem restoration, and challenges for agriculture. J. Environ. Qual. 30:303–320.

Bohlen, P.J., D.M. Pelletier, P.M. Groffman, T.J. Fahey, and M.C. Fisk. 2004. Influence of earthworm invasion on redistribution and retention of soil carbon and nitrogen in northern temperate forests. Ecosystems 7:13–27.

Booth, M.S., J.M. Stark, and E. Rastetter. 2005. Controls on nitrogen cycling in terrestrial ecosystems: A synthetic analysis of literature data. Ecol. Monogr. 75:139–157.

Bormann, F.H., and G.E. Likens. 1979. Pattern and process in a forested ecosystem. Springer-Verlag, New York.

Brookshire, E.N.J., H.M. Valett, S.A. Thomas, and J.R. Webster. 2007. Atmospheric N deposition increases organic N loss from temperate forests. Ecosystems 10:252–262.

Bruland, G.L., C.M. Bliss, S. Grunwald, N.B. Comerford, and D.A. Graetz. 2008. Soil nitrate-nitrogen in forested versus non-forested ecosystems in a mixed-use watershed. Geoderma 148:220–231.

Brumme, R., W. Borken, and S. Finke. 1999. Hierarchical control on nitrous oxide emission in forest ecosystems. Global Biogeochem. Cycles 13:1137–1148.

Burgin, A.J., and S.K. Hamilton. 2007. Have we overemphasized the role of denitrification in aquatic ecosystems? A review of nitrate removal pathways. Front. Ecol. Environ. 5:89–96.

Chapman, S.K., J.A. Langley, S.C. Hart, and G.W. Koch. 2006. Plants actively control nitrogen cycling: Uncorking the microbial bottleneck. New Phytol. 169:27–34.

Clough, T.J., R.R. Sherlock, and D.E. Rolston. 2005. A review of the movement and fate of N_2O in the subsoil. Nutr. Cycl. Agroecosyst. 72:3–11.

Colman, B.P., N. Fierer, and J.P. Schimel. 2008. Abiotic nitrate incorporation, anaerobic microsites, and the ferrous wheel. Biogeochemistry 91:223–227.

Davidson, E.A., S.C. Hart, C.A. Shanks, and M.K. Firestone. 1991. Measuring gross nitrogen mineralization, immobilization, and nitrification by N-15 isotopic pool dilution in intact soil cores. J. Soil Sci. 42:335–349.

Davidson, E.A., and L.V. Verchot. 2000. Testing the hole-in-the-pipe model of nitric and nitrous oxide emissions from soils using the TRAGNET database. Global Biogeochem. Cycles 14:1035–1043.

Eshleman, K.N., R.P. Morgan, J.R. Webb, F.A. Deviney, and J.N. Galloway. 1998. Temporal patterns of nitrogen leakage from mid-Appalachian forested watersheds: Role of insect defoliation. Water Resour. Res. 34:2005–2016.

Fang, Y.T., W.X. Zhu, P. Gundersen, J.M. Mo, G.Y. Zhou, and M. Yoh. 2009. Large loss of dissolved organic nitrogen from nitrogen-saturated forests in subtropical China. Ecosystems 12:33–45.

Fitzhugh, R.D., G.M. Lovett, and R.T. Venterea. 2003. Biotic and abiotic immobilization of ammonium, nitrite, and nitrate in soils developed under different tree species in the Catskill Mountains, New York, USA. Global Change Biol. 9:1591–1601.

Folorunso, O.A., and D.E. Rolston. 1984. Spatial variability of field measured denitrification gas fluxes. Soil Sci. Soc. Am. J. 48:1214–1219.

Foster, D.R., J.D. Aber, J.M. Melillo, R.D. Bowden, and F.A. Bazzaz. 1997. Forest response to disturbance and anthropogenic stress. Bioscience 47:437–445.

Galloway, J.N., J.D. Aber, J.W. Erisman, S.P. Seitzinger, R.W. Howarth, E.B. Cowling, and B.J. Cosby. 2003. The nitrogen cascade. Bioscience 53:341–356.

Galloway, J.N., F.J. Dentener, D.G. Capone, E.W. Boyer, R.W. Howarth, S.P. Seitzinger, G.P. Asner et al. 2004. Nitrogen cycles: Past, present, and future. Biogeochemistry 70:153–226.

Galloway, J.N., A.R. Townsend, J.W. Erisman, M. Bekunda, Z.C. Cai, J.R. Freney, L.A. Martinelli, S.P. Seitzinger, and M.A. Sutton. 2008. Transformation of the nitrogen cycle: Recent trends, questions, and potential solutions. Science 320:889–892.

Gold, A.J., W.R. Deragon, W.M. Sullivan, and J.L. Lemunyon. 1990. Nitrate–nitrogen losses to groundwater from rural and suburban land uses. J. Soil Water Conserv. 45:305–310.

Groffman, P., K. Butterbach-Bahl, R. Fulweiler, A. Gold, J. Morse, E. Stander, C. Tague, C. Tonitto, and P. Vidon. 2009. Challenges to incorporating spatially and temporally explicit phenomena (hotspots and hot moments) in denitrification models. Biogeochemistry 92:49–77.

Guenther, A., C. Geron, T. Pierce, B. Lamb, P. Harley, and R. Fall. 2000. Natural emissions of non-methane volatile organic compounds; carbon monoxide, and oxides of nitrogen from North America. Atmos. Environ. 34:2205–2230.

Hiscock, K.M., A.S. Bateman, I.H. Muhlherr, T. Fukada, and P.F. Dennis. 2003. Indirect emissions of nitrous oxide from regional aquifers in the United Kingdom. Environ. Sci. Technol. 37:3507–3512.

Holland, E.A., B.H. Braswell, J. Sulzman, and J.F. Lamarque. 2005. Nitrogen deposition onto the United States and western Europe: Synthesis of observations and models. Ecol. Appl. 15:38–57.

Jones, K.B., A.C. Neale, M.S. Nash, R.D. Van Remortel, J.D. Wickham, K.H. Riiters, and R.V. O'Neill. 2001. Predicting nutrient discharges and sediment loadings to streams from landscape metrics: A multiple watershed study from the United States Mid-Atlantic region. Landsc. Ecol. 16:301–312.

King, R.S., M.E. Baker, D.F. Whigham, D.E. Weller, T.E. Jordan, P.F. Kazyak, and M.K. Hurd. 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams. Ecol. Appl. 51:137–152.

Korom, S.F. 1992. Natural denitrification in the saturated zone—A review. Water Resour. Res. 28:1657–1668.

Lewis, G.P., and G.E. Likens. 2007. Changes in stream chemistry associated with insect defoliation in a Pennsylvania hemlock-hardwoods forest. For. Ecol. Manage. 238:199–211.

Lovett, G.M., L.M. Christenson, P.M. Groffman, C.G. Jones, J.E. Hart, and M.J. Mitchell. 2002. Insect defoliation and nitrogen cycling in forests. Bioscience 52:335–341.

Lowrance, R., R. Todd, J. Fail, O. Hendrickson, R. Leonard, and L. Asmussen. 1984. Riparian forests as nutrient filters in agricultural watersheds. Bioscience 34:374–377.

Michalzik, B., K. Kalbitz, J.H. Park, S. Solinger, and E. Matzner. 2001. Fluxes and concentrations of dissolved organic carbon and nitrogen—A synthesis for temperate forests. Biogeochemistry 52:173–205.

Mooney, H.A., P.M. Vitousek, and P.A. Matson. 1987. Exchange of materials between terrestrial ecosystems and the atmosphere. Science 238:926–932.

Mosier, A.R., C. Kroeze, C. Nevison, O. Oenema, S. Seitzinger, and O. Van Cleemput. 1998. Closing the global atmospheric N_2O budget: Nitrous oxide emissions through the agricultural nitrogen cycle. Nutr. Cycl. Agroecosyst. 52:225–248.

National Atmospheric Deposition Program. 2008. National Atmospheric Deposition Program 2007 Annual Summary. NADP Data Report 2008-01. Illinois State Water Survey. University of Illinois at Urbana-Champaign, Champaign, IL.

National Research Council. 1993. Soil and water quality: An agenda for agriculture. National Academy Press, Washington, DC.

Nolan, B.T., and J.D. Stoner. 2000. Nutrients in groundwaters of the conterminous United States 1992–1995. Environ. Sci. Technol. 34:1156–1165.

Parkin, T.B. 1993. Spatial variability of microbial processes in soil—A review. J. Environ. Qual. 22:409–417.

Pastor, J., J.D. Aber, C.A. McClaugherty, and J.M. Melillo. 1984. Above-ground production and N and P cycling along a nitrogen mineralization gradient on Blackhawk Island, Wisconsin. Ecology 65:256–268.

Paul, E.A. (ed.). 2007. Soil microbiology, ecology, and biochemistry. 3rd edn. Academic Press, Amsterdam, the Netherlands.

Pellerin, B.A., S.S. Kaushal, and W.H. McDowell. 2006. Does anthropogenic nitrogen enrichment increase organic nitrogen concentrations in runoff from forested and human-dominated watersheds? Ecosystems 9:852–864.

Pinder, R.W., A.B. Gilliland, and R.L. Dennis. 2008. Environmental impact of atmospheric NH3 emissions under present and future conditions in the eastern United States. Geophys. Res. Lett. 35:L12808.

Prather, M., R. Derwent, D. Ehhalt, P.J. Fraser, E. Sanhueza, and X. Zhou. 1995. Other trace gases and atmospheric chemistry, p. 73–126. *In* J.T. Houghton, L.G. Meiro Filho, B.A. Callander, N. Harris, A. Kattenburg, and K. Maskell (eds.) Climate change 1994: Radiative forcing of climate change and an evaluation of the IPCC IS92 emission scenarios. Cambridge University Press, New York.

Rabalais, N.N., R.E. Turner, and W.J. Wiseman. 2002. Gulf of Mexico hypoxia, aka "The dead zone." Annu. Rev. Ecol. Syst. 33:235–263.

Randall, G.W., and D.J. Mulla. 2001. Nitrate nitrogen in surface waters as influenced by climatic conditions and agricultural practices. J. Environ. Qual. 30:337–344.

Rivett, M.O., S.R. Buss, P. Morgan, J.W.N. Smith, and C.D. Bemment. 2008. Nitrate attenuation in groundwater: A review of biogeochemical controlling processes. Water Res. 42:4215–4232.

Robertson, G.P., M.A. Huston, F.C. Evans, and J.M. Tiedje. 1988. Spatial variability in a successional plant community: Patterns of nitrogen availability. Ecology 69:1517–1524.

Ryther, J.H., and W.M. Dunstan. 1971. Nitrogen, phosphorus, and eutrophication in the coastal marine environment. Science 171:1008–1013.

Schaefer, D.A., W.H. McDowell, F.N. Scatena, and C.E. Asbury. 2000. Effects of hurricane disturbance on stream water concentrations and fluxes in eight tropical forest watersheds of the Luquillo Experimental Forest, Puerto Rico. J. Trop. Ecol. 16:189–207.

Schlesinger, W.H. 2009. On the fate of anthropogenic nitrogen. Proc. Natl. Acad. Sci. USA 106:203–208.

Schnabel, R.R., W.J. Gburek, and W.L. Stout. 1994. Evaluating riparian zone control on nitrogen entry into northeast streams, p. 423–435. *In* Riparian ecosystems in the humid U.S.: Functions, values and management. Proc. Conf. March 15–18, 1993. Atlanta, GA. National Association of Conservation Districts, Washington, DC.

Seitzinger, S., J.A. Harrison, J.K. Bohlke, A.F. Bouwman, R. Lowrance, B. Peterson, C. Tobias, and G. Van Drecht. 2006. Denitrification across landscapes and waterscapes: A synthesis. Ecol. Appl. 16:2064–2090.

Seitzinger, S.P., R.W. Sanders, and R. Styles. 2002. Bioavailability of DON from natural and anthropogenic sources to estuarine plankton. Limnol. Oceanogr. 47:353–366.

Sims, J.T., L. Bergstrom, B.T. Bowman, and O. Oenema. 2005. Nutrient management for intensive animal agriculture: Policies and practices for sustainability. Soil Use Manage. 21:141–151.

Stehfest, E., and L. Bouwman. 2006. N_2O and NO emission from agricultural fields and soils under natural vegetation: Summarizing available measurement data and modeling of global annual emissions. Nutr. Cycl. Agroecosyst. 74:207–228.

Tague, C. 2009. Modeling hydrologic controls on denitrification: Sensitivity to parameter uncertainty and landscape representation. Biogeochemistry 92:79–90.

Tonitto, C., M.B. David, C.S. Li, and L.E. Drinkwater. 2007. Application of the DNDC model to tile-drained Illinois agroecosystems: Model comparison of conventional and diversified rotations. Nutr. Cycl. Agroecosyst. 78:65–81.

U.S. EPA. 1990. Criteria document for nitrate/nitrite. Office of Drinking Water, Washington, DC.

van Kessel, C., T. Clough, and J.W. van Groenigen. 2009. Dissolved organic nitrogen: An overlooked pathway of nitrogen loss from agricultural systems? J. Environ. Qual. 38:393–401.

Vidon, P.G.F., and A.R. Hill. 2004. Landscape controls on nitrate removal in stream riparian zones. Water Resour. Res. 40:W03201.

Vidon, P.G., and A.R. Hill. 2006. A landscape-based approach to estimate riparian hydrological and nitrate removal functions. J. Am. Water Resour. Assoc. 42:1099–1112.

Vitousek, P.M., J.D. Aber, R.W. Howarth, G.E. Likens, P.A. Matson, D.W. Schindler, W.H. Schlesinger, and D.G. Tilman. 1997. Human alteration of the global nitrogen cycle: Sources and consequences. Ecol. Appl. 7:737–750.

Vitousek, P.M., J.R. Gosz, C.C. Grier, J.M. Melillo, and W.A. Reiners. 1982. A comparative analysis of potential nitrification and nitrate mobility in forest ecosystems. Ecol. Monogr. 52:155–177.

Vitousek, P.M., and R.W. Howarth. 1991. Nitrogen limitation on land and in the sea—How can it occur? Biogeochemistry 13:87–115.

Wiegner, T.N., S.P. Seitzinger, P.M. Glibert, and D.A. Bronk. 2006. Bioavailability of dissolved organic nitrogen and carbon from nine rivers in the eastern United States. Aquat. Microb. Ecol. 43:277–287.

28
Molecular Techniques

Judith Ascher
University of Florence

Maria Teresa Ceccherini
University of Florence

Yin Chen
University of Warwick

Guo-Chun Ding
*Institute for Epidemiology
and Pathogen Diagnostics*

Holger Heuer
*Institute for Epidemiology
and Pathogen Diagnostics*

Jiri Jirout
University of South Bohemia

Deepak Kumaresan
University of Warwick

J. Colin Murrell
University of Warwick

Giacomo Pietramellara
University of Florence

Kornelia Smalla
*Institute for Epidemiology
and Pathogen Diagnostics*

28.1 Cultivation-Independent Detection of Genes Present in Soil Bacteria

Guo-Chun Ding

Holger Heuer

Kornelia Smalla

The development of methods to extract DNA directly from soil samples opened new dimensions of studies on microbial communities in soil. The molecular analysis of soil DNA can provide information on the microbial community composition or on the presence and on the relative abundance of functional genes, for example, of genes involved in N-cycling, antibiotic resistance genes, degrading genes, or plasmid replicon-related sequences.

DNA-based analysis of soils led to the discovery of many taxa that are abundant in soils but had not been studied before as they do not easily form colonies on plates. This chapter is exclusively devoted to the detection of genes in DNA directly extracted from soil. Crucial prerequisites of the cultivation-independent detection of genes in soil are appropriate sampling strategies and DNA extraction protocols. Therefore, the subsequent two paragraphs will briefly address these topics before discussing the potentials and limitations of methods for targeted or untargeted detection of genes.

28.1.1 Sampling and Sample Processing

The adequate sampling design depends on the hypotheses to be tested and often preexperiments might assist in determining the numbers of samples to be analyzed. Sampling of soil requires a

careful design of a strategy considering the estimated heterogeneity of the soil system and the expected variability (Van Elsas et al., 2002; Smalla and van Elsas, 2010). To overcome variability due to field heterogeneity, composite samples of soil taken from a plot or a field, often in a number of replicates, were analyzed in many of the studies. The main reason for analyzing composite samples is that the sample should be in a representative manner reporting on the structural and functional diversity of a soil under a particular treatment and on the variation among replicates of the same treatment. Each composite sample may consist of several soil cores taken from a plot in a randomized fashion. Mixing of the soil is usually achieved by sieving the material of the various soil cores through a 2 or 4 mm mesh size. A subsample from each composite sample is then used for the analysis. However, the investigation of composite samples may obscure a meaningful biological variation. Thus, in studies aiming to explore field, plot or microscale variability characterization of independent single samples would be required (Sliwinski and Goodman, 2004; Becker et al., 2006).

28.1.1.1 Nucleic Acid Extraction from Soil Samples

Thirty years after the first paper on the extraction of DNA from soil had been published by Torsvik (1980), obtaining nucleic acids from soil matrices that are suitable for molecular analysis remained a challenge (for review; Van Elsas et al., 2000). None of the protocols for DNA extraction from soil seems suitable for all kinds of soils, in particular for soils originating from contaminated sites. Commercial kits for total DNA extraction from soil facilitated largely the procedure in particular in view of a simplification and miniaturization of the method. Soil DNA extraction kits are often less time-consuming and more efficiently remove coextracted humic acids. Nucleic acids can be extracted directly either from the soil matrix or from a microbial pellet obtained after dislodging microbes associated with soil particles or plant roots and subsequent centrifugation. The first protocol for direct extraction of DNA from soil was published by Ogram et al. (1987) and is based on in situ lysis of microbial cells in the presence of the environmental matrix. The second approach, also termed indirect method, was pioneered by Torsvik (1980) and Holben et al. (1988) and involves the extraction of nucleic acids from the microbial fraction. The recovery of the microbial fraction from soils commonly involves repeated homogenization of soil resuspended in aqueous solution and differential centrifugation steps as originally suggested by Fægri et al. (1977). However, protocols differ considerably with respect to the solutions used to break up soil colloids and dislodge surface-attached cells, which adhere to surfaces by various bonding mechanisms such as polymers, electrostatic forces, and water bridging, and of different strengths (Bakken and Lindahl, 1995; Van Elsas et al., 2000). This approach might preferably be used for obtaining nucleic acids from soil adhering to roots (rhizosphere), problematic soils (soils with a high content of humic acids or contaminants), or to obtain high-molecular DNA for generating fosmid or BAC libraries. Homogenization is usually achieved by shaking soil suspensions with gravel or blending, for example, with a Stomacher lab blender. Although a complete dislodgment of cells seems impossible, it is important that cells bound to soil particles with different degrees of strength are released with similar efficiency. A clear advantage of the indirect approach is that nucleic acids recovered are less contaminated with coextracted humic acids, DNA of nonbacterial origin, or extracellular DNA (unless extracellular DNA is tightly adsorbed to microbial cells). In comparison with the indirect method, the direct DNA extraction approach is less time-consuming and much higher DNA yields are achieved (Van Elsas et al., 2000), which might, however, contain considerable amounts of coextracted humic acids.

A crucial step for the recovery of representative DNA that mirrors the genomes of all microbes present in a soil sample (Moré et al., 1994; Miller et al., 1999) is the efficient lysis of the bacterial, archaeal, and fungal cell walls. Cell lysis can be achieved by mechanical cell disruption and by enzymatic or chemical disintegration of cell walls, or a combination of these methods is often used. The efficiency of the different methods used for cell lysis might influence not only the yield but also the presence of genomic DNA of cells difficult to lyse (resting stage, e.g., dwarf cells or spores). However, obviously the strength of lysis needs to be a trade-off as too rigorous lysing methods might shear DNA released from cells that are easy to lyse. Therefore, high-molecular-weight DNA is an important criterion when evaluating and comparing different protocols because sheared DNA can cause PCR artifacts and is not suitable for direct cloning of large DNA fragments. To obtain large DNA fragments is particularly important for generation of large insert libraries (e.g., fosmid or BAC libraries). In plug lysis of the microbial pellet and pulse field gel electrophoresis is most frequently used to obtain large DNA fragments (Robe et al., 2003). The complete removal of coextracted humic acids is critical for the subsequent molecular analysis as humic acids were shown to interfere with DNA hybridization, restriction enzyme digestions, and PCR amplification (Tebbe and Vahjen, 1993). The DNA yield might vary considerably for different DNA extraction kits used for the same soils. As directly extracted DNA can be of different origin, it is recommended to estimate the DNA yield not only on ethidium bromide–stained agarose gels or by measuring the ratio 260/280 but also by determining the 16S or 18S rRNA gene copy numbers by quantitative PCR. A recently performed comparison of two frequently used soil DNA extraction kits revealed that the 16S rRNA gene copy numbers were approximately two orders of magnitude different. Differences in DNA yield might be no problem if bacterial, archaeal, or fungal ribosomal gene fragments are amplified by PCR. However, if the analysis is targeting less abundant taxa or genes that are carried only by a small proportion of the microbial community, then the difference in yield will be critical. Last but not least, it should be stressed that strict precautionary measures need to be taken to prevent contamination of the DNA during the extraction either directly from soil or from the microbial pellet. In particular, when PCR amplification is used to amplify a target gene that occurs less frequently,

for example, antibiotic resistance genes or transgenic DNA in soil, the extraction of DNA, preparation of PCRs, and analysis of PCR products need to be strictly separated.

28.1.2 Targeted Detection of Genes in Nucleic Acids from Soil

28.1.2.1 Detection of Genes Providing Information on the Structural Diversity

PCR-based amplification of 16S, 18S rRNA, or ITS gene fragments from soil DNA and their subsequent analysis either by cloning and sequencing, direct sequencing, or by fingerprinting methods are most frequently used to study the composition of Bacteria, Archaea, or Fungi in soil. A considerable advantage is the rapidly growing database of rRNA gene sequences, which contains presently more than 710,000 long high quality 16S rRNA gene sequence entries deposited in Ribosomal Database (http://rdp.cme.msu.edu, release 10.26). A disadvantage of using rRNA gene fragments is that bacteria possess different numbers of rRNA operons that are assumed to reflect different ecological strategies of bacteria (Klappenbach et al., 2000) and that sequence heterogeneity of the different operons might occur (Nübel et al., 1996; von Wintzingerode et al., 1997). Alternative markers such as the gene coding for the σ factor *rpoB* or *gacA* have been suggested since these genes are present only as a single copy and were shown to allow a higher resolution between species (Dahllöf et al., 2000; Costa et al., 2007). However, no matter which gene is targeted, one major limitation that remains is that gene fragments of less common populations are often not represented in clone libraries or fingerprints, especially when primers targeting all bacteria are used for amplification. Bent and Forney (2008) termed this problem "the tragedy of the uncommon" and critically discussed in this context the use of diversity indices. The application of group- or genus-specific primers targeting the 16S rRNA gene can assist in studying less common populations in soil (Heuer et al., 1997, 2001; Costa et al., 2006a; Weinert et al., 2009). In order to be detected by PCR-based methods, the gene copy number of a population should be above 100 or 1000 cells g⁻¹ of soil as usually one-hundredth or less of the DNA extracted from a gram of soil is used as template for PCR amplification. Thus, the diversity of uncommon populations in soil might become only detectable after soil enrichment steps.

The 16S or 18S rRNA gene or ITS fragments amplified from total community DNA with Bacteria-, Archaea-, or Fungi-specific primers were traditionally characterized by cloning and sequencing. Janssen (2006) compared 32 different clone libraries generated from various soils and found that 16S rRNA gene sequences from soil DNA could be affiliated to 32 phyla but that the dominant phyla in the libraries belonged to nine phyla only, that is, the Proteobacteria, Acidobacteria, Actinobacteria, Verrucomicrobia, Bacteroidetes, Chloroflexi, Planctomycetes, Gemmatimonadetes, and Firmicutes. Interestingly, the phylum Firmicutes, including the classes Bacilli and Clostridia that were considered to be typical soil bacteria, contributed up to only 2% of

the sequences in the clone libraries. However, the composition of the clone libraries, that is, the relatively low abundance of the Firmicutes, might have been strongly affected by DNA extraction biases as mentioned above. Another rather recent methodological development is the use of pyrosequencing to study the diversity of bacterial 16S rRNA gene fragments PCR amplified from soil DNA (Roesch et al., 2007; Fulthorpe et al., 2008; Jones et al., 2009). These studies provided the largest datasets on rRNA gene sequences obtained for soil so far. Furthermore, these data indicated that the number of species per gram of soil estimated by Gans et al. (2005) was an overestimate. More importantly, in-depth analysis of soil microbial diversity by sequencing was shown to be no longer impracticable as suggested by Gans et al. (2005). However, despite the enormous amount of data acquired by pyrosequencing, 10,000 sequences are still a very small fraction of >10⁹ bacteria living in 1 g of soil. Therefore, more attention should be paid to statistical exploration of these datasets. First of all, how well does the dataset reflect the real diversity? Table 28.1 demonstrates the relationship among the relative abundance of a species, the total number of sequences, and the chance to detect this species. For example, the relative abundance of a species needs to be at least 0.03% to be detected in a sample of 10,000 sequences with a chance of 95%. This corresponds to more than a million cells per gram of soil. However, many soil bacteria will have a lower abundance, at least temporarily or locally. A second important question is how well the dataset from pyrosequencing can represent the community structure. The simulation results in Table 28.2 show that there will be a significant error in predicting the relative abundance of rare species. A relative species abundance of less than 0.01% will be predicted with an error of more than 100%. Obviously, it is still very difficult to reliably predict the community structure also with respect to minorities.

To study spatial and temporal variation of microbial communities in relation to environmental factors, perturbation, or experimental treatment, multiple sample analysis is essential (Muyzer and Smalla, 1998; Smalla et al., 2001, 2007; Forney et al., 2004; Kowalchuk et al., 2006). For this purpose, approaches based on cloning and sequencing or direct sequencing of 16S rRNA gene fragments amplified from community DNA are still

TABLE 28.1 Relative Abundance of a Species That Is Needed to Detect Its Sequence within a Total Number of Sequences at a Specified Likelihood, Based on a Poisson Distribution

Likelihood of Detection	Minimal Relative Abundance (%)			
	Total Number of Sequences			
	10	100	1,000	10,000
0.1	1.0	0.11	0.01	0.001
0.3	3.5	0.36	0.04	0.004
0.5	6.7	0.69	0.07	0.007
0.7	11.3	1.20	0.12	0.012
0.9	20.6	2.28	0.23	0.023
0.95	25.9	2.95	0.30	0.030
0.99	36.9	4.50	0.46	0.046

TABLE 28.2 Uncertainty to Predict the Community Structure from Pyrosequencing Data (Coefficient of Variance from 1000 Simulations to Sample Nine Species of a Particular Relative Abundance)

Relative Abundance (%)	Error of Species Abundance Prediction (%)			
	Total Number of Sequences			
	1,000	2,000	10,000	20,000
10	9.7	6.9	3.1	2.2
1	30.9	21.5	9.5	7.0
0.1	95.4	66.8	30.8	21.9
0.01	262.1	209.6	101.7	71.7

too costly and labor intensive, especially for terrestrial habitats of a large microbial diversity. Although a rapid development in the field of high-throughput sequencing techniques is witnessed making pyrosequencing more and more affordable and feasible, the use of molecular fingerprinting techniques like denaturing gradient gel electrophoresis and temperature gradient gel electrophoresis (DGGE and TGGE; Muyzer and Smalla, 1998), single-strand conformation polymorphism (SSCP; Schwieger and Tebbe, 1998), terminal-restriction fragment analysis of 16S ribosomal genes or fragments amplified from total community DNA, that is, terminal-restriction fragment length polymorphism (T-RFLP; Liu et al., 1997; Osborn et al., 2000), or automated rRNA intergenic spacer analysis (ARISA; Ranjard et al., 2001) remains the state of the art here (Table 28.3). The molecular fingerprinting techniques based on 16S or 18S rRNA gene or ITS fragments or on alternative marker genes amplified from total community DNA still represent the best compromise between the number of processed samples and the level of information that is obtained. An advantage of microbial community fingerprint analysis is the quick overview of the composition of dominant sequences amplified, which reflect the relative abundance of bacterial populations or a specific taxonomic group. When combined with the analysis of sequences of differentiating bands, information on the phylogenetic placement can be rapidly gathered. Recently, Smalla et al. (2007) showed that the analysis of 16S rRNA gene fragments amplified from soil DNA with bacteria-specific primers by means of DGGE, SSCP,

or T-RFLP resulted in comparable results. Although the target regions varied, DGGE, T-RFLP, and SSCP analyses all led to similar clustering. In this study, little variability in the bacterial community composition of the four replicate composite samples analyzed per soil was observed while significant differences among the different soils were found that mirrored soil physicochemical properties such as pH, clay content, or organic C. In a continental-scale study, including 98 soil samples from North and South America, T-RFLPs of 16S rRNA gene fragments amplified from total community DNA were used to compare bacterial communities across sites (Fierer and Jackson, 2006). The study showed that the composition of soil bacterial communities differed between ecosystems and that most of the differences could be explained by soil pH.

The resolving power of current PCR-based community fingerprinting methods can be considerably increased using taxon-specific primers. During the PCR with universal primers, a particular target sequence should be present at least at about 0.1%–1% of the total target sequences to be amplified and appear as a visible band in a fingerprint. To analyze the less common ribotypes and also reduce the complexity of the patterns, nested or seminested PCR approaches with taxon-specific primers have been quite successfully used (Heuer et al., 1997; Gomes et al., 2001, 2005; Boon et al., 2002; Salles et al., 2002; Garbeva et al., 2003; Freitag et al., 2005; Costa et al., 2006a, 2006b; Weinert et al., 2009). Different studies that compared DGGE patterns based on 16S rRNA gene fragments amplified from DNA or cDNA (following RNA extraction and a reverse transcription step) from soil, reported that DNA- and RNA-based fingerprints of soil bacterial communities are often rather similar (Nicol et al., 2003; Gomes et al., 2005).

Recently, 16S rRNA gene fragments amplified from soil DNA were also analyzed by so-called PhyloChips that enable the parallel detection of more than 8700 operational taxonomic units (OTUs) and the ability to identify individual OTU varying by five orders of magnitude in abundance (Brodie et al., 2006; DeSantis et al., 2007). Typically, more than 2000 OTU can be detected in soil DNA. In the study by DeAngelis et al. (2009), the rhizosphere effect was studied by PhyloChip analysis. OTUs responding to the root exudates were detected in 17 of the 44 phyla. PhyloChip analysis of 16S rRNA gene

TABLE 28.3 Molecular Techniques Used to Compare and Analyze Microbial Communities from Soil

Technique	Principles	Reference
DGGE and TGGE	PCR-amplified fragments of 16S/18S rRNA, *rpoB*, *gacA*, or other genes are separated based on their melting in a denaturing gradient formed by chemicals (DGGE) or temperature (TGGE)	Muyzer and Smalla (1998)
SSCP	Digestion of one DNA strand of PCR-amplified fragments, and electrophoretic separation based on the sequence determined conformation of the remaining single strand	Schwieger and Tebbe (1998)
T-RFLP	PCR amplification of gene fragments using fluorescently labeled primers, cutting off the labeled terminal fragments using restriction enzymes, and separation on a sequencer based on the size of the terminal fragments	Liu et al. (1997)
ARISA	PCR amplification of intergenic spacers of ribosomal operons using fluorescently labeled primers targeting rRNA genes, and separation on a sequencer based on the size of the fragments	Ranjard et al. (2001)
Pyrosequencing of rRNA genes	Massive sequencing of PCR-amplified 300–500 base fragments by synthesizing a complementary strand and detection of pyrophosphate release on nucleotide incorporation	Roesch et al. (2007)

fragments enables a comparison of multiple samples and a determination of taxa responding to biotic or abiotic factors. The great advantage of PhyloChips compared to other fingerprinting methods is that information on the taxonomic affiliation of these responders can be obtained.

However, the molecular fingerprints described above provide no quantitative data on the abundance of the taxa detected. Here, the real-time determination of 16S rRNA gene copy numbers, for example, of Bacteria or Archaea or of specific groups is the method of choice. Particular characteristics of the 16S rRNA gene, such as those caused by the presence of different numbers of rRNA operons per genome (from 1 to up to 15), in different species prevent that the copy number determined can be directly translated to cell number. Real-time PCR formats that are most commonly used are SYBR Green detection and TaqMan probes. SYBR Green binds unspecifically to all double-stranded DNA, and, thus, product formation is monitored by an increased fluorescence. In contrast, the TaqMan probe assay is based on the specific binding of the double-labeled probe (quencher and reporter dye) to a target sequence and the cleavage of the probe by the Taq polymerase endonuclease activity. Upon cleavage, the reporter dye is no longer quenched and the signal can be detected. Again, a sufficient abundance of the target sequence and absence of potential PCR-inhibiting substances, such as coextracted humic acids, need to be considered for a reliable real-time PCR detection from soil.

28.1.2.2 Targeted Detection of Functional Genes

PCR-based detection of functional genes in soil has widely been used in several studies. The presence of different functional genes in soil bacteria has usually been studied by PCR amplification from total DNA and subsequent cloning and sequencing. If sufficient amounts of PCR products can be obtained, the analysis of PCR amplicons by cloning and sequencing or by molecular fingerprints can provide information on the diversity of respective genes, for example, *pmoA* (Horz et al., 2001), *dsr* (Schmalenberger et al., 2007), bacterial or archaeal *amoA* (Nicol et al., 2008; Jia and Conrad, 2009), *ndo* genes (Gomes et al., 2007; Flocco et al., 2009), or chitinase genes (Metcalfe et al., 2002).

If the gene targeted is present above a critical detection threshold, then its presence can be quantified by real-time PCR (Heuer and Smalla, 2007; Hai et al., 2009; Heuer et al., 2009; Schauss et al., 2009). In particular, the development of real-time PCR assays enabled studies relating gene abundance to various different biotic or abiotic factors. Again the detection limit is an important issue. Only if gene copies occur with more than 10^5–10^6 copies per gram of soil, amplicons visible in ethidium bromide–stained agarose gels are obtained. The detection limit is comparable for real-time PCR. Although SYBR Green is widely employed for the detection of functional genes by quantitative PCR, the use of TaqMan probes might increase specificity and sensitivity of the system. In many studies, the copy number determined is related to a gram of soil or even microliter of DNA. However, it is strongly recommended that the copy number determined is related to 16S rRNA gene copy numbers to overcome biases caused by uncertainties of the total soil DNA composition. Although comparable amounts of DNA are obtained per gram of soil, the number of 16S rRNA gene copy numbers might vary considerably (Heuer et al., 2009). If the gene targeted is less abundant in the soil population, then it is recommendable to increase the sensitivity of its detection by Southern blot hybridization. Even if amplicons are not visible on the agarose gel, they can be detected by means of hybridization with digoxigenin or ^{32}P-labeled probes. Although the data obtained are at the most semiquantitative, the sensitivity of gene detection is often increased by two orders of magnitude. In addition, by choosing the stringency of the hybridization conditions, sequence specificity of the signal can be confirmed. The presence of different functional genes in soil bacteria has been studied by PCR amplification from total DNA and subsequent analysis by Southern blot analysis, for example, *merA*, *trfA* (Smalla et al., 2006), *nahA*, *nahH*, and *phnAc* (Gomes et al., 2005, 2007).

The spectrum of primers is also a very important issue as the diversity of many genes is still unknown and many novel sequences could be frequently found from environmental samples. The potential for microbial degradation of polycyclic aromatic hydrocarbons (PAHs) in the environment is of major interest due to the bioaccumulation and toxicity of these compounds. The diversity of PAH-degrading genes in soils and how it is influenced by different soil types or their history and type of pollution is not yet well explored, because the existing primer systems have severe limitations. In Table 28.4, published primers that detect PAH-ring-hydroxylating dioxygenase genes are summarized. Most of the primers target only a rather narrow range of sequences, for example, *nahAc* or *phnAc* type sequences, or only genes from Gram-negative bacteria. One primer system could detect dioxygenase genes of both Gram-positive and Gram-negative bacteria, but the specificity is broader than PAH dioxygenases and the short amplified fragment does not resolve much of the diversity (Ni Chadhain et al., 2006). To overcome these limitations, recently accumulating sequence data need to be used to develop better primer systems to determine the diversity and abundance of PAH-ring-hydroxylating dioxygenase genes of soil microbial communities.

28.1.3 Nontargeted Detection of Genes

Functional genes have been detected in soil DNA by the so-called functional arrays. However, the performance of microarray hybridization from complex environments such as soil has to be carefully evaluated and obviously a number of technical challenges need to be overcome before this technique can be fully exploited. A microarray with 100 functional genes was used by Wu et al. (2001) to systematically study the specificity, sensitivity, and quantification of microarray hybridization with DNA from complex environmental samples. When environmental DNA is used without prior PCR amplification, this seems to be the most difficult challenge. The level of detection is 1,000- to 10,000-fold lower than with PCR amplification (Wu et al., 2001). In order to achieve a comprehensive investigation of functional

TABLE 28.4 PCR Systems to Study the Diversity and Abundance of PAH-RHDα Genes in Soil

Primers	Target Genes	Amplicon Size (bp)	Sample Sources	Sample Type and Analysis	Reference
NAPH-1F/-1R; NAPH-2F/-2RGC	*nahAc*, *phnAc*, *nagAc*, and *ndo* (Gram-negative bacteria)	896; 740	PAH-contaminated mangrove sediments (Brazil) or soils (Maritime Antarctic)	Total community DNA, DGGE, and clone library	Gomes et al. (2007)
FRT5A/FRT3B; FRT6A/FRT4B	PAH-RHDα genes of Gram-negative bacteria; PAH-RHDα genes of Gram-negative bacteria	437; 491	Long-term and short-term oil-contaminated microbial mats	Total community DNA and clone library	Bordenave et al. (2008)
Nah-for, Nah-rev1/-rev2; Nid-for, Nid-rev1/-rev2	PAH-RHDα genes *nahAc*, *phnAc*, Gram-negative bacteria; PAH-RHDα genes of Gram-positive bacteria	937, 317; ~600, ~310	PAH enrichment of mangrove sediments, China	PAH-degrading isolates, PCR, and sequencing	Zhou et al. (2006)
PAH-RHDα GN, F610a + R16a; PAH-RHDα GP, F641b + R933	PAH-RHDα genes of Gram-negative bacteria; PAH-RHDα genes of Gram-positive bacteria	360; 292	Five kinds of PAH-contaminated soil with different texture, France	Total community DNA, clone library and real-time PCR	Cebron et al. (2008)
Rieske_f; Rieske_r	(1,2,4–6) sulfur center of PAH-RHDα genes	78	PAH enrichment from contaminated soils, USA	Total DNA from enriched cell pellets, clones	Ni Chadhain et al. (2006)
Cyc372F; Cyc854R	*phnA1*-like genes	479–482	PAH-contaminated intertidal sediments	Total community DNA and clone library	Lozada et al. (2008)
nahAcfor/-rev; P8073/P9047	*nahAc*-like genes; *phnAc*-like genes	1009; 974	PAH-contaminated soil microcosms, New Zealand	Total community DNA, clones and real-time PCR	Laurie and Lloyd-Jones (2000)
Ac114F; Ac596R	*nahAc*-like genes	482	Groundwater, possibly contaminated	RNA from extracted cells, real-time PCR, and dot blot	Wilson et al. (1999)
NAH-F; NAH-R	*nahAc*-like genes	377	Isolates	Genomic DNA, real-time PCR, and hybridization	Baldwin et al. (2003)
nidA-F/-R; nahAc-F/-R	*nidA*; *nahAc*	~100	Coal tar-contaminated sediments	Total community DNA and real-time PCR (TaqMan probes)	Debruyn et al. (2007)
nagAc-like-F; nagAc-like-R	*nagAc*-like genes	107	Coal tar-contaminated freshwater sediments	Total community DNA, real-time PCR (TaqMan probe)	Dionisi et al. (2004)

genes, so-called GeoChip microarrays containing over 24,000 probes covering more than 10,000 genes distributed among more than 150 functional groups involved in nitrogen, carbon, sulfur, and phosphorus cycling were recently employed for soil studies (Gentry et al., 2006; Wu et al., 2006; He et al., 2007). The direct hybridizations were, however, not very sensitive, and to increase the sensitivity of the approach, a preamplification step (e.g., using rolling circle amplification) is now included. Despite numerous challenges that need to be solved, the GeoChip will be a powerful tool that allows linking functional microbial dynamics to particular ecosystem processes. Probe development, hybridization quality, and data evaluation are crucial steps for an appropriate use of DNA microarrays to study the soil microbiota. The GeoChip was applied to study N- and C-cycle genes in DNA extracted from soil taken from different sites in the Antarctic (Yergeau et al., 2007). The N- and C-cycle genes detected differed significantly across sampling locations and vegetation types. However, DNA microarrays cannot generate information on new sequence types, and, thus, only the breadth of known functions can be assessed (Yergeau et al., 2007). A recent study of Suenaga et al. (2009) gives an idea on the severity of this limitation. They created a metagenomic library with DNA extracted from activated sludge and sequenced those clones with extradiol dioxygenase activity. By sequence analysis, they found that most of these clones (36 of 38) contained complete aromatic degradation pathways that shared low similarity to those found in known cultured bacteria metabolizing aromatic compounds. Having learned that soil bacterial communities can be extremely diverse, it will be a challenge to include all potentially important genetic variants of a particular soil function on a microarray. Furthermore, the abundance of most bacteria may be below the detection limit of microarrays, some of them having the potential to take over important functions when environmental conditions are advantageous for them. To increase the sensitivity, community DNA was frequently subjected to whole-genome amplification before microarray analysis.

Another nontargeted approach is cloning large fragments of environmental DNA into fosmid or BAC vectors and subsequent screening of the metagenomic libraries. Metagenomics offers access to functional genes in the soil microbiota and thus to novel bioactive products (Sjöling et al., 2007). This approach has been successfully applied to recover DNA coding for so far unknown enzymes or antibiotics directly from soil DNA (Rondon et al., 2000; Gillespie et al., 2002; Sjöling et al., 2007). However, despite the vast potential of this approach, several methodological challenges remain to be solved. To obtain sequence information of particular uncultured taxa, which are assumed to be abundant in the soil sample, in a first step, specific primers are used to identify the clones carrying the respective genes, either based on 16S rRNA or even on functional genes. Using this approach, genomic fragments of uncultured Acidobacteria or Archaea could be obtained and subjected to sequence-based gene detection (Ochsenreiter et al., 2003; Treusch et al., 2005; Van Elsas et al., 2008a, 2008b). Sequencing of a large fosmid library from a grassland soil revealed that a fragment with an rRNA gene, which showed the highest similarity to group 1.1b of the

crenarchaeota, carried genes encoding copper nitrite reductase and two subunits of ammonia monooxygenase or particulate methane monooxygenase (Treusch et al., 2005). The discovery of ammonia-oxidizing Archaea showed the potential of soil metagenomics studies to elucidate the role of uncultured organisms in, for example, soil nutrient cycles (Leininger et al., 2006). Considering that the majority of soil bacteria occur at relatively low abundance but might have an important ecological role, the development of approaches such as targeted access to low-abundance bacteria are of considerable interest. Recently, Hjort et al. (2010) analyzed chitinase genes in DNA directly extracted from a phytopathogen-suppressive soil, in a metagenomic library constructed from microbial cells extracted from soil, and in genomic DNA from bacterial isolates with antifungal and chitinase activities. Although T-RFLP of chitinase genes revealed differences in amplified chitinase genes from the metagenomic library and the directly extracted DNA, approximately 40% of the identified chitinase T-RFs were found in both sources. All of the chitinase T-RFs from the isolates were matched to T-RFs in the directly extracted DNA and the metagenomic library. The authors demonstrated an impressive agreement between three very different screening techniques all of which pointed toward specific *Streptomyces* species that could play a role in suppression of fungi by chitinase production in soil. However, some clusters of chitinase genes were represented depending on the approach used. Therefore, Hjort et al. (2010) concluded that a combination of molecular approaches increases the information obtained and the reliability of the data.

28.1.4 Cultivation-Independent Detection of Genes Carried by Mobile Genetic Elements

The importance of horizontal gene exchange for short-term bacterial adaptability and for successful colonization of new ecological niches has only recently been appreciated (Heuer et al., 2008). Nowadays, mobile genetic elements (MGEs) are recognized as an important and essential component that promotes bacterial diversity. The PCR-based detection of MGE-specific sequences in community DNA was first used by Götz et al. (1996). Primers targeting replicon-specific sequences were designed on the basis of sequenced broad-host range plasmids or integrons. In combination with Southern blot hybridization, a specific and sensitive monitoring of large numbers of environmental samples became possible (Smalla et al., 2000, 2006; Heuer et al., 2002, 2009). While the targeted detection of MGE by endpoint or real-time PCR requires sequence information, capturing MGE directly from soil microbial pellets and subsequent sequencing allow a nontargeted discovery and detection of genes. Direct capturing of MGE by means of so-called exogenous plasmid isolation techniques (Bale et al., 1988; Hill et al., 1992) that have been widely used to capture MGE conferring selectable traits such as mercury or antibiotic resistance have been carried out in a wide range of soils or sediments in Gram-negative recipients functioning as a genetic sink (Heuer et al.,

2002, 2009; Smalla et al., 2006; Binh et al., 2008). Capturing of degradative genes resident on MGE has been demonstrated as well (Top et al., 1995). Increased transfer frequencies have often been observed when the soil environmental sample was previously exposed to pollutants. The PCR-based detection of MGE has been used for monitoring their abundance in soils (Götz et al., 1996; Smalla et al., 2000; Heuer et al., 2009), and "hot spots" with high abundance of MGE could be identified.

28.1.5 Conclusion and Outlook

The tools for cultivation-independent detection of genes in soil bacteria rapidly advanced over the last few years. The technological developments briefly described in this chapter brought us the opportunity to study the enormous complexity of soil microbial communities in more comprehensive and complete terms. The use of molecular tools has dramatically changed our view of the microbes residing in soil. Most importantly, it became possible not only to study who is there but also—although still a methodological challenge—what they are doing. The advances of soil metagenomic tools will continue to improve our understanding of soil microbes. To better understand the enormous microbial diversity and the biotic and abiotic factors shaping their composition and functions, it is certainly needed to study the soil microbiota at different scales and also to develop approaches to target the uncommon or rare populations in soil. The rapidly advancing tools will assist to successfully uncover more secrets of microbial life in soil.

References

Bakken, L.R., and V. Lindahl. 1995. Recovery of bacterial cells from soil, p. 13–27. *In* J.T. Trevors and J.D. van Elsas (eds.) Nucleic acids in the environment. Springer, Heidelberg, Germany.

Baldwin, B.R., C.H. Nakatsu, and L. Nies. 2003. Detection and enumeration of aromatic oxygenase genes by multiplex and real-time PCR. Appl. Environ. Microbiol. 69:3350–3358.

Bale, M.J., M.J. Day, and J.C. Fry. 1988. Novel method for studying plasmid transfer in undisturbed river Epilithon. Appl. Environ. Microbiol. 54:2756–2758.

Becker, J.M., T. Parkin, C.H. Nakatsu, J.D. Wilbur, and A. Konopka. 2006. Bacterial activity, community structure, and centimeter-scale spatial heterogeneity in contaminated soil. Microb. Ecol. 51:220–231.

Bent, S.J., and L.J. Forney. 2008. The tragedy of the uncommon: Understanding limitations in the analysis of microbial diversity. ISME J. 2:689–695.

Binh, C.T.T., H. Heuer, M. Kaupenjohann, and K. Smalla. 2008. Piggery manure used for soil fertilization is a reservoir for transferable antibiotic resistance plasmids. FEMS Microbiol. Ecol. 66:25–37.

Boon, N., W. de Windt, W. Verstraete, and E.M. Top. 2002. Evaluation of nested PCR-DGGE (denaturing gradient gel electrophoresis) with group-specific 16S rRNA primers for the analysis of bacterial communities from different wastewater treatment plants. FEMS Microbiol. Ecol. 39:101–112.

Bordenave, S., M. Goni-Urriza, C. Vilette, S. Blanchard, P. Caumette, and R. Duran. 2008. Diversity of ring-hydroxylating dioxygenases in pristine and oil contaminated microbial mats at genomic and transcriptomic levels. Environ. Microbiol. 10:3201–3211.

Brodie, E.L., T.Z. DeSantis, D.C. Joyner, S.M. Baek, J.T. Larsen, G.L. Andersen, T.C. Hazen et al. 2006. Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. Appl. Environ. Microbiol. 72:6288–6298.

Cebron, A., M.P. Norini, T. Beguiristain, and C. Leyval. 2008. Real-time PCR quantification of PAH-ring hydroxylating dioxygenase (PAH-RHDα) genes from Gram positive and Gram negative bacteria in soil and sediment samples. J. Microbiol. Methods 73:148–159.

Costa, R., N.C.M. Gomes, E. Krögerrecklenfort, K. Opelt, G. Berg, and K. Smalla. 2007. *Pseudomonas* community structure and antagonistic potential in the rhizosphere: Insights gained by combining phylogenetic and functional gene-based analyses. Environ. Microbiol. 9:2260–2273.

Costa, R., M. Götz, N. Mrotzek, J. Lottmann, G. Berg, and K. Smalla. 2006a. Effect of site and plant species on rhizosphere community structure as revealed by molecular analysis of different microbial guilds. FEMS Microbiol. Ecol. 56:236–249.

Costa, R., J.F. Salles, G. Berg, and K. Smalla. 2006b. Cultivation-independent analysis of *Pseudomonas* species in soil and in the rhizosphere of field-grown *Verticillium dahliae* host plants. Environ. Microbiol. 8:2136–2149.

Dahllöf, I., H. Baillie, and S. Kjelleberg. 2000. *rpoB*-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. Appl. Environ. Microbiol. 66:3376–3380.

DeAngelis, K.M., E.L. Brodie, T.Z. DeSantis, G.L. Andersen, S.E. Lindow, and M.K. Firestone. 2009. Selective progressive response of soil microbial community to wild oat roots. ISME J. 3:168–178.

Debruyn, J.M., C.S. Chewning, and G.S. Sayler. 2007. Comparative quantitative prevalence of mycobacteria and functionally abundant *nidA*, *nahAc*, and *nagAc* dioxygenase genes in coal tar contaminated sediments. Environ. Sci. Technol. 41:5426–5432.

DeSantis, T.Z., E.L. Brodie, J.P. Moberg, I.X. Zubieta, Y.M. Piceno, and G.L. Andersen. 2007. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. Microb. Ecol. 53:371–383.

Dionisi, H.M., C.S. Chewning, K.H. Morgan, F.M. Menn, J.P. Easter, and G.S. Sayler. 2004. Abundance of dioxygenase genes similar to *Ralstonia* sp. strain U2 *nagAc* is correlated with naphthalene concentrations in coal tar-contaminated freshwater sediments. Appl. Environ. Microbiol. 70:3988–3995.

Fægri, A., V.L. Torsvik, and J. Goksøyr. 1977. Bacterial and fungal activities in soil: Separation of bacteria and fungi by a rapid fractionated centrifugation technique. Soil Biol. Biochem. 9:105–112.

Fierer, N., and R.B. Jackson. 2006. The diversity and biogeography of soil bacterial communities. PNAS 103:626–631.

Flocco, C.G., N.C.M. Gomes, W.M. Cormack, and K. Smalla. 2009. Occurrence and diversity of naphthalene dioxygenase genes in soil microbial communities from the Maritime Antarctic. Environ. Microbiol. 11:700–714.

Forney, L.J., X. Zhou, and C.J. Brown. 2004. Molecular microbial ecology: Land of the one-eyed king. Curr. Opin. Microbiol. 7:210–220.

Freitag, T.E., L. Chang, C.D. Clegg, and J.I. Prosser. 2005. Influence of inorganic nitrogen management regime on the diversity of nitrite-oxidizing bacteria in agricultural grassland soils. Appl. Environ. Microbiol. 71:8323–8334.

Fulthorpe, R.R., L.F.W. Roesch, A. Riva, and E.W. Triplett. 2008. Distantly sampled soils carry few species in common. ISME J. 2:901–910.

Gans, J., M. Wolinsky, and J. Dunbar. 2005. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science 309:1387–1390.

Garbeva, P., J.A. Van Veen, and J.D. Van Elsas. 2003. Predominant *Bacillus* spp. in agricultural soil under different management regimes detected via PCR-DGGE. Microb. Ecol. 45:302–316.

Gentry, T.J., G.S. Wickham, C.W. Schadt, Z. He, and J. Zhou. 2006. Microarray applications in microbial ecology research. Microb. Ecol. 52:159–175.

Gillespie, D.E., S.F. Brady, A.D. Bettermann, N.P. Cianciotto, M.R. Liles, M.R. Rondon, J. Clardy, R.M. Goodman, and J. Handelsman. 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl. Environ. Microbiol. 68:4301–4306.

Gomes, N.C.M., L.R. Borges, R. Paranhos, F.N. Pinto, E. Krögerrecklenfort, L.C.S. Mendonça-Hagler, and K. Smalla. 2007. Diversity of *ndo* genes in mangrove sediments exposed to different sources of polycyclic aromatic hydrocarbon pollution. Appl. Environ. Microbiol. 73:7392–7399.

Gomes, N.C.M., H. Heuer, J. Schönfeld, R. Costa, L. Hagler-Mendonça, and K. Smalla. 2001. Bacterial diversity of the rhizosphere of maize (*Zea mays*) grown in tropical soil studied by temperature gradient gel electrophoresis. Plant Soil 232:167–180.

Gomes, N.C.M., I.A. Kosheleva, W.-R. Abraham, and K. Smalla. 2005. Effects of the inoculant strain *Pseudomonas putida* KT2442 (pNF142) and of naphthalene contamination on the soil bacterial community. FEMS Microbiol. Ecol. 54:21–33.

Götz, A., R. Pukall, E. Smit, E. Tietze, R. Prager, H. Tschäpe, J.D. van Elsas, and K. Smalla. 1996. Detection and characterization of broad-host range plasmids in environmental bacteria by PCR. Appl. Environ. Microbiol. 62:2621–2628.

Hai, B., N.H. Diallo, S. Sall, F. Haesler, K. Schauss, M. Bonzi, K. Assigbetse, J.-L. Chotte, J.C. Munch, and M. Schloter. 2009. Quantification of key genes steering the microbial nitrogen cycle in the rhizosphere of 2 sorghum cultivars in tropical agro-ecosystems. Appl. Environ. Microbiol. 75:4993–5000.

He, Z., T.J. Gentry, C.W. Schadt, L. Wu, J. Liebich, S.C. Chong, Z. Huang et al. 2007. GeoChip: A comprehensive microarray for investigating biogeochemical, ecological and environmental processes. ISME J. 1:67–77.

Heuer, H., Z. Abdo, and K. Smalla. 2008. Patchy distribution of flexible genetic elements in bacterial populations mediates robustness to environmental uncertainty. FEMS Microbiol. Ecol. 65:361–371.

Heuer, H., C. Kopmann, C.T.T. Binh, E.M. Top, and K. Smalla. 2009. Spreading antibiotic resistance through spread manure: Characteristics of a novel plasmid type with low %G + C content. Environ. Microbiol. 11:937–949.

Heuer, H., E. Krögerrecklenfort, S. Egan, L.S. van Overbeek, G. Guillaume, T.L. Nikolakopoulou, E.M.H. Wellington et al. 2002. Gentamicin resistance genes in environmental bacteria: Prevalence and transfer. FEMS Microbiol. Ecol. 42:289–302.

Heuer, H., M. Krsek, P. Baker, K. Smalla, and E.M.H. Wellington. 1997. Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. Appl. Environ. Microbiol. 63:3233–3241.

Heuer, H., and K. Smalla. 2007. Manure and sulfadiazine synergistically increased bacterial antibiotic resistance in soil over at least two months. Environ. Microbiol. 9:657–666.

Heuer, H., G. Wieland, J. Schönfeld, A. Schönwälder, N.C.M. Gomes, and K. Smalla. 2001. Bacterial community profiling using DGGE or TGGE analysis, p. 177–190. *In* P. Rouchelle (ed.) Environmental molecular microbiology: Protocols and applications. Horizon Scientific Press, Wymondham, UK.

Hill, K.E., A.J. Weightman, and J.C. Fry. 1992. Isolation and screening of plasmids from the Epilithon which mobilize recombinant plasmid pD10. Appl. Environ. Microbiol. 58:1292–1300.

Hjort, K., M. Bergström, M.F. Adesina, J.K. Jansson, K. Smalla, and S. Sjöling. 2010. Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen suppressive soil. FEMS Microbiol. Ecol. 71:197–207.

Holben, W.E., J.K. Jansson, B.K. Chelm, and J.M. Tiedje. 1988. DNA probe method for the detection of specific microorganisms in the soil bacterial community. Appl. Environ. Microbiol. 54:703–711.

Horz, H.-P., M.T. Yimaga, and W. Liesack. 2001. Detection of methanotroph diversity on roots of submerged rice plants by molecular retrieval of *pmoA*, *mmoX*, *mxaF* and 16S rRNA and ribosomal DNA, including *pmoA*-based terminal restriction fragment length polymorphism profiling. Appl. Environ. Microbiol. 67:4177–4185.

Janssen, P.H. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl. Environ. Microbiol. 72:1719–1728.

Jia, Z., and R. Conrad. 2009. Bacteria rather than Archaea dominate microbial ammonia oxidation in an agricultural soil. Environ. Microbiol. 11:1658–1671.

Jones, R.T., M.S. Robeson, C.L. Lauber, M. Hamady, R. Knight, and N. Fierer. 2009. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J. 3:442–453.

Klappenbach, J.A., J.M. Dunbar, and T.M. Schmidt. 2000. rRNA operon copy number reflects ecological strategies of bacteria. Appl. Environ. Microbiol. 66:1328–1333.

Kowalchuk, G.A., W.H.G. Hol, and J.A. van Veen. 2006. Rhizosphere fungal communities are influenced by *Senecio jacobaea* pyrrolizidine alkaloid content and composition. Soil Biol. Biochem. 38:2852–2859.

Laurie, A.D., and G. Lloyd-Jones. 2000. Quantification of *phnAc* and *nahAc* in contaminated New Zealand soils by competitive PCR. Appl. Environ. Microbiol. 66:1814–1817.

Leininger, S., T. Urich, M. Schloter, L. Schwark, J. Qi, G.W. Nicol, J.I. Prosser, S.C. Schuster, and C. Schleper. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442:806–809.

Liu, W.-T., T.L. Marsh, H. Cheng, and L.J. Forney. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. 63:4516–4522.

Lozada, M., J.P. Riva Mercadal, L.D. Guerrero, W.D. Di Marzio, M.A. Ferrero, and H.M. Dionisi. 2008. Novel aromatic ring-hydroxylating dioxygenase genes from coastal marine sediments of Patagonia. BMC Microbiol. 8:50.

Metcalfe, A.C., M. Krsek, G.W. Gooday, J.I. Prosser, and E.M.H. Wellington. 2002. Molecular analysis of a bacterial chitinolytic community in an upland pasture. Appl. Environ. Microbiol. 68:5042–5050.

Miller, D.N., J.E. Bryant, E.L. Madsen, and W.C. Ghiorse. 1999. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. Appl. Environ. Microbiol. 65:4715–4724.

Moré, M.I., J.B. Herrick, M.C. Silva, W.C. Ghiorse, and E.L. Madsen. 1994. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. Appl. Environ. Microbiol. 60:1572–1580.

Muyzer, G., and K. Smalla. 1998. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. Antonie van Leeuwenhoek 73:127–141.

Ni Chadhain, S.M., R.S. Norman, K.V. Pesce, J.J. Kukor, and G.J. Zylstra. 2006. Microbial dioxygenase gene population shifts during polycyclic aromatic hydrocarbon biodegradation. Appl. Environ. Microbiol. 72:4078–4087.

Nicol, G.W., L.A. Glover, and J.I. Prosser. 2003. Spatial analysis of archaeal community structure in grassland soil. Appl. Environ. Microbiol. 69:7420–7429.

Nicol, G.W., S. Leininger, and C. Schleper. 2008. The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. Environ. Microbiol. 10:2966–2978.

Nübel, U., B. Engelen, A. Felske, J. Snaidr, A. Wiesenhuber, R.I. Amann, W. Ludwig, and H. Backhaus. 1996. Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. J. Bacteriol. 178:5636–5643.

Ochsenreiter, T., D. Selezi, A. Quaiser, L. Bonch-Osmolovskaya, and C. Schleper. 2003. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. Environ. Microbiol. 5:787–797.

Ogram, A., G.S. Sayler, and T.J. Barkay. 1987. DNA extraction and purification from sediments. J. Microbiol. Methods 7:57–66.

Osborn, A.M., E.R.B. Moore, and K.N. Timmis. 2000. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ. Microbiol. 2:39–50.

Ranjard, L., F. Poly, J.C. Lata, C. Mougel, J. Thioulouse, and S. Nazaret. 2001. Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: Biological and methodological variability. Appl. Environ. Microbiol. 67:4479–4487.

Robe, P., R. Nalin, C. Capellano, T.A. Vogel, and P. Simonet. 2003. Extraction of DNA from soil. Eur. J. Soil Biol. 39:183–190.

Roesch, L.F.W., R.R. Fulthorpe, A. Riva, G. Casella, A.K.M. Hadwin, A.D. Kent, S.H. Daroub, F.A.O. Camargo, W.G. Farmerie, and E.W. Triplett. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. 1:283–290.

Rondon, M.R., P.R. August, A.D. Bettermann, S.F. Brady, T.H. Grossman, M.R. Liles, K.A. Loiacono et al. 2000. Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. 66:2541–2547.

Salles, J.F., F.A. de Souza, and J.D. van Elsas. 2002. Molecular method to assess the diversity of *Burkholderia* species in environmental samples. Appl. Environ. Microbiol. 68:1595–1603.

Schauss, K., A. Focks, S. Leininger, A. Kotzerke, H. Heuer, S. Thiele-Bruhn, S. Sharma et al. 2009. Dynamics and functional relevance of ammonia-oxidizing archaea in two agricultural soils. Environ. Microbiol. 11:446–456.

Schmalenberger, A., H.L. Drake, and K. Küsel. 2007. High unique diversity of sulfate-reducing prokaryotes characterized in a depth gradient in an acidic fen. Environ. Microbiol. 9:1317–1328.

Schwieger, F., and C.C. Tebbe. 1998. A new approach to utilize PCR-single strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. Appl. Environ. Microbiol. 64:4870–4876.

Sjöling, S., W. Stafford, and D.A. Cowan. 2007. Soil metagenomics: Exploring and exploiting the soil microbial gene pool, p. 409–434. *In* J.D. van Elsas, J. Jansson, and J.T. Trevors (eds.) Modern soil microbiology. 2nd edn. CRC Press, Taylor & Francis Group, Boca Raton, FL.

Sliwinski, M.K., and R.M. Goodman. 2004. Spatial heterogeneity of Crenarchaeal assemblages within mesophilic soil ecosystems as revealed by PCR-single-stranded conformation polymorphism profiling. Appl. Environ. Microbiol. 70:1811–1820.

Smalla, K., A.S. Haines, K. Jones, E. Krögerrecklenfort, H. Heuer, M. Schloter, and C.M. Thomas. 2006. Increased abundance of IncP-1β plasmids and mercury resistance genes in mercury-polluted river sediments: First discovery of IncP-1β plasmids with a complex *mer* transposon as the soil accessory element. Appl. Environ. Microbiol. 72:7253–7259.

Smalla, K., H. Heuer, A. Götz, D. Niemeyer, E. Krögerrecklenfort, and E. Tietze. 2000. Exogenous isolation of antibiotic resistance plasmids from piggery manure slurries reveals a high prevalence and diversity of IncQ-like plasmids. Appl. Environ. Microbiol. 66:4854–4862.

Smalla, K., M. Oros-Sichler, A. Milling, H. Heuer, S. Baumgarte, R. Becker, G. Neuber, S. Kropf, A. Ulrich, and C.C. Tebbe. 2007. Bacterial diversity of soils assessed by DGGE, T-RFLP and SSCP fingerprints of PCR-amplified 16S rRNA gene fragments: Do the different methods provide similar results? J. Microbiol. Methods 69:470–479.

Smalla, K., and J.D. van Elsas. 2010. The soil environment, p. 111–130. *In* W.T. Liu and J.K. Janson (eds.) Environmental molecular microbiology. Caister Academic Press, Norfolk, UK.

Smalla, K., G. Wieland, A. Buchner, A. Zock, J. Parzy, S. Kaiser, N. Roskot, H. Heuer, and G. Berg. 2001. Bulk and rhizosphere soil bacterial communities studied by denaturing gradient gel electrophoresis: Plant-dependent enrichment and seasonal shifts revealed. Appl. Environ. Microbiol. 67:4742–4751.

Suenaga, H., Y. Koyama, M. Miyakoshi, R. Miyazaki, H. Yano, M. Sota, Y. Ohtsubo, M. Tsuda, and K. Miyazaki. 2009. Novel organization of aromatic degradation pathway genes in a microbial community as revealed by metagenomic analysis. ISME J. 3:1335–1348.

Tebbe, C.C., and W. Vahjen. 1993. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. Appl. Environ. Microbiol. 59:2657–2665.

Top, E.M., W.E. Holben, and L.J. Forney. 1995. Characterization of diverse 2,4-dichlorophenoxyacetic acid-degradative plasmids isolated from soil by complementation. Appl. Environ. Microbiol. 61:1691–1698.

Torsvik, V. 1980. Isolation of bacterial DNA from soil. Soil Biol. Biochem. 12:15–21.

Treusch, A.H., S. Leininger, A. Kletzin, S.C. Schuster, H.-P. Klenk, and C. Schleper. 2005. Novel genes for nitrite reductase and *amoA*-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. Environ. Microbiol. 7:1985–1995.

Van Elsas, J.D., R. Costa, J.K. Jansson, S. Sjöling, M.J. Bailey, R. Nalin, T.M. Vogel, and L.S. van Overbeek. 2008a. The metagenomics of disease-suppressive soils—Experiences from the METACONTROL project. Trends Biotechnol. 26:591–601.

Van Elsas, J.D., K. Smalla, A.K. Lilley, and M.J. Bailey. 2002. Methods for sampling soil microbes, p. 505–526. *In* C.J. Hurst, R.L. Crawford, G.R. Knudsen, M.J. McInerney, and L.D. Stetzenbach (eds.) Manual of environmental microbiology. 2nd edn. ASM Press, Washington, DC.

Van Elsas, J.D., K. Smalla, and C.C. Tebbe. 2000. Extraction and analysis of microbial community nucleic acids from environmental matrices, p. 29–51. *In* J.K. Jansson, J.D. van Elsas, and M.J. Bailey (eds.) Tracking genetically-engineered microorganisms. Eurekah, Austin, TX.

Van Elsas, J.D., A.J. Speksnijder, and L.S. van Overbeek. 2008b. A protocol for the metagenomic analysis of suppressive soil. J. Microbiol. Methods 75:515–522.

von Wintzingerode, F., U.B. Göbel, and E. Stackebrandt. 1997. Determination of microbial diversity in environmental samples: Pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. 21:213–229.

Weinert, N., R. Meincke, C. Gottwald, H. Heuer, N.C. Gomes, M. Schloter, G. Berg, and K. Smalla. 2009. Rhizosphere communities of genetically modified zeaxanthin-accumulating potato plants and their parent cultivar differ less than those of different potato cultivars. Appl. Environ. Microbiol. 75:3859–3865.

Wilson, M.S., C. Bakermans, and E.L. Madsen. 1999. *In situ*, real-time catabolic gene expression: Extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater. Appl. Environ. Microbiol. 65:80–87.

Wu, L.Y., X. Liu, C.W. Schadt, and J.Z. Zhou. 2006. Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. Appl. Environ. Microbiol. 72:4931–4941.

Wu, L., D.K. Thompson, G. Li, R.A. Hurt, J.M. Tiedje, and J. Zhou. 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl. Environ. Microbiol. 67:5780–5790.

Yergeau, E., S. Kang, Z. He, J. Zhou, and G.A. Kowalchuk. 2007. Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. ISME J. 1:163–179.

Zhou, H.W., C.L. Guo, Y.S. Wong, and N.F. Tam. 2006. Genetic diversity of dioxygenase genes in polycyclic aromatic hydrocarbon-degrading bacteria isolated from mangrove sediments. FEMS Microbiol. Lett. 262:148–157.

28.2 Expression of Genes in Soil

Giacomo Pietramellara

Judith Ascher

Jiri Jirout

Maria Teresa Ceccherini

28.2.1 Introduction

The detection and monitoring of bacterial gene expression in the environment is important in microbial ecology. The use of molecular techniques for quantifying gene expression by transcriptomic and proteomic analysis in complex soil samples can improve the understanding of the role of many soil microbial species including uncultivable microorganisms, which are the majority of microorganisms inhabiting soil (Torsvik et al., 1996).

The progressive cost reduction of some of these techniques has markedly contributed to the widespread use of molecular tools for analyzing gene expression in soil samples. Among these techniques for the analysis of gene expression of both prokaryotic and eukaryotic cells, there are reverse transcription polymerase chain reaction (RT-PCR), northern and western blotting, RNase protection assays (RPAs), stable-isotope probing (SIP), microarrays, and sequencing methods. The consideration of the advantages and drawbacks of these, as well as the emerging molecular techniques, will be important to evaluate the accuracy and the representativity of the obtained data.

A relevant aspect to consider, that is out of the scope of this chapter, is the opportunity of contemporary analysis of DNA and RNA coextracted from soil in order to have a complete characterization of the soil microbial community, permitting also the assessment of active populations of the soil microbial community and its correlation to specific soil functions.

The aim of this chapter is to give a concise review of the main methods used to study gene expression in soil, including those not yet applied to soil but potentially promising, and to discuss the main factors affecting gene expression in soil.

Some of the techniques discussed in this chapter are also discussed in the specific chapter on gene detection of this section, because they can be used both for evaluating the presence and expression of genes. We think that this overlapping is important for a complete overview of methods for detecting presence and expression of genes.

28.2.2 Gene Expression

28.2.2.1 Definition and Relative Processes

The gene itself is typically a long stretch of DNA and does not perform an active role but is a blueprint for the production of RNA. The process called transcription is performed by RNA polymerases, which add one nucleotide to a growing RNA strand. This RNA is complementary to the DNA nucleotide being transcribed, and mRNA is the information carrier coding for the synthesis of one (common in eukaryotes) or more (common in prokaryotes) proteins. In addition to transcription, gene expression includes RNA splicing, translation, and posttranslational modification of a protein (Kirschner, 1999; Graves and Haystead, 2002). Many factors such as the metabolic state of the cell, the local environment, and chemical signals from other cells affect gene expression. Therefore, an expression profile allows one to deduce the state of a cell in its environment. Gene regulation controls timing, location, and rate of gene expression, determining the versatility and adaptability of any organism.

The aim of transcriptomics is to determine the number of expressed genes through the mRNAs detection.

Consequential to gene expression is the metabolome that represents the collection of all the molecules, which are the end products of gene expression. Metabolomics is the systematic study of the unique chemical fingerprints that specific cellular processes leave behind. Metabolomics, indeed, is the quantitative measure of the phenotype of a biological system and with many other "-omics" sciences represents a key aspect of "system biology."

The main metabolomic molecules studied are proteins. The information carried by proteins is represented by its structures (primary, secondary, tertiary, and quaternary), physical size, and electric charge properties that markedly influence the function of the protein. Thus, the characterization of the proteome by proteomics can be based on expression, functionality, and structure of proteins. The expression permits the quantitative study of proteins in relation to changes in biological or environmental parameters (*expression proteomics*). The functionality can identify specific proteins, their role in individual metabolic activities, and their contribution to the metabolic network operating in the system (*functional proteomics*; Hsieh and Chen, 2008). The third approach permits to characterize the structure of proteins affecting their function (*structural proteomics*; Norin and Sundström, 2002).

28.2.2.2 Gene Expression in Soil

Soil is an extremely complex environment in terms of biochemical activities and contains highly diverse and complex microbial communities (Pietramellara et al., 2002, Graeme et al., 2003). For these reasons, to monitor gene expression of the whole soil microbial community, both RNA and proteins should be extracted in sufficient amounts to be representative of the overall microbial activity of soil (Ogram et al., 1995). Moreover, mRNAs show higher instability and represent a small fraction (~1%–5%) of the total cellular RNA (Neidhardt and Umbarger, 1996). Furthermore, the molecular resolution of soil proteins for elucidating species diversity, functions, and level of activity is extremely difficult due to the large number of different proteins present in soil, and also because one gene can codify the synthesis of several proteins. It is relevant to note that a high percentage of soil proteins, approximately 30%–45% of the soil total N, is present as extracellular proteins (Stevenson, 1986), while intracellular proteins only represent, on average, 4% (Nannipieri, 2006). However, characterization of intracellular proteins should give insights into microbial gene expression in soil. Nannipieri (2006)

named this proteomic approach *functional soil proteomics*, whereas the characterization of extracellular proteins stabilized by soil colloids was named *structural soil proteomics*. The latter characterization could give insights into past biological events.

As for soil DNA extraction, both RNA and proteins can be extracted directly from soil or indirectly after separation of microorganisms from soil (Bakken and Frostegård, 2006; Nannipieri, 2006). It is reasonable to hypothesize that the indirect extraction gives high-quality RNA and proteins but the efficiency is much lower than that of the direct extraction (Nannipieri and Paul, 2009). However, the direct extraction presents some problems because both RNA and proteins can be adsorbed by soil colloids and/or degraded by soil microflora (Norde, 2008). Indeed, soil humic substances and clay particles can bind RNA and proteins interfering with their extraction and purification procedures (Ogram et al., 1995; Trevors and van Elsas, 1995; Trevors, 1996; Griffiths et al., 2000; England et al., 2001; Hurt et al., 2001; Sayler et al., 2001; Sessitsch et al., 2002; Bürgmann et al., 2003; England and Trevors, 2003). Enzyme inhibitors in soil can also inhibit culture-independent analyses (Mendum et al., 1998). Furthermore, high temperatures, extreme pH values, and ionic strength can decrease both RNA and protein extraction efficiency and affect the structural integrity of the protein compromising its information content (Saleh-Lakha et al., 2005). Concerning the pH effect on RNA extraction from soil both Mettel et al. (2010) and Wang et al. (2009) suggest the utilization of low pH lysisi buffer (pH 5.0) and organic solvent (pH 4.5) to minimized the humic acids coextraction with any reduction in the RNA recovery.

28.2.3 Methods for Studying Gene Expression in Soil by Transcriptomic Analysis

Methods for studying gene expression can be divided into (1) culture-independent approaches such as DNA microarrays, northern blotting, fluorescent in situ hybridization (FISH; Parker and Barnes, 1999), ribonuclease protection assay (RPA; Hod, 1992; Saccomanno et al., 1992), RT-PCR (Weis and Reinberg, 1992) coupled with real-time PCR and competitive RT-PCR, or (2) culture-dependent approaches such as SIP (see the specific chapter of this book and Section 28.2.3.3), serial analysis of gene expression (SAGE), and reporter genes. Except for FISH and reporter gene techniques, the other techniques require the extraction of mRNA from samples. The problem of the RNA degradation by RNases, eventually present as contaminants in glassware and reagents, can be avoided by sterilization of glassware and by RNase inhibitors such as DEPC (diethylpyrocarbonate) prior to the extraction procedures (Trayhurn, 1996).

28.2.3.1 mRNA Extraction

28.2.3.1.1 Direct Extraction of Total RNA and mRNA from Soil

RNA isolation protocols by direct extraction include cell lysis, inactivation of RNase activity to prevent losses of RNA, and purification of the extracted RNA with removal of organic contaminants that can be coextracted with nucleic acids (Ogram et al., 1995). Cell lysis for the RNA extraction from soil

can be mechanical (ballistic disintegration of cells using glass or zirconium beads), chemical (solubilization of cell membranes by detergents), enzymatic (digestion of the cell wall membranes), or physical (boiling and freeze–thawing cycles) (Ogram et al., 1995; Borneman and Triplett, 1997; Hurt et al., 2001; Bürgmann et al., 2003). The appropriate extraction technique depends on the soil properties (Borneman and Triplett, 1997). A key step in the RNA extraction procedure is the mRNA separation from total RNA. Subtractive hybridization Ribosomal RNA (rRNA) was utilized by Gilbert et al. (2008) to enhance the mRNA enrichment capturing rRNA from total RNA solution by hybridization on magnetic beads coated with selective probes. It is relevant to note that this approach failed to completely remove rRNA due to insufficient compatibility of the capture probes (Mettel et al. 2010) and also due to not fully substraction of damaged rRNA molecule from total RNA (Gilbert et al. 2008). Frias-Lopez et al. (2008) operated a selective RNA enzymatic digestion in order to obtain an mRNA enrichment through a processive 5′-3′ exonuclease treatment. The enzymes utilized in the rRNA digestion were able to discriminate between the rRNA monophosphate 5′-end and mRNA tri-phosphate 5′-end, that characterize most archeal and bacterial mRNAs, degrading the former RNA (Mettel et al. 2010). Relevant to note that more than 30% of mRNA are present as monophosphate 5′-end in consequence of molecular decay after transcription processes (Celesnik et al. 2007). Cosequently the molecular analysis of exonuclease enriched mRNA result mainly direct to unprocessed mRNA fractions, that are triphosphorylated (Mettel et el. 2010). McGrath et al. (2008) developed a simple method to isolate mRNA from total RNA by size separation using gel electrophoresis, and this method yielded a high number of unique transcripts directly extracted from soil.

28.2.3.1.2 Indirect Extraction of Total RNA and mRNA from Soil

The indirect extraction method is based on previous extraction of microbial cells from soil providing an opportunity to examine the physiological state of extracted cells and to identify individual genes. This extraction requires physical or chemical dispersion of soil followed by a separation of microbial cells from soil particles. The main physical technique used to separate microbial cells from soil particles are shaking, blending, stomacher, and ultrasonication, whereas the chemical techniques include the use of ion-exchange resins, which can reduce the electrostatic attraction between cells and soil particles. To separate cells from soil particles, *buoyant density* can be also used, with the assumption that most cells are not aggregated and can be detached from soil particles. This method is based on ultracentrifugation and the most commonly used density material is Nycodenz (Bakken and Lindahl, 1995; Unge et al., 1999; Berry et al., 2003). Another cell–soil separation technique is the immunomagnetic capture based on the capture of microbial cells by microscopic magnetic beads coated with monoclonal or polyclonal antibodies (Morgan et al., 1991; Fluit et al., 1993; Porter et al., 1998).

It is important to underline that the yield and quality of extracted cells depend on the choice of the most suitable method according to the soil characteristics.

28.2.3.2 Reverse Transcription Polymerase Chain Reaction Technologies

Reverse transcription polymerase chain reaction permits the detection of low amounts of RNA and is easy to be performed with small reaction volumes; the total isolated RNA is subjected to reverse transcriptase to produce complementary DNA (cDNA) copies that are used as templates in a PCR with probes designed to amplify the target gene(s).

28.2.3.2.1 Competitive Reverse Transcription PCR

The competitive reverse transcription polymerase chain reaction (cRT-PCR) permits to quantify the expression of specific genes (McGrath et al., 2000). The technique is based on the competitive coamplification in the same reaction tube of a known amount of an internal standard (competitor) with a sequence of interest and a known amount of the total RNA extracted from the sample. The internal standard has the same primer-binding sites as the target sequence but is usually modified by a small deletion, insertion, or mutation to distinguish it by electrophoresis (Zimmerman and Mannhalter, 1996). Quantification is accomplished by the intensity of the PCR signal of the internal standard; as the amount of internal standard increases, the signal of the target sequence decreases (McGrath et al., 2000; Sharkey et al., 2004). In cRT-PCR, internal standard and target sequences are amplified with equal efficiency, and consequently, their products accumulate with the same kinetics, even if the PCR reagents are present in limiting concentrations.

28.2.3.2.2 Quantitative PCR

Quantitative PCR (qPCR) can be done either by measuring the amount of PCR products at a given number of cycles (endpoint qPCR; Gertsch et al., 2002; Spackman and Suarez, 2008; Udvardi et al., 2008; Van Guilder et al., 2008) or by measuring the amount of products during several PCR cycles (real-time qPCR; Nailis et al., 2006; Logan et al., 2009). These two techniques have been reviewed by Lantz et al. (1999). Here, we only discuss the real-time qPCR as the most utilized quantitative PCR technique.

The real-time qPCR monitors DNA amplification during each PCR cycle through the use of the fluorescence, whose intensity is directly related to the amount of amplicons at the end of each PCR cycle. Therefore, qPCR overcomes the shortcomings of conventional PCR, allowing the quantification of the initial number of gene copies present in the sample, even at a concentration of one genome copy per microliter of soil DNA extract (Gruntzig et al., 2001; Lerat et al., 2005).

28.2.3.2.3 Reverse Transcription PCR/PCR Suppression Subtractive Hybridization

Suppression subtractive hybridization (SSH) permits to detect the influence of environmental variables on gene expression in selected bacteria. This technique is based on the selective hybridization of cDNA, obtained from different environmental samples, with mRNA extracted from a reference sample. Successively, double-stranded cDNA–mRNA hybrids are selectively removed

and the composition of single-stranded cDNA (cDNAss) in the supernatant is analyzed (Utt et al., 1995; Rebrikov, 2003).

28.2.3.2.4 Arbitrarily Primed PCR

Several strategies have been developed for the analysis of small samples in which the amount of available RNA is insufficient for massive gene expression studies (Trenkle et al., 1998; Welford et al., 1998; Sakai et al., 2000; Vernon et al., 2000). A very promising technique is the *RNA arbitrarily primed PCR* (RAP-PCR) method based on the use of nonstoichiometric reduced complexity probes for hybridization to cDNA array membranes (Trenkle et al., 1998). RAP-PCR is an unbiased fingerprinting PCR that samples a reproducible subset of message population based on the best matches with arbitrary primers (Welsh et al., 1992). This technique is capable for the construction of a probe with reduced complexity, which increases the representativity of rare messages starting from small amounts of total RNA or mRNA (Grau et al., 2005).

Aneja et al. (2004) utilized the RAP-PCR technique to investigate the metabolic profiles of the microbial community involved in leaf litter degradation without using selective primer systems. Both mRNA and rRNA transcripts have been successfully targeted in the described protocol, which has the advantage of simultaneous analysis of both active taxa (rRNA) and expression of functional genes (mRNA).

28.2.3.2.5 Rapid Amplification of cDNA Ends

Rapid amplification of cDNA ends (RACE) is a procedure for amplification of nucleic acid sequences from a messenger RNA (mRNA) template located between a defined internal site and unknown sequences at either the 3′ or the 5′ end of the mRNA. This methodology of amplification with single-side specificity has been also described as "one-side" PCR or "anchored" PCR. In general, PCR amplification of a low number of target molecules in a complex mixture requires two sequence-specific primers that flank the region of sequence to be amplified. However, the amplification and characterization of regions of unknown sequences are subjected to severe limitations (Loh et al., 1989), which could be overcome by the 3′- and 5′-RACE methodologies.

The 3′ RACE takes advantage of the natural poly(A) tail in mRNA as a generic priming site for PCR amplification (Park, 2004) whereas 5′ RACE, or "anchored" PCR, is a technique that facilitates the isolation and characterization of 5′ ends from low-copy messages (Schramm et al., 2000).

28.2.3.2.6 Reverse Transcription PCR—Temperature/Denaturing Gradient Gel Electrophoresis

Temperature gradient gel electrophoresis and denaturing gradient gel electrophoresis are molecular fingerprinting methods that separate PCR products of same length based on different sequence compositions resulting in different denaturing characteristics (Muyzer and Smalla, 1998). In the case of soil microbial community studies, these techniques generate banding patterns where each band represents a predominant population of the community. The development of reverse transcriptase PCR (RT-PCR),

transforming RNA in cDNA, empowers the application of T/DGGE techniques to compare the patterns of selected microbial populations (rRNA) with the respective patterns of functional genes through their expression (mRNA–cDNA).

28.2.3.2.7 *The Increase of the RT-PCR Efficiency*

Botero et al. (2005) increased the extraction efficiency of prokaryotic mRNA from environmental samples by coupling poly(A) polymerase with the reverse transcriptase PCR (RT-PCR). Modification of native mRNA molecules with a tail of known oligonucleotide sequences [poly(A)] permits the synthesis of their cDNA. Unexpectedly, no Archaea was detected from poly(A)-modified RNA, whereas additional RT-PCRs performed with universal and Archaea-biased primers and unmodified RNA demonstrated the presence of novel Archaea in the soil. Gao et al. (2007) obtained a sufficient amount of mRNA by reverse transcription with a random hexamer primer with an attached T7 RNA promoter sequence. The synthesized cDNAs were then used as templates for linear RNA amplification with T7 RNA polymerase. Finally, Polidoros et al. (2006) improved inverse-RACE method, which uses CircLigase™ (Epicentre Biotechnologies, Madison, WI) for cDNA circularization, followed by rolling circle amplification (RCA) of the circular cDNA with Φ29 DNA polymerase (New England Biolabs, Ipswich, MA). In this way, a large amount of the PCR template is produced, allowing the simultaneous isolation of the 3′ and 5′ unknown ends of a virtually unlimited number of transcripts after a single reverse transcription reaction. The authors named this method RCA-RACE.

28.2.3.3 Stable-Isotope-Probing Techniques

The application of SIP to nucleic acid studies is among the most promising approaches to link the genetic structure and diversity of autochthonous communities with the physiological status of their representatives or with their own function in situ (Radajewski et al., 2002; Dumont et al., 2006; Bernard et al., 2007; Buckley et al., 2007). The extension of the SIP approach to RNA, which is considered to be a responsive biomarker (Manefield et al., 2002; Lueders et al., 2004), enables the detection of actively metabolizing microorganisms via the expression of the target genes (Radajewski et al., 2003; Ginige et al., 2004; Podar et al., 2007) through the incorporation of stable isotopes' (^{13}C, ^{15}N, ^{32}P, and ^{18}O) rRNA with separation of labeled (heavy) and unlabeled (light) forms by ultracentrifugation in an appropriate density gradient (Radajewski et al., 2003; Schwartz, 2007). The assimilation of the labeled substrate by functional community members can be obtained by adding specific substrates or choosing conditions, under which only certain populations are active (Whiteley et al., 2006). The RNA-SIP method has the potential for wide application in rhizosphere microbial ecology, in particular for tracking carbon flow through different compartments of the rhizosphere microbiota and for assessing the effect of environmental conditions on this flow (Wardle et al., 2004; Lu and Conrad, 2005). The stable ^{13}C-isotope-probing RNA-SIP analysis was capable to highlight an unsuspected diversity of microorganisms living in roots (Vandenkoornhuyse et al., 2007).

The application of combined DNA–RNA-SIP approach showed to be a novel means of linking microbial community function to phylogeny as reported by Bernard et al. (2007) in assessing the dynamics and the diversity of soil bacterial populations actively assimilating ^{13}C derived from labeled plant residues.

A detailed discussion of the use of SIP in the soil–plant system will be present in the specific chapter of this section.

28.2.3.4 Hybridization Techniques

Nucleic acid hybridization techniques are used for the identification of DNA and RNA species with varying degree of homology and for the estimation of relative amounts of nucleic acid with known homology.

28.2.3.4.1 *Northern Blot*

Northern blot involves (1) the use of electrophoresis to separate RNA molecules by size, (2) the transfer of RNA from the electrophoretic gel to the blotting nylon membrane through a capillary or vacuum blotting system, (3) the RNA immobilization on the membrane through covalent linkage promoted by UV light or heat, and (4) the RNA detection with a hybridization probe complementary to a part or to the entire target sequence (Alwine et al., 1977). After hybridization, the membrane is washed to ensure that the probe has bound specifically and to avoid background signals. The hybrid signals are detected by x-ray film and can be quantified by densitometry. The efficiency and the specificity of hybridization can be affected by ionic strength, viscosity, duplex length, mismatched base pairs, and base composition (Engler-Blum et al., 1993). The probes can be labeled either with radioactive isotopes or with chemiluminescence; in this case, alkaline phosphatase (AP) or horseradish peroxidase (HRP) catalyzes the breakdown of chemiluminescent substrates with emission of light (Streit et al., 2008). The northern blot can detect smaller changes in gene expression, and due to its high specificity, it can reduce false-positive results (Engler-Blum et al., 1993).

Concerning soil, Mendum et al. (1998) have tested a quick and reliable technique to extract RNA of a quality suitable for RT-PCR in a fresh sieved clay-rich soil microcosm inoculated with *Rhizobium leguminosarum* CT0370, genetically modified by the insertion in its chromosome of the *uidA* gene codifying for the enzyme L-glucuronidase. The presence of the target transcript, encoding for the L-glucuronidase gene, in the RNA extracted poll was detected by RT-PCR and verified by successive hybridization of the amplified products. This method worked well in the model system; however, the protocol will have to be further optimized to detect the lower levels of mRNA likely to be observed in the field.

Felske et al. (2000) have monitored the composition of predominant soil bacteria during grassland succession at the level of major bacterial taxa coupling multiple competitive RT-PCR with quantitative dot blot hybridization. The dot blot hybridization technique is a simplification of the northern blot, where the mixture containing the target molecule is not firstly separated by gel electrophoresis but directly applied on a membrane as a dot. The monitored predominant ribosome types (16S rRNA)

represented approximately half of all bacterial soil rRNA, which was estimated by dot blot hybridizations of soil rRNA with the bacteria probe EUB338. These ribosome types mainly represented Bacillus and members of the Acidobacterium cluster and the alpha-subclass of Proteobacteria.

28.2.3.4.2 Fluorescent In Situ Hybridization

Fluorescent in situ hybridization permits the detection of a gene of interest through the hybridization of a target gene with a specific fluorescently labeled probe (Bakermans and Madsen, 2002; Ginige et al., 2004) and also the identification of active bacteria within a community or population directly in the environment of interest.

The application of FISH in nutrient-poor environments such as roots and soil is more problematic due to lower cell numbers and the presence of interfering autofluorescence from roots and soil particles. Briones et al. (2002) employed FISH using rRNA-targeted oligonucleotide probes providing, for the first time, population estimates of ammonium oxidizer bacteria (AOB) on the rice root surface. These results were supported by DGGE of the *amoA* gene and analysis of libraries of cloned *amoA*.

A promising application of FISH is its coupling with microautoradiography (MAR) and with stable-isotope probe [MAR–SIP and fluorescent DAPI (4,6-diamino-2-phenyl-indole)] providing information about the bacterial community density, activity, and specific substrate uptake (Ginige et al., 2004).

Compared with other environments, FISH has not frequently been applied in soil and showed quite low detection rates related to total DAPI cell counts (Bouvier and del Giorgio, 2003). It was reported that the detection of FISH-stained cells in soil was mostly affected by autofluorescent soil particles (Hahn et al., 1993; Zarda et al., 1997). This effect could be avoided by the extraction of bacteria with Nycodenz prior to FISH and the use of laser scanning microscopy (Bertaux et al., 2007). Recent applications of FISH in soil were used to analyze shifts in the soil microbial community (Stein et al., 2005) and to visualize microorganisms in undisturbed soil samples (Eickhorst and Tippkötter, 2008a) by its combination with microcalorimetric measurements and micropedological methods, respectively. Eickhorst and Tippkötter (2008b), coupling FISH with catalyzed reporter deposition (CARD), drastically improved the in situ hybridization increasing the fluorescence intensity at the target binding site of the probe by tyramide signal amplification (TSA). As a result, CARD–FISH-stained cells were suitable for automated counting using digital image analysis.

Other FISH-based techniques, like Raman microscopy in combination with FISH and multi-isotope imaging mass spectrometry (MS) (or nano-SIMS), actually in development may permit a marked increase in the detection capacity.

Raman microscopy detects stable-isotope incorporation into cellular constituents, offering a possible advantage over FISH–MAR by obviating the need to handle radioisotopes and facilitating the cumbersome quantification of silver grains. Furthermore and in contrast to FISH–MAR, Raman provides information into which cellular compounds the labeled substrate was incorporated. Recently, FISH and Raman confocal microscopy have been directly combined to confirm the involvement of *Pseudomonas* spp. in the assimilation of $^{13}C_{10}$–naphthalene in groundwater (Huang et al., 2007).

The nano-SIMS technology analyzes the stable- or radioactive-isotope content of single cells at a resolution higher than both Raman microscope and MAR (Kuypers and Jørgensen, 2007). As already achieved for Raman–FISH, a combination of FISH–nano-SIMS would enable phylogenetic and isotopic analysis of an environmental sample in a single scan. Furthermore, it might even become possible to couple isotope array experiments, performed after the addition of stable-isotope-labeled substrates, with nano-SIMS. Such combination would provide high throughput, stable-isotope-based analysis of microbial communities, and their functions. Owing to the high spatial resolution of nano-SIMS, these analyses could possibly be performed in a massively parallel manner with high-density arrays (Lechene et al., 2006).

28.2.3.4.3 Flow Cytometry—Fluorescence-Activated Cell Sorting

Flow cytometry—fluorescence-activated cell sorting (FC-FACS) is a powerful method for the detection and quantification of fluorescence from individual cells within a population. FACS provides data on optical properties of thousands of single cells per second as they pass through a laser beam (Nebe-von-Caron et al., 2000).

The FC-FACS method for in situ analysis and quantification of gene expression at a single-cell level opens up new possibilities for coherent studies of activity and diversity of the soil microbial community.

De Werra et al. (2008) demonstrated the capability and efficiency of a novel FACS-based method to quantify the expression of antifungal genes in *Pseudomonas fluorescens* CHA0 colonizing plant roots.

28.2.3.4.4 RNase Protection Assay

RNase protection assay permits the detection and quantification of specific mRNA transcripts in a complex mixture of total RNA or mRNA molecules (Müller et al., 2001). The RNA probe, complementary to the gene sequence of interest, with incorporated radioactive ($CT^{32}P$) or biotinylated nucleotides (Biotin-N^4-CTP), is synthesized through an in vitro transcription reaction. This labeled probe is incubated with a sample of total RNA or mRNA to facilitate the hybridization reaction. The hybridization mixture is then treated with ribonuclease (RNase) to eliminate not hybridized single-stranded RNA molecules. Usually, the sample is electrophoresed on a denaturing Tris–borate–EDTA (TBE)–urea polyacrylamide gel and detected by methods specific for radiolabeled or biotinylated probes. With respect to the former, the latter probes can be used without gel purification and are more stable (years vs. weeks). Müller et al. (2001) proposed multiprobe RPA sets that are valuable tools for the simultaneous quantitative determination of gene expression.

28.2.3.4.5 Microarrays

DNA microarrays have been used for bacterial detection and to monitor the physiological status and functional activities of microbial community (Ball and Trevors, 2002; Dubois et al.,

2004). In microarrays, the probes, oligonucleotide, cDNA, or small fragments of PCR products that correspond to mRNAs are synthesized prior to deposition on the array surface and then "spotted" onto glass. The resulting "grid" of probes represents the nucleic acid profiles of the prepared probes and is ready to receive complementary cDNA or RNA "targets" derived from experimental samples. By hybridizing cDNA with different fluorophores, the detection of different target cDNAs is possible (Shalon et al., 1996).

The whole-genome microarrays, through target bacteria transcriptional profiles, permit to reveal the extent and complexity of the wild-type *Agrobacterium tumefaciens* adaptation and responses to plant signaling under soil acidic condition identifying several acid-inducible genes, which may be directly involved in agrobacterium–plant interactions (Yuan et al., 2008).

Tiling arrays are a subtype of microarrays frequently utilized to find expressed genes, characterized by the utilization of overlapping probes in very close proximity to obtain an unbiased analysis with higher resolution, permitting the detection of small and rare molecules. Different degrees of resolution depend on the lengths and spacing of the utilized probes.

28.2.3.4.6 *Selective Capture of Transcribed Sequences*

Selective capture of transcribed sequences (SCOTS) permit the direct identification of specifically expressed genes of any microorganism from which genomic DNA can be obtained, without the need of specialized genetic techniques (e.g., DNA microarrays, libraries, species-specific cloning vectors). SCOTS allows the selective capture of bacterial cDNAs from total cDNA using hybridization to specific biotinylated bacterial genomic DNA (Graham and Clark-Curtiss, 1999).

28.2.3.5 Sequencing Technologies

To sequence transcriptomes, various methods have been developed to directly determine cDNA sequences based mostly on traditional (and more expensive) Sanger's sequencing, while others include methodologies such as SAGE and RNA-sequencing (RNA-seq).

28.2.3.5.1 *Serial Analysis of Gene Expression*

Serial analysis of gene expression produces a snapshot of the mRNA population in a sample of interest in the form of small tags that correspond to fragments of those transcripts. The tagged mRNAs are converted to cDNA by RT-PCR and then ligated together into concatemers, after linkers digestion with the appropriate restriction enzymes. Successively to concatemerization, the fragments are ligated into plasmids and are used to transform bacteria; the growth of the transformed bacterial population generates many copies of the plasmid containing the inserts. Those may be sequenced to identify the present mRNA, as well as analyzing expression levels of a given mRNA by counting the present number of copies (Velculescu et al., 1995).

28.2.3.5.2 *RNA-Sequencing*

RNA-sequencing is a revolutionary tool for transcriptomics (Subramanian et al., 2005) based on the use of high-throughput sequencing technologies used to sequence cDNA.

RNA-sequencing can be done with a variety of platforms but here we briefly illustrate only the pyrosequencing 454 technique due to its wide diffusion, and the most promising new techologies, such as Pacific Biosciences, Ion Torrent and Oxford Nanopore.

In pyrosequencing, the cDNAss molecule fragments (max 500 bp) are physically bound to a surface and sequenced in parallel using DNA polymerase and adding one nucleotide species at a time. The number of nucleotides incorporated by the synthesis is quantified by light emitted by the released pyrophosphates. Advantages of the pyrosequencing 454 technique mainly consist in a significant reduction of time and cost of the metagenome approach, permitting to sequence genomes without its isolation by cloning (Marsh, 2007).

Parallel tagged sequencing (PTS) is a molecular barcoding method designed to adapt the recently developed high-throughput 454 parallel sequencing technology to multiple samples (Meyer et al., 2007). The next generation sequencing technology are represented by Pacific Bioscience, Ion Torrent and Oxford Nanopores that darstically decreasing the DNA detection threshold. The Pacific Bioscience maintain the utilization of the fluorescence probes to detect the nucleotides added during the DNA sequencing, wherease the latter two are defined as post-light sequencing technology detecting the nucleotides on the characteristic variation induced on pH values and ionic current flow of the solution, respectively. These latter two techologies should permit a drastic reduction of sequencing machine and reaction costs, increasing also the sequencing speed (Clarke et al. 2009; Metzker 2010; Pennisi 2010).

After cDNA sequencing, the sequence must be aligned. Sequence alignment consists in arranging cDNA to identify regions of similarity. The main alignment methods are based on computational algorithm approaches and could be classified as *global alignment*, where the alignment spans the entire length of all query sequences; *local alignment*, which identifies regions of similarity within long sequences that are overall often widely divergent; and *hybrid methods*, a combination of the global and local alignment methods.

28.2.3.5.3 *Signature-Tagged Mutagenesis*

Recent advances in genome sequencing have permitted to catalog genomes of a large variety of organisms, but the function of their genes is still largely unknown. Signature-tagged mutagenesis (STM) permits to infer what function the product of a particular gene has by disabling it and observing the effect on the organism. The original and most common use of STM is to discover the genes in a pathogen, which are involved in virulence in its host, so that better curative treatments can be designed (Saenz and Dehio, 2005).

28.2.3.6 Bioreporter and Related Reporter Genes

Reporter genes provide a mean of studying and quantifying gene activity in situ (Sayler et al., 2001). Reporter genes are usually fused to the gene of interest for which the activity needs to be quantified. There are various types of reporter genes available, and their selection depends on the sensitivity of the promoter and the type of detection with which the reporter gene can be quantified (Köhler et al., 2000). They can generate colorimetric, fluorescent, luminescent, chemiluminescent, or electrochemical signals, which have the common characteristic to be easily measurable

and proportional to the rate at which the gene is expressed and to the concentration of the inducing substrate. Environmental factors (temperature, pH, moisture, initial substrate concentrations, etc.) can influence the relationship between expression of the gene of interest and luminescence response of the reporter gene (bioluminescence); these factors can significantly affect the quantification of gene expression (Dorn et al., 2003).

The following section briefly outlines some of the available reporter gene systems and their applications.

28.2.3.6.1 Luciferase

Luciferase (*Lux*) is a generic name for an enzyme, found in prokaryota and eukaryota, that catalyzes a light-emitting reaction (bioluminescence). Three variants of *Lux* are available, whose activity differs for the optimal temperature (<30°C, <37°C, and <45°C). Jensen et al. (1998) hypothesized that the luciferase enzyme was stable in bacterial cells inhabiting soil, and subsequently, this could be a limitation to the use of the *luxAB* system as a monitoring tool of the gene of interest. Another limitation to the use of the *Lux* genetic system for studies of soil microbial populations is its energetic demand on cell metabolism.

28.2.3.6.2 Green Fluorescent Protein

Advantages of the green fluorescent protein (GFP) are its easy expression and detection in both prokaryotic and eukaryotic organisms; its use at the single-cell level does not require exogenous substrates or cofactors and requires a lower energetic demand than luminescent markers. Further advantages are its sensitivity, high stability, and resistance to denaturants. Additionally, the ability to alter GFP to produce light emissions besides blue (i.e., cyan, red, and yellow) can allow the use of this technique as a multianalyte detector. Disadvantages include expression variability, lower sensitivity than *luxB* gene, the request of oxygen, the unknown influence of environmental conditions, and the background interference by other fluorescent particles or bacteria (Errampalli et al., 1999; Kohlmeier et al., 2007).

28.2.3.6.3 β-Galactosidase (lacZ Gene)

The *lacZ* gene of *Escherichia coli* encodes the enzyme β-galactosidase, capable of cleaving lactose into glucose and galactose and also the synthetic compound X-gal (5-bromo-4-chloro-3-indolyl-b-d-galactoside) producing a blue pigment (Jensen et al., 1998; Koch et al., 2001; Dorn et al., 2003). The advantages of *lacZ* reporter genes are their efficiency in both prokaryotic and eukaryotic cells, their high turnover rate, their strong signals with different substrates, and the high detection sensitivity of the relative technique (Köhler et al., 2000). The main limitation of this reporter gene is the background level of endogenous β-galactosidase activity within cells, which vary greatly (Lewis et al., 1998), and the need to culture the microorganisms on X-gal medium to detect the expression of the β-galactosidase enzyme.

28.2.3.6.4 Bioreporter Functioning

To be active, the reporter gene needs to be inserted in a host organism, the so-called bioreporter, which is a living cell that has been genetically engineered to produce a measurable signal in response to a specific chemical or physical agent in their environment. Concerning soil, the bioreporter has to be representative of soil-inhabiting microorganisms or to be a specific microorganism for a specific purpose when the strain is constructed. *Pseudomonads* are generally the preferred microorganisms due to their abundance in soil environments (Jensen et al., 1998). Originally developed for fundamental analysis of factors affecting gene expression, bioreporters have been used for the detection of environmental contaminants (Brueckner et al., 2009) and have evolved in fields such as precision agriculture and food safety assurance. Their versatility is due to the high number of reporter gene systems capable to generate a variety of signals (vectors).

To assemble a bioreporter, the promoter gene-trapping technique (PGT) is very efficient. This approach involves the introduction of a promoter gene fused with a reporter gene into the genome, which leads to the isolation and activity detection of the trapped gene of interest (Liu et al., 2006).

Among the PGT are those based on in vitro and in vivo expression technology (IVET) and on differential fluorescence induction (DFI).

The in vitro IVET is an approach to identify gene products that are turned off in one environment (e.g., in laboratory media) but turned on in a different environment (e.g., during infection of a host). This technique has been utilized to establish the genetic determinants important for the life style of target organisms (Ramos-Gonzales et al., 2005).

The PGT techniques based on in vivo IVET and on DFI are mainly utilized to identify and study genes of interest with elevated levels of expression in complex environments. This capacity permits to explore niche-specific gene expression, like identification of genes involved in microbe–plant interactions (Silby and Levy, 2004; Ramos-Gonzales et al., 2005) and also provide useful information about how an organism perceives its environment.

Several variations on the original IVET have emerged, based on different selection strategies such as recombinase selection and system-specific selection. The former is defined as recombinase-based in vivo expression technology (RIVET), founded on the activation of a site-specific DNA resolvase that permits to isolate transiently or weakly expressed genes in a selected environment. The latter one permits to identify genes (promoters) differentially expressed during particular stage of the interaction between a bacterium and its eukaryotic host.

Complementary strategies to PGT-IVET/DFI are signature tagged mutagenesis (STM) (Hensel et al., 1995; Saenz and Dehio, 2005), differential display using arbitrarily primed PCR (McClelland et al., 1995; Fislage, 1998), suppression subtractive hybridization (SSH) (Ito and Sakaki, 1996, 1997), and selective capture of transcribed sequences (SCOTS) (Graham and Clark-Curtiss, 1999).

28.2.3.7 Main Advantages and Disadvantages of the Reported Methods for Gene Expression Studies by Transcriptomic Analysis

Reverse transcription PCR is the most sensitive and flexible of the quantification methods (Wang and Brown, 1999), and it permits to avoid cloning steps and generate reagents, such as full-length

cDNA inserts for cloning (Borson et al., 1992), or arbitrarily primed enhanced sequence tag cDNA libraries (Neto et al., 1997).

Suppression subtractive hybridization is characterized by a simple and efficient strategy to obtain rare cDNAs by the selective digestion of dominant double-stranded cDNA–mRNA hybrids, followed by successive amplification of not annealed fragments, which will grow faster becoming dominant (Rebrikov, 2003).

RNA arbitrarily primed PCR is useful if no prior knowledge is available about functional genes of the microbial community to be analyzed and also permits to increase the array techniques efficiency by reducing the probe complexity and increasing the number of screened differentially expressed genes (10 times higher) respect to those detectable with total cDNA probes (Trenkle et al., 1998; Sakai et al., 2000).

The main drawback of RACE-PCR is the difficulty to bind the unknown RNA fragment to the terminal end of known RNA/cDNAss fragments. Recently, an innovative RACE-PCR based on the utilization of bacteriophage T4 ligase has been proposed (German et al., 2008). Bacteriophage T4 ligase covalently attaches the unknown RNA fragment to the 5′ terminal end of a known mRNA, increasing the efficiency of the successive RT-PCR.

RT-PCR-D/TGGE advantages include simple performance, high sensibility, and reproducibility, whereas the main disadvantage is the PCR dependency, with detection of only dominant target genes.

The SIP technique (only) permits detecting cells that are actively replicating in situ during the pulse with the exclusion of the slow-growing microorganisms that outcompete during enrichment (Manefield et al., 2002, 2006; Radajewski et al., 2003). Other limitations of SIP are the time request by microorganisms for the assimilation of the labeled substrate and the possibility of cross-feeding because labeled products and intermediates of the substrate metabolism can be taken up by nontarget microorganisms (Radajewski et al., 2000, 2003; Whitby et al., 2001; Morris et al., 2002). A further critical step in applying DNA- or RNA-SIP techniques to environmental samples is the need to physically separate labeled from unlabeled molecules using density gradient centrifugation (Friedrich, 2006). The necessity of adding a ^{13}C-labeled substrate to the environmental sample may cause bias by physical disturbance and a change of in situ substrate concentrations (Lu and Conrad, 2005). An advantage of SIP-RNA respect to SIP-DNA is based on fact that RNA synthesis is rapid and its labeling can occur without the need of DNA synthesis or replication of the organism (Manefield et al., 2002; Radajewski et al., 2003).

Concerning the hybridization techniques, the advantages of using northern blotting include the opportunity to verify quality (size) and quantity of RNA on the gel prior to blotting, the use of probes with partial homology, and the storage and reprobing of the membranes for years after blotting (Streit et al., 2008). In the case of RPA, the hybridization of the probe and target mRNA is made directly in solution to maximize the target availability and thus permits detecting rare messages, discriminating between related mRNAs of similar size, which migrate at similar positions on a northern blot, and probing multiple transcripts in one RNA sample. The main limit of the RPA efficiency is the size

of the probes due to the destruction of the nonhybridized RNA during the nuclease digestion step.

With respect to other hybridization methods, microarrays permit visualizing thousands of genes at a time. In comparison, RT-PCR northern blotting has a low sensitivity but it is characterized by a high specificity, which is important to reduce false-positive results (Streit et al., 2008). A disadvantage of microarrays is the lack of standardization in platform fabrication, assay protocols, and analytical methods that makes microarray data difficult to be interpreted, also due to their huge number (especially for tiling array).

The SCOTS technique, unlike other methods for analysis of in vivo gene expression (e.g., IVET, DFI, and STM), permits to identify potentially important genes rather than promoter regions and is not confounded by polar effects when genes are arranged in prokaryotic polycistronic operons. This method overcomes the problems related to very small numbers of target microorganisms, instability of bacterial mRNA, and difficulties involved in separating bacterial mRNA from ribosomal RNA (rRNA) and host RNA. SCOTS is suggested to introduce only a limited bias in gene expression analysis (Graham and Clark-Curtiss, 1999).

The in situ hybridization and SIP techniques are the most complex assays, but they are the only ones capable of localizing transcripts in living cells.

The application of FISH to soil is still problematic for the aspecific effects of soil particles (Nannipieri et al., 2003). The main limit of the FISH method applied to soil studies is due to the autofluorescence of soil particles and the small size of soil bacteria, which can be shown only by *dim fluorescence* (Hahn et al., 1992; Christensen et al., 1999). The problems related to the use of FISH and FISH–MAR in soils and sediments have been discussed by Rogers et al. (2007).

FC-FACS permits to quantify both the bacterial gene expression in soil populations at a single-cell level and the size of a bacterial population without using traditional culture-dependent methods such as plating serial dilutions on selective agar. A consistent problem with the use of flow cytometry in environmental applications is the background noise originating from basal autofluorescence of bacteria or from soil, plant, substrate, and other contaminating particles emitting fluorescence that may interfere with analysis of relevant data (Nebe-von-Caron et al., 2000; De Werra et al., 2008).

Regarding sequencing technologies, SAGE is able to quantify gene expression through a direct count of the number of transcripts. In addition, as the mRNA sequences do not need to be known a priori, unknown genes or gene variants can be discovered. At present, major limitations of SAGE are the high cost and that the gene assignment is limited on the identified cDNA sequences (database) (Lash et al., 2000; Boheler and Stern, 2003; Pleasance et al., 2003).

High-throughput sequencing technologies (454 sequencing), parallelizing the sequencing process, can produce thousands or millions of sequences at once, reducing the cost of DNA sequencing with respect to standard dye-terminator sequencing methods.

The principal characteristics of bioreporter and related reporter genes are the capacity of generating measurable signals

that are proportional to the concentration of the target chemical or physical agent to which are exposed (Brueckner et al., 2009). The main advantages of this technique are to not utilize microbial mRNA, to not require extensive knowledge of the genome of the target microorganisms, and to permit the detection of differences in gene expression occurring within a population inhabiting different environments (e.g., biofilm, rhizosphere, animal host). The main drawbacks of the method are to be unable to isolate repressed promoters, to detect genes that are regulated posttranscriptionally and proteins constitutively expressed but only activated in the wild. Another disadvantage, especially in soil, is represented by the sensitivity of bioreporters to target substrate concentration and availability. Concerning the type of the reporter gene, it has also to be considered that the signal can occur only in the presence of a secondary substrate, which has to be added to the bioassay (*luxAB*, *Luc*, and aequorin), or cannot be self-induced, like *luxCDABE* reporter genes, requiring an external light source (GFP and UMT). Finally, the reporter gene host has to be a representative environmental organism or a specific organism for a specific purpose when it is constructed.

A clear advantage of the IVET technology is that investigations are carried out in the appropriate environment of the target organisms; there is no extrapolation from in vitro model systems. The technique as normally used, however, excludes genes that show significant expression in vitro in minimal medium and identifies only those that are upregulated in the rhizosphere (Silby and Levy, 2004; Ramos-Gonzales et al., 2005).

28.2.4 Gene Expression and Soil Proteomics

28.2.4.1 Protein Extraction Methods

The proteomic approach applied to soil is based on the extraction of proteins from soil. Several extraction protocols for soil proteins have been published (Wright and Upadhyaya, 1996; Rilling et al., 2002; Bolygo et al., 2003; Singleton et al., 2003; Barzaghi et al., 2004; Harner et al., 2004; Ogunseitan and LeBlanc, 2004; Nannipieri, 2006; Ogunseitan, 2006). The extraction strategy varies according to the targeted protein fraction (i.e., prokaryotic/eukaryotic and extracellular/intracellular) and by the subsequent methods of protein analysis (i.e., 2D comparative protein maps or measurement detection of specific polypeptides or enzymatic activities) (Maron et al., 2007). Here, we briefly discuss the indirect soil protein extraction method by Maron et al. (2008) and the direct extraction method by Benndorf et al. (2007). The indirect approach is based on the extraction of microorganisms from soil prior to protein extraction, whereas the direct extraction of soil proteins involves the cell lysis in the soil sample. Both approaches present advantages and disadvantages. The indirect extraction gives purer proteins than the direct extraction but a lower yield and thus a scarce representativity of the proteins codified by soil genes. In addition, proteome can change during the extraction of microorganisms and from soil.

Maron et al. (2008) extracted bacteria from soil suspensions by high-speed centrifugation with a nonionic density gradient created by Nycodenz (Nycomed Pharma, Norway) (Bakken and

Lindahl, 1995). Bacterial cells were lysed by sonication and the protein concentration was determined by the Bradford method (Degens et al., 2000). Colorimetric methods such as the Bradford method have been used to determine the protein concentration of soil extracts in several studies but they can give artifacts as discussed by Nannipieri and Paul (2009).

Benndorf et al. (2007) extracted proteins from soil using a 0.1 M NaOH–phenol solution. The NaOH solution extracts humic acids and proteins from soil minerals by disruption of microbial cells. The subsequent phenol extraction separates the proteins from the humic organic matter. Protein extracts were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and 2D electrophoresis (2DE). Spots and bands were excised from the 2DE gel, solubilized, and hydrolyzed by trypsin for protein characterization. Tryptic peptides were analyzed by MS and the isolated proteins were assessed by database research.

To increase the protein extraction efficiency, an ultrasonication treatment of soil may be applied; this treatment solubilizes the protein fraction immobilized on soil colloids. To avoid the risk of protein denaturation, it was suggested to work at low temperature (Ogunseitan and LeBlanc, 2004). Furthermore, Renella et al. (2002) have proposed the use of a protease inhibitor cocktail to inhibit proteases during cell lysis. The combination of the autoclave treatment and the protein extraction with buffer has been used to extract glomalin, a glycoprotein produced by soil mycorrhizae (Wright and Upadhyaya, 1996, 1998; Rilling et al., 2002). It has been shown that such extractions give humic-like compounds rather than glycoproteins (Schindel et al., 2007).

28.2.4.2 Protein Characterization after Their Extraction and Purification

The expression level of encoding protein genes can be directly assessed by assays similar to those used for quantifying mRNA.

28.2.4.2.1 Hybridization Techniques

28.2.4.2.1.1 Western Blot   Western blot permits to identify proteins of interest using gel electrophoresis to separate native or denatured proteins by the length of the polypeptide (denaturing conditions) or by the 3D structure of the protein (native/nondenaturing conditions), giving also information about their size. The proteins are then transferred to a membrane, typically nitrocellulose or polyvinylidene fluoride (PVDF), where they are detected using antibodies specific to the target protein (Renart et al., 1979; Towbin et al., 1979). The antibody can either be conjugated to a fluorophore or to HRP for imaging and/or quantification.

28.2.4.2.1.2 Enzyme-Linked Immunosorbent Assay   The enzyme-linked immunosorbent assay (ELISA) uses antibodies immobilized on a microtiter plate to capture proteins of interest from samples added to the well. Using a detection antibody conjugated to an enzyme or fluorophore, the quantity of bound proteins can be accurately measured by fluorimetric or colorimetric detection. ELISA can be performed with several modifications of the basic

procedure. The key step, immobilization of the antigen of interest, can be accomplished by direct adsorption to the assay plate or indirectly via a capture antibody that has been attached to the plate. The antigen is then detected either directly (labeled primary antibody) or indirectly (labeled secondary antibody). The most powerful ELISA assay format is the sandwich assay. The analyte to be measured is bound between two primary antibodies, the capture antibody and the detection antibody. The sandwich format is preferentially used because it is sensitive and robust.

The most commonly used enzyme are HRP and AP. Other enzymes have been used as well (β-galactosidase, acetylcholinesterase, and catalase), but they have not gained widespread acceptance because of limited substrate options. The choice of substrate depends upon the required assay sensitivity and the instrumentation available for signal detection (spectrophotometer, fluorometer, or luminometer).

28.2.4.2.1.3 Protein Microarrays

Three types of protein microarrays are currently used in proteomic research: analytical, functional, and reverse-phase microarrays.

The *analytical protein microarrays* are typically used to characterize a complex mixture of proteins according to their binding affinities, specificities, and expression level (Bertone and Snyder, 2005). The *functional protein microarrays* are based on chips of full-length functional proteins or protein domains and are used to study protein–protein, protein–DNA, protein–RNA, protein–phospholipid, and protein–small molecule interactions (Zhu et al., 2001; Hall et al., 2004). In the reverse-phase protein microarray (RPA), the lysate of isolated cells is arrayed onto a nitrocellulose slide and then probed with antibodies against the target protein of interest. The antibodies are detected by chemiluminescent, fluorescent, or colorimetric assays. Reference peptides are printed on the slides for protein quantification of the sample lysates. This technique discriminates between proteins altered by posttranslational modifications, for example, as a result of disease.

Surface plasmon resonance, carbon nanotubes, carbon nanowires, and microelectromechanical system cantilevers techniques are not yet suitable for the detection of high-throughput protein interactions but they seem to be very promising (Ramachandran et al., 2005).

28.2.4.2.2 Gel Electrophoresis

28.2.4.2.2.1 Two-Dimensional Gel Electrophoresis

Two-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (2D-SDS-PAGE) involves the migration and separation of polypeptides in two dimensions according to their isoelectric point and molecular weight, respectively. Recent advances in 2D gel electrophoresis (2DGE) are represented by difference gel electrophoresis (DIGE) that permits to compare different proteins on the same gel by the contemporary utilization of two different protein fluorescent tags (Unlu et al., 1997). Bands of the electrophoretic patterns can be excised and the relative proteins solubilized and analyzed. Usually after solubilization proteins are hydrolyzed by trypsin and the tryptic peptides are analyzed by MS as discussed before. Another interesting technique is the isotope-coded affinity

tag (ICAT) assay, where proteins are treated with the same chemical reagent differing in the mass values due to the different isotopic composition in the linker region. This technique permits to quantify the level of the protein expression (Graves and Haystead, 2002).

28.2.4.2.3 Mass Spectrometry

MS is based on the ionization of chemical compounds to generate charged molecules or molecule fragments, followed by the measurement of their mass to charge ratios. The type of ionization depends on the sample to be analyzed by MS. In the case of liquid and solid biological samples, two soft laser desorption techniques can be used: the matrix-assisted laser desorption/ionization (MALDI) and the surface-enhanced laser desorption/ionization (SELDI). Both MALDI and SELDI are also used for the fragmentation of proteins and peptides. Usually, proteins are digested by trypsin before MS analysis (Nannipieri and Paul, 2009).

28.2.4.2.4 Two-Dimensional Gel Electrophoresis and Liquid Chromatography Combined with MS and MS/MS

In order to increase our protein identification ability especially in difficult environment, such as soil with the presence of unknown sequences species and complex mixture, it has been proposed to combine the separation power of 2DGE and liquid chromatography (LC) with MS or MS/MS peptide information based on molecular mass (MS) and both molecular mass and fragmentation (MS/MS).

Challenges for the use of 2DGE–MS include limitations in the molecular sizes, pI ranges, and hydrophobicities of the proteins that can be analyzed. Furthermore, even with automated techniques, the ability to identify more than a few hundred proteins from a gel plate containing a thousand spots is still very difficult and generally low throughput (Keller and Hettich, 2009).

Benndorf et al. (2003) used a combination of SDS-PAGE and LC–ESI–MS to analyze proteins extracted from contaminated soil and groundwater. To assess the suitability of this approach for the functional analysis of environmental metaproteomes, it was applied to soil that had been enriched in chlorophenoxy acid-degrading bacteria by incubation with 2,4-dichlorophenoxyacetic acid for 22 days. The identification of enzymes such as chlorocatechol dioxygenases was consistent with bacterial metabolic pathways expected to be expressed in these samples.

28.2.4.2.5 Proteogenomics

The ability of the proteomics approaches integrating online high-performance LC-MS/MS coupled with a computational matching algorithms to compare the experimental fragmentation spectrum with the predicted fragmentation spectrum of the putative peptide sequence suggested by the genome information can be used to verify coding regions of a genomic sequence, to validate predicted genes at the level of translation, to detect novel genes, and also to confirm translation of hypothetical proteins. These opportunities are named proteogenomics (Banfield et al., 2005; Ansong et al., 2008).

Ram et al. (2005) reported the efficiency of LC-MS/MS-based proteomics to metagenomics to evaluate gene expression, identify key activities, and examine partitioning of metabolic functions in a natural acid mine drainage (AMD) microbial biofilm community.

Ansong et al. (2008) have highlighted several ways in which high-throughput tandem MS-based proteomics can improve the quality of genome annotations, with emphasis on microbial genomes, concluding however that further improvements are still needed to make this search routine less problematic and more widely applicable. Gilbert et al. (2008) in their studies on the metatranscriptome of a complex marine microbial communities provides evidence that metagenomic and metatranscriptomic combined approache could, offer an unprecedented opportunity to obtain a functional metabolic profile predicting the expressed proteins by translate the sequence informations in peptides that are compared against know databases.

28.2.4.3 Advantages and Disadvantages of the Methods Used in Proteomic Analysis

As discussed before, proteomic analysis can be done by different methods such as western blot, ELISA, 2DGE, MS, in situ hybridization, and protein microarrays. The advantages and disadvantages of these methods are summarized in the following.

The gel-based nature of the western blot assay makes the quantification less accurate but it has the advantage of being able to identify later modifications (proteolysis or ubiquitination) of the protein due to changes in molecular size.

The ELISA detection process is similar to the western blot, but, since there are no gel steps, a more accurate quantification can be achieved. Although the ELISA *direct* detection involves fewer steps than western blotting, it is expensive and time-consuming due to the need of a specific labeling antibody for each protein with a significant reduction of flexibility. The ELISA *indirect* detection assay can use a wide variety of labeled secondary antibodies (available commercially), leading to higher flexibility and higher detection capacity than the ELISA *direct* detection assay. The occurrence of cross reactivity with secondary antibodies, producing nonspecific signals, and the use of complex protocol represent the main drawbacks.

The 2D-SDS-PAGE is expensive, time-consuming, and has a limited resolution capacity that does not permit the detection of all proteins in a complex mixture, particularly those present in low quantities (Graves and Haystead, 2002).

The MS-based proteomics is widely used in studies of proteome or subproteome of cells, whereas its use in environmental research is rare because of the assumption that free proteins are not sufficiently stable in the environment. MS is expensive; trained technicians are required for its functioning and bioinformatic experts are needed to determine the type of protein (Han et al., 2008).

Protein microarrays are commonly used to contemporarily determine several proteins, while western blotting is usually monitoring a small number of proteins. As reported above, the drawback of microarrays applied to transcriptomic analysis is the lack of standardization of platform fabrication, assay protocols, and analysis methods, making the comparison and interpretation of microarray data difficult; it is an interoperable problem in bioinformatics. However, constructing a proteome microarray not only allows an overexpression library, but also provides a high-throughput expression and purification protocol producing thousands of pure and functional proteins that can be utilized in microarray analysis (Hall et al., 2007).

Concerning proteogenomics, which should permit to identify both the gene and its complement proteins from the dominant organisms, there are technical obstacles currently restricting this type of approach to soil. The main drawback is the low DNA/protein extraction efficiency that do not permit to capture the full range of genetic potential of the soil microbial communities. Furthermore, many of the identified proteins have unknown function and peptides that differ from those predicted from gene sequences can be measured but they generally cannot be identified by database matching, even if the difference is only a single amino acid residue. Thus, it is necessary to develop bioinformatics approaches to generate relatively comprehensive gene inventories for each organism type, which are critical for expression and functional analyses (Banfield et al., 2005).

28.2.5 Environmental Factors Affecting Gene Expression in Soil

Numerous environmental factors can activate or repress gene expression, thus affecting microbial activity, richness and evenness (Daubaras and Chakrabarty, 1992) of microorganisms inhabiting complex environments like soil, characterized by a variety of fluctuating conditions in physical and chemical properties such as texture, structure, nutrient levels, temperature, water content, pH, and oxygen status (Graeme et al., 2003; Silby and Levy, 2004; Hamamura et al., 2008). It is therefore important to study the environmental conditions influencing the growth of microorganisms and their ability to express activities under in situ conditions (Fleming et al., 1993; Borneman and Triplett, 1997; Jensen et al., 1998).

28.2.5.1 Soil Sampling

Microhabitats and microenvironments resulting from heterogeneity are familiar concepts in microbial ecology, and communities and their activities are determined by local environmental conditions at the submillimeter scale. In practical terms, this has implications for sampling strategies (Grundmann and Gourbiere, 1999). For example, molecular analyses are frequently carried out using $1–10\,g^{-1}$ soil samples, which are often homogenized by pooling subsamples prior to analysis, with little knowledge of the extent to which such samples are representative of soil microbial structure, activity, molecular diversity between treatments, and the extent to which this approach leads to loss of important information.

Graeme et al. (2003) have assessed the effects of scale level (macro and micro) and both sample size (10, 1, and 0.1 g)

and distance between samples on the heterogeneity of soil Crenarchaeota community structure using PCR-DGGE analysis of 16S rRNA- and ribosomal DNA-derived profiles. The results showed a higher heterogeneity for Crenarchaeota DGGE profiles of $1\,g^{-1}$ samples and $0.1\,g^{-1}$ samples, excluding any relationship between the archaeal profiles and the distance between 1 and $0.1\,g^{-1}$ samples; although relationships between community structure and distance of separation may occur at a microscale level, attention has to be paid in the interpretation of molecular data in terms of the representativity of the microbial community structure in the soil environment.

28.2.5.1.1 Soil pH

Although the soil pH has numerous effects on physiological, morphological, and metabolic responses of soil microorganisms, as well as the availability of mineral nutrients, only few studies have examined the effects of soil pH on gene expression and most of them were conducted in pure cultures of microorganisms isolated from soil (Richardson et al., 1988; Soto et al., 2004). The influence of soil pH on the diversity, abundance, and transcriptional activity of ammonia-oxidizing archaea and bacteria in acid soil was studied by Graeme et al. (2008), evidencing the effect of pH through the selection of different bacterial and archaea ammonia-oxidizer phylotypes with consequences for nitrification in acid soils. Čuhel et al. (2010) studying the effects of pH on grassland soil denitrifier community have indirectly derived their size and activity through the correlation of $N_2O/(N_2O + N_2)$ fluxes ratio and the *narG*, *napA*, *nirS*, *nirK*, and *nosZ* denitrification genes amount. Results have pointed out how the N_2 emissions were mostly responsible for the differences in the total N fluxes between the pH treatments, as the N_2O emissions were not significantly different.

28.2.5.1.2 Soil Moisture and Temperature

Soil microorganisms are subjected to large variations in water availability; these variations occur at various scales, such as seasonal fluctuations in soil structure, which then impact water availability at the microhabitat level (Torsvik and Øvreas, 2002). The soil temperature varies in response to physical processes occurring at the soil surface. It is thought that fluctuations in temperature can have a profound effect on microbial gene expression in soil. However, studies monitoring effects of water availability and temperature on gene expression in soil are scarce (Axtell and Beattie, 2002). Wertz et al. (2009) have evidenced the weak effects of seasonal climate changes on total and active *nirK* denitrifying community variations in a potato crop field.

28.2.5.1.3 Oxygen Concentrations

Oxygen was found to limit the expression of genes codifying enzymes involved in denitrification and to inhibit the process itself (Baumann et al., 1996; Metz et al., 2003). Certain proteins are expressed at higher activity levels under anaerobic than under aerobic conditions (Lynch and Lin, 1996). However, studies about the effect of oxygen on gene expression in soil are scarce.

28.2.5.2 Nutrient Status

28.2.5.2.1 Carbon

Even though carbon is essential for microbial growth and survival, few studies have examined the effects of carbon substrates on the expression of microbial genes in soil. Van Overbeek et al. (1997) and Koch et al. (2001) studied the carbon limitation effects on the β-galactosidase activity of a biosensor (*P. fluorescens*) with *lacZ* reporter system. The biosensor responded to carbon limitation by increasing the β-galactosidase activity but the response was more variable in soil than in pure culture.

28.2.5.2.2 Nitrogen

The amount and form of soil nitrogen can strongly influence the expression of genes codifying proteins involved in the N cycle. Magasanik (1996) found an increment of activated *glnN* gene in microorganisms, encoding the enzyme glutamine synthetase (GS) and other nitrogen regulatory genes in response to a decrease in ammonium availability. The activation of these genes increased the ability of microorganisms to use small amounts of available ammonium by synthesizing several enzymes (GS and others) and permeases catalyzing N reactions and uptake of N compounds, respectively. The activation of N-regulatory genes in the absence of easily assimilated nitrogen has also been observed by Jensen et al. (1998), who measured the expression of an ammonium-starvation-inducible gene in *P. fluorescens* using promoterless *luxAB* genes. Studies on enzyme activity coupled with gene expression have also shown that nitrogen fixation is induced when nitrogen sources are not available (Bürgmann et al., 2003).

28.2.5.2.3 Phosphorus

Soluble phosphate is often present at levels lower than other essential macronutrients, and consequently, it is often a limiting factor for soil microorganisms (Paul and Clark, 1996) that usually overexpress genes codifying proteins of phosphorus uptake involved in the phosphate-specific transport systems, those codifying phosphatases and those codifying compounds involved in the dissolution of insoluble phosphate minerals (Al-Niemi et al., 1997; Welch et al., 2001).

Studies have been conducted with *E. coli*, where about 30 genes involved in phosphate regulation have been sequenced (Wanner, 1993). It is known that when inorganic phosphate is limiting, the proteins PhoB and PhoR are activated and, in turn, act as transcriptional regulators for the Pho regulon, which consists of many genes involved in the transport and assimilation of phosphate (Makino et al., 1988, 1989).

Data about the direct measure of gene expression in relation to phosphate availability are lacking.

28.2.5.3 Pollutants

A number of microorganisms satisfy their nutritional and energy requirements by degrading pollutants not naturally present in soil. Studies have focused on activity measurements of

functional genes codifying proteins involved in pollutant degradation in soil, whereas others focused on evaluating the bioavailability of heavy metals in contaminated soils before, during, and after bioremediation.

Holden et al. (2002) did not find correlations between surface-active gene expression and pollutant degradation in hexadecane-contaminated sand. Devers et al. (2004), investigating the expression of atrazine-degrading genes (*atz*) in *Pseudomonas* sp. ADP and *Chelatobacter heintzii* using real-time qPCR, detected a twofold higher atrazine degradation efficiency of *Pseudomonas* sp. with respect to *Chelatobacter* sp.

Nota et al. (2008) reported the synthesis of antibiotics in *Folsomia candida* as a response to increased cadmium-induced susceptibility to invading pathogens, probably caused by repression of genes involved in the immune system. Jiang et al. (2008) isolated and characterized heavy metal-resistant *Burkholderia* sp. from Pb- and Cd-contaminated soil.

Extraction of mRNA and RT-PCR analysis, as well as the use of microarrays and bioreporters, can be useful for evaluating bioremediation (Devers et al., 2004; Sharkey et al., 2004).

28.2.6 Conclusion and Outlook

There are information gaps about microbial gene expression in soil due to limits of the present techniques, the complexity of the soil environment, and interpretation of data analysis. To better understand the regulation of microbiological activity, we need to link gene expression with the numbers and activity of the corresponding protein.

Unfortunately, the soil proteomics is still in the infancy stage, and we need to overcome several methodological problems in order to monitor microbial synthesis of proteins in soil. Extraction procedures should give both RNA and proteins representative of both pools of active living soil microbial cells (Bürgmann et al., 2003; Ogunseitan, 2006). In addition, both the transcriptome and proteome should not change during the extraction (Nannipieri, 2006). A better understanding of microbial functioning of soil requires to link gene expression with the relative proteome, and this can have positive effects on the understanding of important processes such as pollutant and waste degradation, resistance mechanisms, in situ toxin production, and cell-to-cell communications.

Concerning metabolomics approaches, recently, a very promising new one, based on the volatile organic compounds (VOCs), named volatilomics has been proposed to support proteomics with the intent to better characterize soil microbial communities structure and activity (Insam and Seewald, 2010). Until now, the investigation of soil VOCs is complicated by a large variability in soil chemical and physical parameters (soil type, nutrient, and oxygen availability) influencing their production by altering the physiological state of the microorganisms. However, once the above reported interrelationships will be robustly established, volatilomics could be a very useful tool to characterize soil microbial communities (McNeal and Heribert, 2009; Insam and Seewald, 2010).

Acknowledgment

References

Al-Niemi, S., M.L. Kahn, and T.R. McDermott. 1997. P metabolism in the bean–*Rhizobium tropici* symbiosis. Plant Physiol. 113:1233–1242.

Alwine, J.C., D.J. Kemp, and G.R. Stark. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Soc. USA 74:5350–5354.

Aneja, M.K., S. Sharma, J. Mayer, J.C. Munch, and M. Schloter. 2004. RNA fingerprinting—A new method to screen for differences in plant litter degrading microbial communities. J. Microbiol. Methods 59:223–231.

Ansong, C., S.O. Purvine, J.N. Adkins, M.S. Lipton, and R.D. Smith. 2008. Proteogenomics: Needs and roles to be filled by proteomics in genome annotation. Brief. Funct. Genomic. Proteomic. 7:50–62.

Axtell, C.A., and G.A. Beattie. 2002. Construction and characterization of a proU-gfp transcriptional fusion that measures water availability in a microbial habitat. Appl. Environ. Microbiol. 68:4604–4612.

Bakermans, C., and E.L. Madsen. 2002. Detection in coal tar waste contaminated groundwater of mRNA transcript related to naphthalene dioxygenase by fluorescent in situ hybridization with tyramide signal amplification. J. Microbiol. Methods 50:75–84.

Bakken, L., and A. Frostegård. 2006. Nucleic acids extraction from soil, p. 49–74. *In* P. Nannipieri and K. Smalla (eds.) Nucleic acids and proteins in soil. Soil Biology, Vol. 8. Springer, Berlin, Germany.

Bakken, L.R., and V. Lindahl. 1995. Recovery of bacterial cells from soil, p. 9–27. *In* J.D. van Elsas and J.T. Trevors (eds.) Nucleic acids in the environment: Methods and applications. Springer, Berlin, Germany.

Ball, K.D., and J.T. Trevors. 2002. Bacterial genomics: DNA microarrays and bacterial artificial chromosomes. J. Microbiol. Methods 49:275–284.

Banfield, J.F., N.C. Verberkmoes, R.L. Hettich, and M.P. Thelen. 2005. Proteogenomic approaches for the molecular characterization of natural microbial communities. OMICS 9:301–333.

Barzaghi, D., J.D. Isbiter, K.P. Lauer, and T.L. Born. 2004. Use of surface-enhanced laser desorption/ionization time of flight to explore bacterial proteomes. Proteomics 4:2622–2628.

Baumann, B., M. Snozzi, A.J.B. Tehnder, and J.R. van der Meer. 1996. Dynamics of denitrification activity of *Paracoccus denitrificans* in continuous culture during aerobic–anaerobic changes. J. Bacteriol. 178:4367–4374.

Benndorf, D., G.U. Balcke, H. Harms, and M. von Bergen. 2007. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. ISME J. 1:224–234.

Bernard, L., C. Mougel, P.A. Maron, V. Nowak, J. Lévêque, C. Henault, F.Z. Haichar et al. 2007. Dynamics and identification of soil microbial populations actively assimilating carbon from ¹³C-labelled wheat residue as estimated by DNA- and RNA-SIP techniques. Environ. Microbiol. 9:752–764.

Berry, A.E., C. Chiocchini, T. Selby, M. Sosio, and E.M. Wellington. 2003. Isolation of high molecular weight DNA from soil for cloning into BAC vectors. FEMS Microbiol. Lett. 173:15–20.

Bertaux, J., U. Gloger, M. Schmid, A. Hartmann, and S. Scheu. 2007. Routine fluorescence in situ hybridization in soil. J. Microbiol. Methods 69:451–60.

Bertone, P., and M. Snyder. 2005. Advances in functional protein microarray technology. FEBS J. 272:5400–5411.

Boheler, K.R., and M.D. Stern. 2003. The new role of SAGE in gene discovery. Trends Biotechnol. 21:55–57.

Bolygo, E., D. Ricketts, F. Moffatt, M. Jessop, and S. Attenborough. 2003. Determination of antifungal proteins in soil by liquid chromatography. Anal. Bioanal. Chem. 376:701–705.

Borneman, J., and E.W. Triplett. 1997. Rapid and direct method for extraction of RNA from soil. Soil Biol. Biochem. 29:1621–1624.

Borson, N.D., W.L. Salo, and L.R. Drewes. 1992. A lock-docking oligo(dT) primer for 5′ and 3′ RACE PCR. PCR Methods Appl. 2:144–148.

Botero, L.M., S. D'Imperio, M. Burr, T.R. McDermott, M. Young, and D.J. Hassett. 2005. Poly(A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. Appl. Environ. Microbiol. 71:1267–1275.

Bouvier, T., and P.A. del Giorgio. 2003. Factors influencing the detection of bacterial cells using fluorescence in situ hybridization (FISH): A quantitative review of published reports. FEMS Microbiol. Ecol. 44:3–15.

Briones, A.M., S. Okabe, Y. Umemiya, N.-B. Ramsing, W. Reichardt, and H. Okuyama. 2002. Influence of different cultivars on populations of ammonia-oxidizing bacteria in the root environment of rice. Appl. Environ. Microbiol. 68:3067–3075.

Brueckner, F., K.J. Armache, and A. Cheung. 2009. Structure-function studies of the RNA polymerase II elongation complex. Acta Crystallogr., Sect. D: Biol. Crystallogr. 65:112–120.

Buckley, D.H., V. Huangyutitham, S.F. Hsu, and T.A. Nelson. 2007. Stable isotope probing with ¹⁵N₂ reveals novel non-cultivated diazotrophs in soil. Appl. Environ. Microbiol. 73:3196–3204.

Bürgmann, H., F. Widmer, W.V. Sigler, and J. Zeyer. 2003. mRNA extraction and reverse transcription PCR protocol for detection of *nifH* gene expression by *Azotobacter vinelandii* in soil. Appl. Environ. Microbiol. 69:1928–1935.

Celesnik, H., A. deana and J.C. Belasco 2007. Initiation of RNA decay in Eschericia coli by 5′ pyrophosphate removal. Molecular cell 27:79–90.

Christensen, L.J., B. Korsgaard, and P. Bjerregaard. 1999. The effect of 4-nonylphenol on the synthesis of vitellogenin in the flounder *Platichthys flesus*. Aquat. Toxicol. 46:211–219.

Clarke, J., Hai-Chen Wu, L. Jayasinghe, A. Patel, S. Heid, H. Bayley. 2009. Continuous base identification for single molecule nanopore DNA sequencing. Nature Nanotechnology 4:265–270.

Čuhel, J., M. Šimek, R.J. Laughlin, D. Bru, D. Chèneby, C.J. Watson, and L. Philippot. 2010. Insights into the effect of soil pH on N₂O and N₂ emissions and denitrifier community size and activity. Appl. Environ. Microbiol. 76:1870–1878.

Daubaras, D., and A.M. Chakrabarty. 1992. The environment, microbes and bioremediation: Microbial activities modulated by the environment. Biodegradation 3:125–135.

De Werra, P., E. Baehler, A. Huser, C. Keel, and M. Maurhofer. 2008. Detection of plant-modulated alterations in antifungal gene expression in *Pseudomonas fluorescens* CHA0 on roots. Appl. Environ. Microbiol. 74:1339–1349.

Degens, B.P., L.A. Schipper, G.P. Sparling, and M. Vojvodic-Vukovic. 2000. Decreases in organic C reserves in soils can reduce the catabolic diversity of soil microbial communities. Soil Biol. Biochem. 32:189–196.

Devers, M., G. Soulas, and F. Martin-Laurent. 2004. Real-time reverse transcription PCR analysis of expression of atrazine catabolism genes in two bacterial strains isolated from soil. J. Microbiol. Methods 56:3–15.

Dorn, J.G., R.J. Frye, and R.M. Maier. 2003. Effect of temperature, pH, and initial cell number on *luxABCDE* and *nah* gene expression during naphthalene and salicylate catabolism in the bioreporter organism *Pseudomonas putida* RB1353. Appl. Environ. Microbiol. 69:2209–2216.

Dubois, J.W., S. Hill, L.S. England, T. Edge, L. Masson, J.T. Trevors, and R. Brousseau. 2004. A DNA microarray-based assay for the characterization of commercially formulated microbial products. J. Microbiol. Methods 58:251–262.

Dumont, M.G., J.D. Neufeld, and J.C. Murell. 2006. Isotopes as tools for microbial ecologists. Curr. Opin. Biotechnol. 17:57–58.

Eickhorst, T., and R. Tippkötter. 2008a. Detection of microorganisms in undisturbed soil by combining fluorescence *in situ* hybridization (FISH) and micropedological methods. Soil Biol. Biochem. 40:1284–1293.

Eickhorst, T., and R. Tippkötter. 2008b. Improved detection of soil microorganisms using fluorescence in situ hybridization (FISH) and catalyzed reporter deposition (CARD–FISH). Soil Biol. Biochem. 40:1883–1891.

England, L.S., S.B. Holmes, and J.T. Trevors. 2001. Extraction and detection of baculoviral DNA from lake water, detritus and forest litter. J. Appl. Microbiol. 90:194–205.

England, L.S., and J.T. Trevors. 2003. The microbial DNA cycle in soil. Biol. Forum 96:249–258.

Engler-Blum, G., M. Meier, J. Frank, and G.A. Muller. 1993. Reduction of background problems in nonradioactive northern and southern blot analysis enables higher sensitivity than ³²P-based hybridizations. Anal. Biochem. 210:235–244.

Errampalli, D., K. Leung, M.B. Cassidy, M. Kostrynska, M. Blears, H. Lee, and J.T. Trevors. 1999. Applications of the green fluorescent protein as a molecular marker in environmental microorganisms. J. Microbiol. Methods 35:187–199.

Felske, A., A. Wolterink, R. Van Lis, W.M. De Vos, and A.D.L. Akkermans. 2000. Response of a soil bacterial community to grassland succession as monitored by 16S rRNA levels of the predominant ribotypes. Appl. Environ. Microbiol. 66:3998–4003.

Fislage, R. 1998. Differential display approach to quantitation of environmental stimuli on bacterial gene expression. Electrophoresis 19:613–616.

Fleming, J.T., J. Sanseverino, and G.S. Sayler. 1993. Quantitative relationship between naphthalene catabolic gene frequency and expression in predicting PAH degradation in soils at town gas manufacturing sites. Environ. Sci. Technol. 27:1068–1074.

Fluit, A.C., M.N. Widjojoatmodjo, A.T. Box, R. Torensma, and J. Verhoef. 1993. Rapid detection of salmonellae in poultry with the magnetic immuno-polymerase chain reaction assay. Appl. Environ. Microbiol. 59:1342–1346.

Frias-Lopez, J., Y. Shi, G.W. Tyson, M.I: Coleman, S.C. Schuster, S.W. Chisholm, E.F. DeLong. 2008. Microbial community gene expression in ocean surface water. PNAS 105: 3805–3810.

Friedrich, M.W. 2006. Stable-isotope probing of DNA: Insights into the function of uncultivated microorganisms from isotopically labeled metagenomes. Curr. Opin. Biotechnol. 17:59–66.

Gao, H., Z.K. Yang, T.J. Gentry, L. Wu, C.W. Schadt, and J. Zhou. 2007. Microarray based analysis of microbial community RNAs by whole community RNA amplification. Appl. Environ. Microbiol. 73:563–571.

German, M., M. Pillay, D.H. Jeong, A. Hetawal, S. Luo, P. Janardhanan, V. Kannan et al. 2008. Global identification of microRNA–target RNA pairs by parallel analysis of RNA ends. Nat. Biotechnol. 26:941–946.

Gertsch, J., M. Güttinger, O. Sticher, and J. Heilmann. 2002. Relative quantification of mRNA levels in Jurkat T cells with RT-real-time-PCR (RT-rt-PCR): New possibilities for the screening of anti-inflammatory and cytotoxic compounds. Pharm. Res. 19:1236–1243.

Gilbert, J.A., D. Field, Y. Huang, R. Edwards, W. Li, P. Glina, I. Joint. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoSone 3(8) e3042:1–13.

Ginige, M.P., P. Hugenholtz, H. Damis, M. Wagner, J. Keller, and L. Blackall. 2004. Use of stable-isotope probing, full cycle rRNA analysis, and fluorescence in situ hybridization–microautoradiography to study a methanol-fed denitrifying microbial community. Appl. Environ. Microbiol. 70:588–596.

Graeme, G.W., L.A. Glover, and J.I. Prosser. 2003. Spatial analysis of archaeal community structure in grassland soil. Appl. Environ. Microbiol. 69:7420–7429.

Graeme, W.N., S. Leininger, C. Schleper, and J.I. Prosser. 2008. The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. Environ. Microbiol. 10:2966–2978.

Graham, J.E., and J.E. Clark-Curtiss. 1999. Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). Proc. Natl. Acad. Soc. USA 96:11554–11559.

Grau, M., X. Sole, A. Obrador, G. Tarafa, E. Vendrell, J. Valls, V. Moreno, M.A. Peinado, and G. Capella. 2005. Validation of RNA arbitrarily primed PCR probes hybridized to glass cDNA microarrays: Application to the analysis of limited samples. Clin. Chem. 51:93–101.

Graves, P.R., and T.A.J. Haystead. 2002. Molecular biologists' guide to proteomics. Microbiol. Mol. Biol. Rev. 66:39–63.

Griffiths, R.I., A.S. Whiteley, A.G. O'Donnell, and M.J. Bailey. 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA and rRNA based microbial community composition. Appl. Environ. Microbiol. 66:5488–5491.

Grundmann, G.L., and F. Gourbiere. 1999. A micro-sampling approach to improve the inventory of bacterial diversity in soil. Appl. Soil Ecol. 13:123–126.

Gruntzig, V., S.C. Nold, J. Zhou, and J.M. Tiedje. 2001. *Pseudomonas stutzeri* nitrite reductase gene abundance in environmental samples measured by real-time PCR. Appl. Environ. Microbiol. 67:760–768.

Hahn, D., R.I. Amann, and J. Zeyer. 1993. Whole-cell hybridization of *Frankia* strains with fluorescence- or dioxigenin-labeled, 16S rRNA-targeted oligonucleotide probes. Appl. Environ. Microbiol. 59:1709–1716.

Hahn, M.E., A. Poland, E. Glover, and J.J. Stegeman. 1992. The Ah receptor in marine animals: Phylogenetic distribution and relationship to cytochrome P4501A inducibility. Mar. Environ. Res. 34:87–92.

Hall, D.A., J. Ptacek, and M. Snyder. 2007. Protein microarray technology. Mech. Ageing Dev. 128:161–167.

Hall, D.A., H. Zhu, X. Zhu, T. Royce, M. Gerstein, and M. Snyder. 2004. Regulation of gene expression by a metabolic enzyme. Science 306:482–484.

Hamamura, N., M. Fukuy, D. Ward, and W. Inskeep. 2008. Assessing soil microbial populations responding to crude-oil amendment at different temperatures using phylogenetic, functional gene (*alkB*) and physiological analyses. Environ. Sci. Technol. 42:7580–7586.

Han, X., A. Aslanian, and J.R. Yates, III. 2008. Mass spectrometry for proteomics. Curr. Opin. Chem. Biol. 12:483–490.

Harner, M.J., P.W. Ramsey, and M.C. Rillig. 2004. Protein accumulation and distribution in floodplain soils and river foam. Ecol. Lett. 7:829–836.

Hensel, M., J.E. Shea, M.D. Jones, E. Dalton, C. Gleeson, and D.W. Holden. 1995. Simultaneous identification of bacterial virulence genes by negative selection. Science 269:400–403.

Hod, Y. 1992. A simplified ribonuclease protection assay. Biotechniques 13:852–854.

Holden, P.A., M.G. LaMontagne, A.K. Bruce, W.G. Miller, and S.E. Lindow. 2002. Assessing the role of *Pseudomonas aeruginosa* surface-active gene expression in hexadecane biodegradation in sand. Appl. Environ. Microbiol. 68:2509–2518.

Hsieh, J.F., and S.T. Chen. 2008. A functional proteomic approach to the identification and characterization of protein composition in wheat leaf. Curr. Proteomics 5:253–266.

Huang, W.E., K. Stoecker, R. Griffiths, L. Newbold, H. Daims, A.S. Whiteley, and M. Wagner. 2007. Raman–FISH: Combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. Environ. Microbiol. 9:1878–1889.

Hurt, R.A., X. Qiu, L. Wu, Y. Roh, A.V. Palumbo, J.M. Tiedje, and J. Zhou. 2001. Simultaneous recovery of RNA and DNA from soils and sediments. Appl. Environ. Microbiol. 67:4495–4503.

Insam, H., and M.S.A. Seewald. 2010. Volatile organic compounds (VOCs) in soils. Biol. Fertil. Soils 46:199–213.

Ito, T., and Y. Sakaki. 1996. Toward genome-wide scanning of gene expression: A functional aspect of the genome project. Essays Biochem. 31:11–21.

Ito, T., and Y. Sakaki. 1997. Fluorescent differential display. Methods Mol. Biol. 85:37–44.

Jensen, L.E., L. Kragelund, and O. Nybroe. 1998. Expression of a nitrogen regulated *lux* gene fusion in *Pseudomonas fluorescens* DF57 studied in pure culture and in soil. FEMS Microbiol. Ecol. 25:23–32.

Jiang, C.Y., X.F. Sheng, M. Qian, and Q.Y. Wang. 2008. Isolation and characterization of a heavy metal-resistant *Burkholderia* sp. from heavy metal-contaminated paddy field soil and its potential in promoting plant growth and heavy metal accumulation in metal-polluted soil. Chemosphere 72:157–164.

Keller, M., and R. Hettich. 2009. Environmental proteomics: A paradigm shift in characterizing microbial activities at the molecular level. Microbiol. Mol. Biol. Rev. 73:62–70.

Kirschner, M. 1999. Intracellular proteolysis. Trends Cell Biol. 9:M42–M45.

Koch, B., J. Worm, L.E. Jensen, O. Højberg, and O. Nybroe. 2001. Carbon limitation induces js-dependent gene expression in *Pseudomonas fluorescens* in soil. Appl. Environ. Microbiol. 67:3363–3370.

Köhler, S., S. Belkin, and R.D. Schmid. 2000. Reporter gene bioassay in environmental analysis. Fresenius J. Anal. Chem. 366:769–779.

Kohlmeier, S., M. Mancuso, R. Tecon, H. Harms, J.R. van der Meer, and M. Wells. 2007. Bioreporters: *gfp* versus *lux* revisited and single-cell response. Biosens. Bioelectron. 22:1578–1585.

Kuypers, M.M.M., and B.B. Jørgensen. 2007. The future of single-cell environmental microbiology. Environ. Microbiol. 9:6–7.

Lantz, O., E. Bonney, F. Griscelli, and Y. Taoufik. 1999. Kinetic quantitative PCR vs. end-point quantitative PCR with internal standard, p. 61–86. *In* B. Kochanowski and U. Reischl (eds.) Quantitative PCR protocols. Humana Press Inc., Tortowa, NJ.

Lash, A.E., C.M. Tolstoshev, and L. Wagner. 2000. SAGE map: A public gene expression resource. Genome Res. 7:1051–1060.

Lerat, S., L.S. England, M.L. Vincent, K.P. Pauls, C.J. Swanton, J.N. Klironomos, and J.T. Trevors. 2005. Real-time polymerase chain reaction (PCR) quantification of the transgenes for Roundup ready corn and Roundup ready soybean in soil samples. J. Agric. Food Chem. 53:1337–1342.

Lewis, J.C., A. Feltus, C.M. Ensor, S. Ramanathan, and S. Daunert. 1998. Applications of reporter genes. Anal. Chem. 70:579–585.

Liu, X.H., J.P. Lu, J.Y. Wang, H. Min, and F.C. Lin. 2006. Promoter trapping in *Magnaporthe grisea*. J. Zhejiang Univ. Sci. B 7:28–33.

Logan, J., K. Edwards, and N. Saunders. 2009. Real-time PCR: Current technology and applications. Applied and Functional Genomics, Health Protection Agency. Caister Academic Press, London, UK.

Loh, E.Y., J.E. Elliott, S. Cwirla, L.L. Lanier, and M.M. Davis. 1989. Polymerase chain reaction with single-sided specificity: Analysis of T cell receptor delta chain. Science 243:217.

Lu, Y., and R. Conrad. 2005. In situ stable isotope probing of methanogenic Archaea in the rice rhizosphere. Science 309:1088–1090.

Lueders, T., M. Manefield, and M.W. Friedrich. 2004. Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. Environ. Microbiol. 6:73–78.

Lynch, A.S., and E.C.C. Lin. 1996. Responses to molecular O_2, p. 1527–1538. *In* F.C. Neidhardt et al. (eds.) *Escherichia coli* and *Salmonella*: Cellular and molecular biology. 2nd edn. American Society of Microbiology, Washington, DC.

Magasanik, B. 1996. Regulation of nitrogen utilization, p. 1345–1356. *In* F.C. Neidhardt et al. (eds.) *Escherichia coli* and *Salmonella*: Cellular and molecular biology. 2nd edn. American Society of Microbiology, Washington, DC.

Makino, K., H. Shinagawa, M. Amemura, T. Kawamoto, M. Yamada, and A. Nakata. 1989. Signal transduction in the phosphate regulon of *Escherichia coli* involves phosphotransfer between PhoR and PhoB proteins. J. Mol. Biol. 210:551–559.

Makino, K., H. Shinagawa, M. Amemura, S. Kimura, and A. Nakata. 1988. Regulation of the phosphate regulon of *Escherichia coli*: Activation of *pstS* transcription by PhoB protein in vitro. J. Mol. Biol. 203:85–95.

Manefield, M., R.I. Griffiths, A. Whiteley, and M. Bailey. 2006. Stable isotope probing: A critique of its role in linking phylogeny and function, p. 205–255. *In* P. Nannipieri and K. Smalla (eds.) Nucleic acids and proteins in soil. Soil Biology, Vol. 8. Springer, Berlin, Germany.

Manefield, M., A.S. Whiteley, R.I. Griffiths, and M.J. Bailey. 2002. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. Appl. Environ. Microbiol. 68:5367–5373.

Maron, P.A., M. Maitre, A. Mercier, D.P. Lejon, V. Nowak, and L. Ranjard. 2008. Protein and DNA fingerprinting of a soil bacterial community inoculated into three different sterile soils. Res. Microbiol. 159:231–236.

Maron, P.A., C. Moguel, S. Siblot, P. Lemanceau, and L. Ranjard. 2007. Protein extraction and fingerprinting optimization of bacterial communities in natural environment. Microb. Ecol. 53:426–434.

Marsh, S. 2007. Pyrosequencing applications. Methods Mol. Biol. 373:15–24.

McClelland, M., F. Mathieu-Daude, and J. Welsh. 1995. RNA fingerprinting and differential display using arbitrarily primed PCR. Trends Genet. 11:242–246.

McGrath, S., J. Dooley, and R. Haylock. 2000. Quantification of toxin gene expression by competitive reverse transcription PCR. Appl. Environ. Microbiol. 66:1423–1428.

McGrath, K., S.R. Thomas-Hall, C.T. Cheng, L. Leo, A. Alexa, S. Schmidt, and P.M. Schenk. 2008. Isolation and analysis of mRNA from environmental microbial communities. J. Microbiol. Methods 75:172–176.

McNeal, K.S., and B.E. Heribert. 2009. Volatile organic compounds as indicators of soil microbial activity and community composition shift. Soil Sci. Soc. Am. J. 73:579–588.

Mendum, T.A., E. Sockett, and P.R. Hirsch. 1998. The detection of Gram-negative bacterial mRNA from soil. FEMS Microbiol. Lett. 164:369–373.

Mettel, C., Y. Kim, P.M. Shresta and W. Liesack. 2010. Extraction of mRNA from soil. Applied Environmental Microbiology 76:5995–6000.

Metz, S., B. Wolfgang, A. Hartmann, and M. Schloter. 2003. Detection methods for the expression of the dissimilatory copper-containing nitrite reductase gene (*DnirK*) in environmental samples. J. Microbiol. Methods 55:41–50.

Metzker, M.L. 2010. Sequencing technologies – the next generation. Nature Reviews/Genetics 11:31–46.

Meyer, M., U. Stenzel, S. Myles, K. Prufer, and M. Hofreiter. 2007. Targeted high-throughput sequencing of tagged nucleic acid samples. Nucleic Acids Res. 35:e97. doi:10.1093/nar/gkm566. Available online: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1976447/pdf/gkm566.pdf

Morgan, J.A.W., C. Winstanley, R.W. Pickup, and J.R. Saunders. 1991. Rapid immunomagnetic capture of *Pseudomonas putida* cells from lake water by using bacterial flagella. Appl. Environ. Microbiol. 57:503–509.

Morris, S.A., S. Radajewski, T.W. Willison, and J.C. Murrell. 2002. Identification of the functionally active methanotroph population in a peat soil microcosm by stable-isotope probing. Appl. Environ. Microbiol. 68:1446–1453.

Müller, K., S. Ehlers, W. Solbach, and T. Laskay. 2001. Novel multi-probe RNase protection assay (RPA) sets for the detection of murine chemokine gene expression. J. Immunol. Methods 249:155–165.

Muyzer, G., and K. Smalla. 1998. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. Antonie van Leeuwenhoek 73:127–141.

Nailis, H., T. Coenye, F. van Nieuwerburgh, D. Deforce, and H.J. Nelis. 2006. Development and evaluation of different normalization strategies for gene expression studies in *Candida albicans* biofilms by real-time PCR. BMC Mol. Biol. 7:25.

Nannipieri, P. 2006. Role of stabilized enzymes in microbial ecology and enzyme extraction from soil with potential applications in soil proteomics, p. 75–94. *In* P. Nannipieri and K. Smalla (eds.) Nucleic acids and proteins in soil. Soil Biology, Vol. 8. Springer, Berlin, Germany.

Nannipieri, P., J. Ascher, M.T. Ceccherini, L. Landi, G. Pietramellara, and G. Renella. 2003. Microbial diversity and soil functions. Eur. J. Soil Sci. 54:655–670.

Nannipieri, P., and E. Paul. 2009. The chemical and functional characterization of soil organic N and its biotic components. Soil Biol. Biochem. 41:2357–2369.

Nebe-von-Caron, G., P.J. Stephens, C.J. Hewitt, J.R. Powell, and R.A. Badley. 2000. Analysis of bacterial function by multi-colour fluorescence flow cytometry and single cell sorting. J. Microbiol. Methods 42:97–114.

Neidhardt, F.C., and H.E. Umbarger. 1996. Chemical composition of *Escherichia coli*, p. 13–16. *In* F.C. Neidhardt et al. (eds.) *Escherichia coli* and *Salmonella*: Cellular and molecular biology. ASM Press, Washington, DC.

Neto, E.D., R. Harrop, R. Correa-Oliveira, R.A. Wilson, S.D. Pena, and A.J. Simpson. 1997. Minilibraries constructed from cDNA generated by arbitrarily primed RT-PCR: An alternative to normalized libraries for the generation of ESTs from nanogram quantities of mRNA. Gene 186:135–142.

Norde, W. 2008. My voyage of discovery to proteins in flatland and beyond. Colloids Surf. B 61:1–9.

Norin, M., and M. Sundström. 2002. Structural proteomics: Developments in structure to function predictions. Trends Biotechnol. 20:79–84.

Nota, B., M.J.T.N. Timmermans, O. Franken, K. Montagne-Wajer, J. Mariën, M.E. Boer, T.E. Boer, B. Ylstra, N.M. Straalen, and D. Roelofs. 2008. Gene expression analysis of collembola in cadmium containing soil. Environ. Sci. Technol. 42:8152–8157.

Ogram, A., W. Sun, F.J. Brockman, and J.K. Frederickson. 1995. Isolation and characterization of RNA from low biomass deep subsurface sediments. Appl. Environ. Microbiol. 61:763–768.

Ogunseitan, O.A. 2006. Soil proteomics: Extraction and analysis of proteins from soils, p. 95–116. *In* P. Nannipieri and K. Smalla (eds.) Nucleic acids and proteins in soil. Soil Biology, Vol. 8. Springer, Berlin, Germany.

Ogunseitan, O.A., and J.F. LeBlanc. 2004. Environmental proteomics: Methods and applications for aquatic ecosystem, p. 1027–1046. *In* G.A. Kowalchuk, F.J.D. Bruijn, I.M. Head, A.D. Akkermans, and J.D. van Elsas (eds.) Molecular microbial ecology manual. 2nd edn. Springer, Berlin, Germany.

Park, D.J. 2004. 3′ RACE LaNe: A simple and rapid fully nested PCR method to determine 3′-terminal cDNA sequence. Biotechniques 36:586–590.

Parker, R.M., and N.M. Barnes. 1999. mRNA detection by *in situ* and northern hybridization. Methods Mol. Biol. 106:247–283.

Paul, E.A., and F.E. Clark. 1996. Soil microbiology and biochemistry. 2nd edn. Academic Press Inc., Toronto, Canada.

Pennisi, E. 2010. Semiconductors inspire new sequencing technologies. Science 327:1190.

Pietramellara, G., J. Ascher, M.T. Ceccherini, and G. Renella. 2002. Soil as a biological system. Ann. Microbiol. 52:119–131.

Pleasance, E.D., M.A. Marra, and S.J.M. Jones. 2003. Assessment of SAGE in transcript identification. Genome Res. 13:1203–1215.

Podar, M., C.B. Abulencia, M. Walcher, D. Hutchison, K. Zengler, and J.A. Garcia. 2007. Targeted access to the genomes of low-abundance organisms in complex microbial communities. Appl. Environ. Microbiol. 73:3205–3214.

Polidoros, A.N., K. Pasentsis, and A.S. Tsaftaris. 2006. Rolling circle amplification—RACE: A method for simultaneous isolation of 5′ and 3′ cDNA ends from amplified cDNA templates. Biotechniques 41:35–42.

Porter, J., J. Robinson, R. Pickup, and C. Edwards. 1998. An evaluation of lectin-mediated magnetic bead cell sorting for the targeted separation of enteric bacteria. J. Appl. Microbiol. 84:722–732.

Radajewski, S., P. Ineson, N.R. Parekh, and J.C. Murrell. 2000. Stable-isotope probing as a tool in microbial ecology. Nature 403:646–649.

Radajewski, S., R. McDonald, and J.C. Murrell. 2003. Stable-isotope probing of nucleic acids: A window to the function of uncultured microorganisms. Curr. Opin. Biotechnol. 14:269–302.

Radajewski, S., G. Webster, D.S. Reay, S.A. Morris, P. Ineson, D.B. Nedwell, J.I. Prosser, and J.C. Murrell. 2002. Identification of active methylotroph populations in an acidic forest soil by stable-isotope probing. Microbiology 148:2331–2342.

Ram, R.J., N.C. Verberkmoes, and M.P. Thelen. 2005. Community proteomics of a natural microbial biofilm. Science 308:1915–1920.

Ramachandran, N., D.N. Larson, P.R. Stark, E. Hainsworth, and J. LaBaer. 2005. Emerging tools for real-time label free detection of interactions on functional protein microarrays. FEBS J. 272:5412–5425.

Ramos-Gonzales, M., M. Campos, and J.L. Ramos. 2005. Analysis of *Pseudomonas putida* KT2440 gene expression in the maize rhizosphere: In vitro expression technology capture and identification of root-activated promoters. J. Bacteriol. 187:4033–4041.

Rebrikov, D.V. 2003. Identification of differential genes by suppression subtractive hybridization, p. 297–325. *In* C.W. Dieffenbach and G.S. Dveksler (eds.) PCR primer: A laboratory manual. 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Renart, J., J. Reiser, and G.R. Stark. 1979. Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: A method for studying antibody specificity and antigen structure. Proc. Natl. Acad. Sci. USA 76:3116–3120.

Renella, G., A.M. Chaudri, and P.C. Brookes. 2002. Fresh additions of heavy metals do not model long-term effects on microbial biomass and activity. Soil Biol. Biochem. 34:121–124.

Richardson, A.E., M.A. Djordjevic, B.G. Rolfe, and R.J. Simpson. 1988. Effects of pH, Ca and Al on the exudation from clover seedlings of compounds that induce the expression of nodulation genes in *Rhizobium trifolli*. Plant Soil 109:37–47.

Rilling, M.C., S.F. Wright, M.R. Shaw, and G.B. Field. 2002. Artificial climate warming positively affects arbuscolar mycorrhizae but decreases soil aggregates water stability in an annual grassland. Oikos 97:52–58.

Rogers, S.W., T.B. Moorman, and S.K. Ong. 2007. Fluorescent *in situ* hybridization and microautoradiography applied to ecophysiology in Soil. Soil Sci. Soc. Am. J. 71:620–631.

Saccomanno, C.F., M. Bordonaro, J.S. Chen, and J.L. Nordstrom. 1992. A faster ribonuclease protection assay. Biotechniques 13:846–850.

Saenz, H.L., and C. Dehio. 2005. Signature-tagged mutagenesis: Technical advances in a negative selection method for virulence gene identification. Curr. Opin. Microbiol. 8:612–619.

Sakai, K., H. Higuchi, K. Matsubara, and K. Kato. 2000. Microarray hybridization with fractionated cDNA: Enhanced identification of differentially expressed genes. Anal. Biochem. 287:32–37.

Saleh-Lakha, S., M. Miller, R.G. Campbell, K. Schneider, P. Elahimanesh, M.M. Hart, and J.T. Trevors. 2005. Microbial gene expression in soil: Methods, applications and challenges. J. Microbiol. Methods 63:1–19.

Sayler, G.S., J.T. Fleming, and D.E. Nivens. 2001. Gene expression monitoring in soil by mRNA analysis and gene *lux* fusions. Curr. Opin. Biotechnol. 12:455–460.

Schindel, F.V., E.J. Mercer, and J.A. Rice. 2007. Chemical characteristics of glomalin-related soil protein (GRSP) extracted from soils of varying organic matter content. Soil Biol. Biochem. 39:320–329.

Schramm, G., I. Bruchaus, and T. Roeder. 2000. A simple and reliable 5′ RACE approach. Nucleic Acids Res. 28:1–4.

Schwartz, E. 2007. Characterization of growing microorganisms in soil through stable isotope probing with $H_2^{18}O$. Appl. Environ. Microbiol. 73:2541–2546.

Sessitsch, A., S. Gyamfi, N. Stralis-Pavese, A. Weilharter, and U. Pfeifer. 2002. RNA isolation from soil for bacterial community and functional analysis: Evaluation of different extraction and soil conservation protocols. J. Microbiol. Methods 51:171–179.

Shalon, D., S.J. Smith, and P.O. Brown. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. 6:639–645.

Sharkey, F.H., I.M. Banat, and R. Marchat. 2004. Detection and quantification of gene expression in environmental bacteriology. Appl. Environ. Microbiol. 70:3795–3806.

Silby, M.W., and S.B. Levy. 2004. Use of *in vivo* expression technology to identify genes important in growth and survival of *Pseudomonas fluorescens* Pf0-1 in soil: Discovery of expressed sequences with novel genetic organization. J. Bacteriol. 186:7411–7419.

Singleton, I., G. Merrington, S. Colvan, and J.S. Delahunty. 2003. The potential of soil protein-based methods to indicate metal contamination. Appl. Soil Ecol. 23:25–32.

Soto, M.J., P. van Dillewijn, F. Martinez-Abarca, J.I. Jimenez-Zurdo, and N. Toro. 2004. Attachment to plant roots and nod gene expression are not affected by pH or calcium in the acid-tolerant alfalfa-nodulating bacteria *Rhizobium* sp. LPU83. FEMS Microbiol. Ecol. 48:71–77.

Spackman, E., and D.L. Suarez. 2008. Type A influenza virus detection and quantitation by real-time RT-PCR. Methods Mol. Biol. 436:19–26.

Stein, S., D. Selesi, R. Schilling, I. Pattis, M. Schmid, and A. Hartmann. 2005. Microbial activity and bacterial composition of H_2-treated soils with net CO_2 fixation. Soil Biol. Biochem. 37:1938–1945.

Stevenson, F.J. 1986. Cycles of soil: Carbon, nitrogen, phosphorus, sulfur and micronutrients. John Wiley & Sons, New York.

Streit, S., C.W. Michalski, M. Erkan, J. Kleef, and H. Friess. 2008. Northern blot analysis for detection of RNA in pancreatic cancer cells and tissues. Nat. Protoc. 4:37–43.

Subramanian, A., P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, and J.P. Mesirov. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102:15545–15550.

Torsvik, V., and L. Øvreas. 2002. Microbial diversity and function in soil: From genes to ecosystems. Curr. Opin. Microbiol. 5:240–245.

Torsvik, V.L., R. Sørheim, and J. Goksoyr. 1996. Total bacterial diversity in soil and sediment communities—A review. J. Ind. Microbiol. 17:170–178.

Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. Proc. Natl. Acad. Sci. USA 76:4350–4354.

Trayhurn, P. 1996. Northern blotting. Proc. Nutr. Soc. 55:583–589.

Trenkle, T., J. Welsh, B. Jung, F. Mathieu-Daude, and M. McClelland. 1998. Non-stoichiometric reduced complexity probes for cDNA arrays. Nucleic Acids Res. 26:3883–3891.

Trevors, J.T. 1996. Nucleic acids in the environment. Curr. Opin. Biotechnol. 7:331–336.

Trevors, J.T., and J.D. van Elsas. 1995. Nucleic acids in the environment: Methods and applications. Springer, Berlin, Germany.

Udvardi, M.K., T. Czechowski, and W.R. Scheible. 2008. Eleven golden rules of quantitative RT-PCR. Plant Cell 20:1736–1737.

Unge, A., R. Tombolini, L. Mblbak, and J.K. Jansson. 1999. Simultaneous monitoring of cell number and metabolic activity of specific bacterial populations with a dual *gfp–luxAB* marker system. Appl. Environ. Microbiol. 65:813–821.

Unlu, M., M.E. Morgan, and J.S. Minden. 1997. Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. Electrophoresis 18:2071–2077.

Utt, E.A., J.P. Brousal, L.C. Kikuta-Oshima, and F.D. Quinn. 1995. The identification of bacterial gene expression differences using mRNA based isothermal subtractive hybridization. Can. J. Microbiol. 41:152–156.

Van Guilder, H.D., K.E. Vrana, and W.M. Freeman. 2008. Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques 44:619–626.

Van Overbeek, L.S., J.D. van Elsas, and J.A. van Veen. 1997. *Pseudomonas fluorescens* Tn5-B20 mutant RA92 responds to carbon limitation in soil. FEMS Microbiol. Ecol. 24:57–71.

Vandenkoornhuyse, P., S. Mahè, P. Ineson, P. Staddon, N. Ostle, J.B. Cliquet, A.J. Francez, A.H. Fitter, and J. Young. 2007. Active root-inhabiting microbes identified by rapid incorporation of plant-derived carbon into RNA. Proc. Natl. Acad. Sci. USA 43:16970–16975.

Velculescu, V.E., L. Zhang, B. Vogelstein, and K.W. Kinzler. 1995. Serial analysis of gene expression. Science 270:484–487.

Vernon, S.D., E.R. Unger, M. Rajeevan, I.M. Dimulescu, R. Nisenbaum, and C.E. Campbell. 2000. Reproducibility of alternative probe synthesis approaches for gene expression profiling with arrays. J. Mol. Diagn. 2:124–127.

Wang, T., and M.J. Brown. 1999. mRNA quantification by real time TaqMan polymerase chain reaction: Validation and comparison with RNase protection. Anal. Biochem. 269:198–201.

Wang, Y., S. Morimoto, N. Ogawa, T. Oomori and T. Fuji. 2009. An improved method to extract RNA from soil with efficient removal of humic acids. Journal of Applied Microbiology 107:1168–1177.

Wanner, B.L. 1993. Gene regulation by phosphate in enteric bacteria. J. Cell Biochem. 51:47–54.

Wardle, D.A., R.D. Bardgett, J.N. Klironomos, H. Setälä, W.H. van der Putten, and D.H. Wall. 2004. Ecological linkages between aboveground and belowground biota. Science 304:1629–1633.

Weis, L., and D. Reinberg. 1992. Transcription by RNA polymerase II: Initiator-directed formation of transcription-competent complexes. FASEB J. 6:3300–3309.

Welch, S.A., A.E. Taunton, and J.F. Banfield. 2001. Effect of microorganisms and microbial metabolites on apatite dissolution. Geomicrobiol. J. 19:343–367.

Welford, S.M., J. Gregg, E. Chen, D. Garrison, P.H. Sorensen, and C.T. Denny. 1998. Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization. Nucleic Acids Res. 26:3059–3065.

Welsh, J., K. Chada, S.S. Dalal, R. Cheng, D. Ralph, and M. McClelland. 1992. Arbitrarily primed PCR fingerprinting of RNA. Nucleic Acids Res. 20:4965–4970.

Wertz, S., C.E. Dandie, C. Goyer, J.T. Trevors, and C.L. Patten. 2009. Diversity of *nirK* denitrifying genes and transcripts in an agricultural soil. Appl. Environ. Microbiol. 75:7365–7377.

Whitby, C.B., G. Hall, R. Pickup, J.R. Saunders, P. Ineson, N.R. Parekh, and A. McCarthy. 2001. ^{13}C incorporation into DNA as a means of identifying the active components of ammonia-oxidizer populations. Lett. Appl. Microbiol. 32:398–401.

Whiteley, A.S., M. Manefield, and T. Lueders. 2006. Unlocking the "microbial black box" using RNA based stable isotope probing technologies. Curr. Opin. Biotechnol. 17:67–71.

Wright, S.E., and A. Upadhyaya. 1996. Extraction of an abundant and unusual protein from soil and comparison with hyphal protein of arbuscular mycorrhizal fungi. Soil Sci. 161:575–586.

Wright, S.E., and A. Upadhyaya. 1998. A survey of soils for aggregate stability and glomalin, a glycoprotein product produced by hyphae of arbuscolar mycorrhizal fungi. Plant Soil 198:97–107.

Yuan, Z.C., P. Liu, P. Saenkham, K. Kerr, and E.W. Nester. 2008. Transcriptome profiling and functional analysis of *Agrobacterium tumefaciens* reveals a general conserved response to acidic conditions (pH 5.5) and a complex acid-mediated signaling involved in agrobacterium–plant interactions. J. Bacteriol. 190:494–507.

Zarda, B., D. Hahn, A. Chatzinotas, W. Schönhuber, A. Neef, R.I. Amann, and J. Zeyer. 1997. Analysis of bacterial community structure in bulk soil by in situ hybridization. Arch. Microbiol. 168:185–192.

Zhu, H., M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan et al. 2001. Global analysis of protein activities using proteome chips. Science 293:2101–2105.

Zimmerman, K., and W. Mannhalter. 1996. Technical aspects of quantitative competitive PCR. Biotechniques 21:268–279.

28.3 Stable-Isotope Probing and Its Application in Soil Microbiology

Yin Chen

Deepak Kumaresan

J. Colin Murrell

28.3.1 Introduction

Soil is such a complex environment and the functions of many microorganisms inhabiting it are difficult to predict. In the last few decades, environmental microbiologists have applied many techniques in order to study the structure and function of these organisms in soil (Adamczyk et al., 2003; Huang et al., 2004, 2007; Wagner, 2004; Lechene et al., 2006; Ottesen et al., 2006; Hashsham et al., 2007; Li et al., 2008). Stable isotopes, particularly nitrogen (N) and carbon (C), have been used for decades to understand the function of microorganisms in soil ecosystems. Labeling methodologies enable the introduction of the compounds of interest, and the fate of these compounds in soil can then be followed. Norman and Werkman (1943) performed the first tracer experiment using labeled ^{15}N fertilizer and demonstrated that only 20% of the soybean residues were

taken up by the plant and the rest were incorporated into soil organic matter. Early contributions of ^{15}N trace studies to our knowledge on nitrogen mineralization and immobilization in soil have been reviewed by Winsor (1958). The use of ^{15}N in experiments revealed that nitrogen added to the soil decreased in availability with time (Legg et al., 1971) and the majority of nitrogen was immobilized as amino acids (Cheng and Kurtz, 1963) and amino sugars (Isiriah and Keeney, 1973). Since then, a number of studies have used ^{15}N to understand nitrogen cycling in soil. This has been reviewed in detail by Nannipieri and Paul (2009). Moreover, a variety of labeled carbon substrates such as CO_2, CH_4, glucose, amino acids, and plant organic matter have now been used to address fundamental questions in soil ecology, particularly to understand the carbon transfer in soil–plant system (Stewart and Metherell, 1999), soil microbial respiration (van Vuuren et al., 2000), and soil food web (Middelburg et al., 2000). The role of carbon isotopes in understanding functional soil ecology has been extensively reviewed by Staddon (2004).

Stable-isotope probing is a method for identifying active microorganisms involved in a specific bioprocess. The method relies on the incorporation of stable isotopes (e.g., ^{13}C, ^{15}N) into DNA, RNA, phospholipid fatty acid (PLFA), or protein (Boschker et al., 1998; Bull et al., 2000; Radajewski et al., 2000; Manefield et al., 2002; Jehmlich et al., 2008). Further separation and identification of those labeled compounds can then be carried out to infer the function of the organisms involved. Since the first publication on DNA-SIP (Radajewski et al., 2000), this technique has been used to study active microorganisms in the soil environment. Here, we summarize recent applications of SIP in the study of soil microorganisms.

28.3.2 DNA/RNA-SIP

DNA and RNA stable-isotope probing are extremely powerful tools in analyzing active microorganisms in a particular process. SIP involves several key steps, including incubations with labeled substrate, DNA/RNA extraction and purification, ultracentrifugation, and fractionation (Figure 28.1). Since the first publication of DNA-SIP by Radajewski et al. (2000) and RNA-SIP by Manefield et al. (2002), the number of publications using SIP has increased considerably. Table 28.5 lists a number of
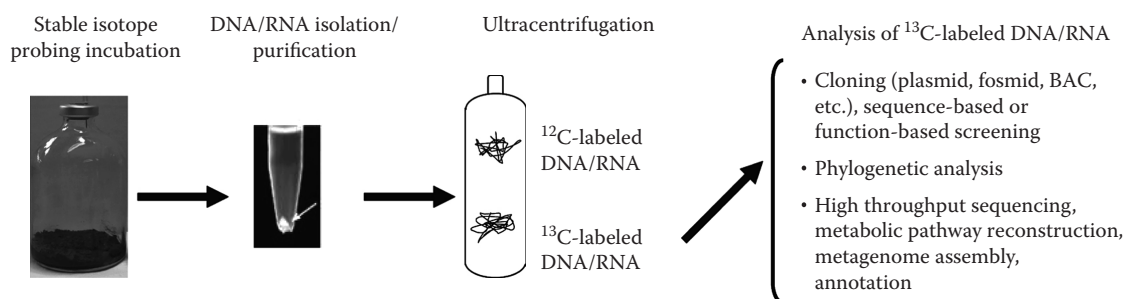
FIGURE 28.1 Flowchart showing major steps in DNA/RNA-SIP analyses. (1) SIP incubations with replicates and ^{12}C-controls; (2) DNA/RNA extraction from incubated samples; (3) separation of labeled and unlabeled DNA/RNA by ultracentrifugation and fractionation and precipitation of DNA/RNA; and (4) downstream analyses of ^{13}C-labeled DNA/RNA.

TABLE 28.5   Recent Studies Using DNA/RNA-SIP to Identify Active Microorganisms in Soils

Substrate	Habitat	Phylogenetic Groups Identified	Marker Genes	Reference
$^{13}CH_4$	Landfill soil with worms	*Methylobacter*; *Methylosarcina*; *Methylocystis*; *Methylomonas*	16S rRNA; *pmoA*; *mmoX*	Héry et al. (2008)
$^{13}CH_4$	Landfill cover soil originally from a peat bog	*Methylobacter*; *Methylomonas*; *Methylocystis*; *Methylocella*	16S rRNA; *pmoA*; *mmoX*	Cébron et al. (2007b)
$^{13}CH_4$	Landfill soil	*Methylobacter*; *Methylomicrobium*; *Methylocystis*	16S rRNA; *pmoA*	Cébron et al. (2007a)
$^{13}CH_4$	Rice field soil	*Methylococcaceae*; *Methylocystaceae*; *Lobosea*; *Heterolobosea*; *Colpodea*; *Cercozoa*	16S rRNA; 18S rRNA	Murase and Frenzel (2007)
$^{13}CH_4$	Forest soil	*Methylocystis*	16S rRNA; *pmoA*	Dumont et al. (2006)
$^{13}CH_4$	Peatland soil	*Methylocystis*; *Methylocella*	16S rRNA	Chen et al. (2008b)
$^{13}CH_4$	Rice field soil	*Methylomicrobium*; *Methylocaldum*	16S rRNA	Noll et al. (2008)
$^{13}CH_4$	Rice field soil	Type I methanotrophs	16S rRNA	Qiu et al. (2008)
$^{13}CH_3Cl$	Soil	*Hyphomicrobium*; *Aminobacter*	*cmuA*	Borodina et al. (2005)
$^{15}N_2$	Soil	*Rhizobiales*; *Methylosinus*; *Methylocystis*	*nifH*	Buckley et al. (2008)
$^{15}N_2$	Soil	*Rhizobiales*; *Actinobacteria*; *Alphaproteobacteria*	16S rRNA; *nifH*	Buckley et al. (2007b)
$^{13}CO_2$	Rice soil	Rice cluster-1 Archaea; *Methanosarcineae*; *Methanomicrobiaceae*; *Methanosaetaceae*	16S rRNA	Lu and Conrad (2005)
$^{13}CO_2$	Upland grassland soil	*Sphingomonas*; *Mycobacterium*; *Sistotrema*; *Rhodotorula*	16S rRNA	Rangel-Castro et al. (2005)
^{13}C-acetate	Rice filed soil	*Geobacter*; *Anaeromyxobacter*	16S rRNA	Hori et al. (2007)
^{13}C-acetate	Soil	*Syntrophus*; *Propionibacterium*; *Geobacter*; *Methanosaeta*; *Methanosarcina*	16S rRNA	Chauhan and Ogram (2006)
^{13}C-xylose; ^{13}C-glucose	Fen soil	*Pseudomonadaceae*; *Aeromonadaceae*; *Acidaminococcaceae*; *Clostridiaceae*; *Enterobacteriaceae*; *Actinomycetales*; *Methanosarcinaceae*; *Methanobacteriaceae*; and *Crenarchaeota*	16S rRNA	Hamberger et al. (2008)
^{13}C-cellulose	Soil	*Dyella*, *Mesorhizobium*; *Sphingomonas*; *Myxobacteria*	16S rRNA	Haichar et al. (2007)
^{13}C-labeled wheat residue	Soil	Beta- and Gamma-proteobacteria	16S rRNA	Bernard et al. (2007)
^{13}C-pyrene	Contaminated soil	Uncultivated Gamma-proteobacteria	16S rRNA	Jones et al. (2008)
$^{13}C_6$-benzene	Coal gasification soil	Deltaproteobacteria; *Clostridia*; *Actinobacteria*	16S rRNA	Kunapuli et al. (2007)
^{13}C-benzoic acid	Agriculture soil	*Burkholderia*	16S rRNA	Pumphrey and Madsen (2008)
^{13}C-polychlorinated biphenyls	Pine tree soil	*Pseudonocardia*; *Kirbella*; *Nocardiodes*; *Sphingomonas*	16S rRNA; ARHD[a]	Leigh et al. (2007)
^{13}C-phenanthrene; ^{13}C-pyrene	Contaminated soil	*Acidovorax*	16S rRNA	Singleton et al. (2007)
^{13}C-labeled 2,4-dichlorophenoxyacetic acid	Agriculture soil	Beta-proteobacteria related to *Ramlibacter* (*Comamonadaceae*)	16S rRNA	Cupples and Sims (2007)
$^{12}C_6$ salicylate; ^{13}C naphthalene phenanthrene	Bioreactor treating contaminated soil	*Acidovorax*; *Pseudomonas*; *Ralstonia*	16S rRNA	Singleton et al. (2005)
^{13}C-labeled naphthalene and glucose	Soil	*Acidovorax*; *Pseudomonas*; *Intrasporangium*	16S rRNA	Yu and Chu (2005)
^{13}C-pentachlorophenol	Pristine grassland soil	*Pseudomonas*; *Burkholderia*; *Sphingomonas*	16S rRNA	Mahmood et al. (2005)
^{13}C-phenol	Agriculture soil	*Kocuria*; *Staphylococcus*; *Pseudomonas*	16S rRNA	DeRito et al. (2005)
^{13}C-phenol	Agriculture soil	*Trichosporon*	18S–28S internal transcribed spacer	DeRito and Madsen (2008)
^{13}C-labeled *E. coli*	Agriculture soil	*Lysobacter*; *Myxococcales*; *Bacteroidetes*; *Microascaceae*	16S rRNA; 18S rRNA	Lueders et al. (2006)

[a] Aromatic ring-hydroxylating dioxygenase.

recent publications using DNA/RNA-SIP technology to analyze a variety of soil microorganisms involved in different processes.

DNA-SIP relies on the incorporation of a substrate that is highly enriched in a stable isotope, such as ^{13}C, and the identification of active organisms by the selective recovery and analysis of isotope-enriched DNA (Radajewski et al., 2000; Dumont and Murrell, 2005; Manefield et al., 2006). DNA-SIP is a culture-independent method; thus, uncultivated organisms can be identified. DNA-SIP also offers direct access to genomic information of potential novel microorganisms involved in a particular process (Dumont et al., 2006; Chen et al., 2008b; Neufeld et al., 2008). One of the drawbacks of DNA-SIP, however, is the relatively long incubation times that are required for DNA replication and incorporation of the ^{13}C-label into newly synthesized DNA (reviewed in Dumont and Murrell, 2005). This long incubation time can result in potential crossfeeding problems, which makes data interpretation more complicated (reviewed by Dumont and Murrell, 2005; Manefield et al., 2006). However, recent studies have employed minimum labeling to shorten incubation time and eliminate crossfeeding of ^{13}C (e.g., Cadisch et al., 2005; Neufeld et al., 2008).

RNA-SIP, on the other hand, offers greater sensitivity because RNA synthesis occurs at a faster rate than DNA synthesis and it is possible to obtain ^{13}C-labeled RNA more quickly than ^{13}C-labeled DNA (Manefield et al., 2002; Lueders et al., 2004). Enhanced sensitivity by RNA-SIP has been demonstrated in a number of studies (Table 28.5). In addition, since only small amounts of isotope need to be incorporated, near in situ concentrations of substrate can be used in RNA-SIP experiments, therefore minimizing potential biases caused by high substrate concentrations during earlier DNA-SIP experiments (Radajewski et al., 2000, 2002). Since its development, RNA-SIP has been applied in a number of studies, especially in the study of plant rhizosphere–microbe interactions using ^{13}CO$_2$ (Table 28.5). This has yielded a wealth of information on rhizosphere bacterial populations that are sequestering carbon through plant exudates, a process that is poorly understood. Compared with DNA-SIP, one of the drawbacks of RNA-SIP is that only 16S rRNA genes can be analyzed since only isotope-labeled RNA is obtained.

Combining RNA-SIP and DNA-SIP is a particularly powerful strategy since these two techniques are complementary. An excellent example was a study by Lueders et al. (2004), in which they simultaneously analyzed the incorporation of the carbon isotope in ^{13}CH$_3$OH in microcosms containing oxic rice field cover soil for 42 days. RNA-SIP allowed the identification of methylotrophs that were active first in the microcosm when only small amounts of ^{13}CH$_3$OH had been metabolized. DNA-SIP, however, shed light on the primary methanol utilizing communities, their enrichment during the incubation period, and their interactions with fungi in the microbial food web.

Previously, only ^{13}C-labeled substrates were used for SIP. More recently, ^{15}N-labeled substrates have also been used (Cadisch et al., 2005; Buckley et al., 2007a, 2007b) to identify those microorganisms metabolizing nitrogen compounds. Compared with ^{13}C-SIP, the separation of DNA bands between ^{15}N-DNA and ^{14}N-DNA was only 4 mm, which is much less than when labeled

with ^{13}C substrates (10 mm). Thus, a second ultracentrifugation was applied to separate ^{15}N-DNA and ^{14}N-DNA (Buckley et al., 2007a, 2007b). The possibility of using ^{18}O for SIP was also studied recently. Incubating *E. coli* cultures or soil samples with H$_2$^{18}O yielded DNA from actively growing cells that was sufficiently labeled to separate ^{18}O-containing DNA from the background (1.5–3.5 mm; Schwartz, 2007).

28.3.3 Combining DNA-SIP and Metagenomics

Metagenomics is a method to study microorganisms without the prerequisite for cultivation (Handelsman et al., 1998) and can be used to retrieve genetic information directly from the environment by either direct large-scale DNA sequencing or by functional screening of relevant bioactive compounds (reviewed in Handelsman, 2005; Lasken, 2007; Schmeisser et al., 2007; and references therein). In combination with metagenomics, DNA-SIP facilitates the selective isolation of DNA from functionally relevant microorganisms to construct metagenomic libraries in a directed fashion that has not been possible previously (Schloss and Handelsman, 2003; Wellington et al., 2003; Dumont and Murrell, 2005; Friedrich, 2006; Neufeld et al., 2007). This goal was first achieved by exposing a soil sample to ^{13}CH$_4$, retrieving high-quality labeled DNA without damaging through UV exposure, and generating a modest bacterial artificial chromosome (BAC) library for the discovery of a BAC clone, which contained a *pmoCAB* operon (Dumont et al., 2006). However, the major concern of combining DNA-SIP with metagenomics has been the challenge of obtaining sufficient ^{13}C-labeled "heavy" DNA for construction of a metagenomic library, without using excessively high concentrations of substrate and reducing the risk of enrichment bias. This has been recently overcome by applying multiple displacement amplification (MDA) to generate large quantities of DNA from minute quantities of ^{13}C-labeled "heavy" DNA (Chen et al., 2008b; Neufeld et al., 2008). In addition, by combining DNA-SIP and high-throughput sequencing technology, Kalyuzhnaya et al. (2008) have managed to reconstruct the genome sequence of a key methylamine degrader from Lake Washington, which had not been previously cultivated.

28.3.4 PLFA-SIP

Lipids found in cellular membrane with other biomarkers such as sterols, hopanoic acids, and ether lipids can be used to differentiate bacteria, fungi, and algae (Boschker and Middelburg, 2002), among which phospholipid fatty acids (PLFAs) have been extensively used as microbial biomarker. The use of stable isotopes such as ^{13}C-CH$_4$ or acetate as substrates and subsequent analysis of the isotope-labeled PLFA revealed the phylogenetic identity of microbes actively utilizing the substrate. Gas chromatography–combustion–isotope ratio mass spectrometry (GC/C/IRMS) can be used to measure the natural variation in isotope ratios of compounds such as ^{13}C and ^{12}C carbon isotopes. GC/C/IRMS can detect minor changes in stable-isotope composition, and traditionally, it has been used to compare the

TABLE 28.6   Recent Studies Using PLFA-SIP to Identify Active Microorganisms in Soils

Substrate	Habitat	Phylogenetic Groups Identified	Reference
$^{13}CH_4$	Rice field soil	Type I methanotrophs	Qiu et al. (2008)
$^{13}CH_4$	Peatland soil	Methylocystis	Chen et al. (2008a)
$^{13}CH_4$	Rice field soil	Type I methanotrophs	Shrestha et al. (2008)
$^{13}CH_4$	Forest soil	Type II methanotrophs	Singh and Tate (2007)
$^{13}CH_4$	Mineral soil	High-affinity methanotrophs	Maxfield et al. (2008)
$^{13}CO_2$	Rice soil	Microorganisms associated with rhizosphere	Lu et al. (2007)
$^{13}CH_4$	Upland brown earth soil	High-affinity methanotrophs	Maxfield et al. (2006)
$^{13}CH_4$	Rice field and forest soil	*Methylocystis; Methylomicrobium; Methylosarcina; Methylobacter;*	Mohanty et al. (2006)

δ13C values (relative abundance of ^{13}C to ^{12}C) of microbial lipids used as biomarkers in geological studies to study ancient environments (Boschker and Middelburg, 2002). The extraction of PLFA from natural samples and the analysis of PLFA by GC/C/IRMS have been reviewed in detail by Evershed et al. (2006) and Boschker and Middelburg (2002).

Boschker et al. (1998) demonstrated the first use of PLFA-SIP to show that type I methanotrophs dominate methane oxidation at a freshwater site using ^{13}C-CH_4 as substrate. Moreover, the authors also revealed that *Desulfotomaculum acetoxidans* was the predominant organism involved in the sulfur cycle in the estuarine and brackish sediments using ^{13}C-acetate. Since then, a large number of labeled substrates (such as bicarbonates, CO_2, toluene, and propionate) have been used to understand the role of microbes in different environments (see review by Evershed et al., 2006).

One of the major contributions of PLFA-SIP is to link function with microbial diversity particularly with methanotrophic bacteria in soil environment (Table 28.6). The presence of distinct PLFAs in methanotrophic bacteria allows us to differentiate between type I (16 carbon fatty acids: 16:0, 16:1) and type II (monosaturated 18 carbon fatty acids: 18:1ω9c, 18:1ω8) methanotrophs and from all other organisms. PLFA-SIP has been successfully used to investigate methanotroph population in sediments (Boschker et al., 1998), soils (Knief et al., 2003; Chen et al., 2008a; Maxfield et al., 2008; Shrestha et al., 2008), microbial mat (Blumenberg et al., 2005), and peat bogs (Raghoebarsing et al., 2005). Owing to the sensitivity of the PLFA-SIP technique (the need for low incorporation of labeled substrate), it has been a preferred tool to study high-affinity methanotrophs, which oxidize methane at atmospheric concentration (2 ppmv) (Bull et al., 2000; Maxfield et al., 2006, 2008). Moreover, PLFA-SIP needs significantly lower amounts of stable-isotope incorporation into microbial biomass than DNA- and RNA-SIP. This eliminates any possibility of crossfeeding, which is one of the potential drawbacks of DNA-SIP.

PLFA-SIP has also been extensively used to study plant–microorganism interactions either in laboratory incubations or in situ studies using ^{13}C-CO_2 pulsed labeling of growing plants (Treonis et al., 2004; Prosser et al., 2006; Williams et al., 2006; Lu et al., 2007). Butler et al. (2003) used ^{13}C-CO_2 pulsing to analyze the incorporation of plant exudates into PLFAs in the bulk and rhizosphere soils of annual ryegrass (*Lolium multiflorum* Lam. var. Gulf) grown in a greenhouse. By analyzing the PLFAs, the authors were able to show differences in the ways in which plant exudates are cycled through microbial communities during different stages of plant development. Prosser et al. (2006) have reviewed the application of stable isotopes in studying plant–microorganism interactions. PLFA-SIP has also been used to link members of the microbial community with degradation of specific xenobiotic compounds such as toluene (Hanson et al., 1999) and styrene (Alexandrino et al., 2001). Alexandrino et al. (2001) used ^2H to identify styrene-degrading organisms in a biofilter.

A key limitation of the PLFA-SIP technique is that even in well-characterized microbial groups such as methanotrophs and sulfate reducers with an extensive PLFA database, it is only possible to detect organisms at the genus level. In addition, recent studies by Dedysh et al. (2007) and Dunfield et al. (2007) have revealed some discrepancies in the PLFAs present in methanotrophs. Dedysh et al. (2007) demonstrated that the type II methanotroph *Methylocystis heyeri* contains 16:1ω8c PLFA, which was previously considered to be a signature PLFA for type I methanotrophs. Therefore, studies have overcome this limitation by combining PLFA-SIP with other complimentary techniques. Chen et al. (2008a) successfully combined PLFA-SIP and mRNA-based analysis to investigate the active methanotroph community structure in peatlands and moorlands under the plant cover of *Sphagnum/Eriophorium* and *Calluna*, respectively. By combining PLFA-SIP with mRNA-based analysis, the authors were able to take the advantage of the sensitivity of the PLFA-SIP while getting a better taxonomic resolution with mRNA-based analysis. Another study by Webster et al. (2006) used PLFA- and DNA-SIP with low concentrations of ^{13}C substrate to provide information on carbon flow and functional diversity in sulfate-reducing marine sediment enrichment slurries.

PLFA-SIP technique provides a powerful culture-independent approach to link function with microbial diversity in different environments and is more sensitive than DNA- and RNA-SIP. The methodology allows us to link microbial diversity with processes at significantly lower concentrations of substrate, which could potentially mimic the substrate concentration in natural environments (such as atmospheric concentrations of methane). However, the PLFA database is rather limited, and, therefore, the taxonomic resolution is not as significant as can be achieved

TABLE 28.7   Comparison of DNA, RNA, PLFA, and Protein-SIP

SIP	Advantages	Disadvantages
DNA-SIP	• Less labor intensive • Instrument requirement is minimal • Whole genomic DNA is available for downstream analyses (e.g., functional gene analyses, genome reconstruction)	• Long incubations required • Potential crossfeeding of labeling • Less sensitive
RNA-SIP	• More sensitive than DNA-SIP • Requires less labeling therefore minimizes crossfeeding	• Only ribosomal RNA gene can be analyzed • Labor intensive • Difficulty in RNA extraction from soil
PLFA-SIP	• Very sensitive • Can be used for quantification of relative bacterial abundance	• Taxonomic assignment of PLFA is troublesome due to the lack of complete database • Labor intensive • Require special instrument
Protein-SIP	• Very sensitive due to the high sensitivity of MS	• Requires metagenome sequences prior to experimental setup • Extremely labor intensive • Requires special instrumentation

with DNA/RNA-based methods. As the use of isotope labeling expands with new substrates and novel environments, PLFA-SIP in combination with other complimentary techniques should continue to provide new insights into microbial ecology, as the PLFA database for microorganisms continues to grow.

28.3.5 Recent Advances and Outlook

There is no doubt that SIP will continue to be an important tool for soil microbiologists to elucidate the function and structure of soil microorganisms. All three SIP techniques have their advantages and disadvantages (Table 28.7) and need to be carefully selected. This largely depends on specific situations. PLFA-SIP is the most sensitive technique and can be used to quantify the activity of certain organisms. However, the lack of a complete database for PLFAs makes it difficult for taxonomic assignment. RNA-SIP, on the other hand, is a powerful tool to reveal the diversity of microorganisms assimilating labeled substrate since labeled RNA can be separated from unlabeled RNA. However, DNA-SIP is desirable in some cases where the whole genomic DNA is needed, although it is the least sensitive technique and requires longer incubation times for sufficient labeling of DNA. One of the main advantages of DNA-SIP is that the labeled DNA can be used for functional gene analyses and for whole-genome/metagenome analyses. This method therefore can considerably link the function of microorganisms to specific processes. The best approach, where possible, will be to combine all these techniques to study the functions of microbes in an environmental sample.

Although DNA/RNA and PLFA-SIP offer links between microbial community and function, the most direct link is proteins since they actually catalyze the biochemical reactions. Very recently, a proof of concept for protein-SIP has been published (Jehmlich et al., 2008). The authors analyzed protein from a single bacterial species (*Aromatoleum aromaticum* strain EbN1, a toluene degrader) within a microbial consortium (the other members of the consortium cannot metabolize toluene but can use gluconate), which were fed with $^{13}C_7$-toluene and ^{12}C-gluconate.

Labeled proteins were separated by two-dimensional gel electrophoresis and MS to follow the ^{13}C carbon atoms into proteins of strain EbN1 and to identify these proteins. The authors demonstrated that incorporation of ^{13}C was exclusively found in proteins from strain EbN1, the toluene degrader, which acquired carbon from ^{13}C-toluene. They therefore demonstrated the suitability of protein-SIP to identify metabolic active species within a mixed culture. Although a powerful technique, protein-SIP requires metagenome sequences prior to the identification of labeled proteins, and this can be time-consuming and expensive. With the continuous progress in high-throughput DNA sequencing technology, this might be affordable in the near future.

In summary, SIP is a useful tool for soil microbiologists and can be used to characterize many functions of soil microorganisms and to reveal the dynamic changes in soil microorganisms in response to changes in soil environments.

References

Adamczyk, J., M. Hesselsoe, N. Iversen, M. Horn, A. Lehner, P.H. Nielsen, M. Schloter, P. Roslev, and M. Wagner. 2003. The isotope array, a new tool that employs substrate-mediated labeling of rRNA for determination of microbial community structure and function. Appl. Environ. Microbiol. 69:6875–6887.

Alexandrino, M., C. Knief, and A. Lipski. 2001. Stable-isotope-based labeling of styrene-degrading microorganisms in biofilters. Appl. Environ. Microbiol. 67:4796–4804.

Bernard, L., C. Mougel, P.A. Maron, V. Nowak, J. Lévêque, C. Henault, F.A. Haichar et al. 2007. Dynamics and identification of soil microbial populations actively assimilating carbon from ^{13}C-labelled wheat residue as estimated by DNA- and RNA-SIP techniques. Environ. Microbiol. 9:752–764.

Blumenberg, M., R. Seifert, K. Nauhaus, T. Pape, and W. Michaelis. 2005. In vitro study of lipid biosynthesis in an anaerobically methane-oxidizing microbial mat. Appl. Environ. Microbiol. 71:4345–4351.

Borodina, E., M.J. Cox, I.R. McDonald, and J.C. Murrell. 2005. Use of DNA-stable isotope probing and functional gene probes to investigate the diversity of methyl chloride-utilizing bacteria in soil. Environ. Microbiol. 7:1318–1328.

Boschker, H.T.S., and J.J. Middelburg. 2002. Stable isotopes and biomarkers in microbial ecology. FEMS Microbiol. Ecol. 40:85–95.

Boschker, H.T.S., S.C. Nold, P. Wellsbury, D. Bos, W. de Graaf, R. Pel, R.J. Parkes, and T.E. Cappenberg. 1998. Direct linking of microbial populations to specific biogeochemical processes by ^{13}C labelling of biomarkers. Nature 392:801–805.

Buckley, D.H., V. Huangyutitham, S. Hsu, and T.A. Nelson. 2007a. Stable isotope probing with ^{15}N achieved by disentangling the effects of genome G + C content and isotope enrichment on DNA density. Appl. Environ. Microbiol. 73:3189–3195.

Buckley, D.H., V. Huangyutitham, S. Hsu, and T.A. Nelson. 2007b. Stable isotope probing with ^{15}N$_2$ reveals novel non-cultivated diazotrophs in soil. Appl. Environ. Microbiol. 73:3196–3204.

Buckley, D.H., V. Huangyutitham, S. Hsu, and T.A. Nelson. 2008. ^{15}N$_2$–DNA–stable isotope probing of diazotrophic methanotrophs in soil. Soil Biol. Biochem. 40:1272–1283.

Bull, I.D., N.R. Parekh, G.H. Hall, P. Ineson, and R.P. Evershed. 2000. Detection and classification of atmospheric methane oxidizing bacteria in soil. Nature 405:175–178.

Butler, J.L., M.A. Williams, P.J. Bottomley, and D.D. Myrold. 2003. Microbial community dynamics associated with rhizosphere carbon flow. Appl. Environ. Microbiol. 69:6793–6800.

Cadisch, G., M. Espana, R. Causey, M. Richter, E. Shaw, J.A.W. Morgan, C. Rahn, and G.D. Bending. 2005. Technical considerations for the use of ^{15}N-DNA stable-isotope probing for functional microbial activity in soils. Rapid Commun. Mass Spectrom. 19:1424–1428.

Cébron, A., L. Bodrossy, Y. Chen, A.C. Singer, I.P. Thompson, J.I. Prosser, and J.C. Murrell. 2007a. Identity of active methanotrophs in landfill cover soil as revealed by DNA-stable isotope probing. FEMS Microbiol. Ecol. 62:12–23.

Cébron, A., L. Bodrossy, N. Stralis-Pavese, A.C. Singer, I.P. Thompson, J.I. Prosser, and J.C. Murrell. 2007b. Nutrient amendments in soil DNA stable isotope probing experiments reduce observed methanotroph diversity. Appl. Environ. Microbiol. 73:789–807.

Chauhan, A., and A. Ogram. 2006. Phylogeny of acetate consuming microorganisms in soils along a nutrient gradient in the Florida Everglades. Appl. Environ. Microbiol. 72:6837–6840.

Chen, Y., M.G. Dumont, N.P. McNamara, P.M. Chamberlain, L. Bodrossy, N. Stralis-Pavese, and J.C. Murrell. 2008a. Diversity of the active methanotrophic community in acidic peatlands as assessed by mRNA and SIP-PLFA analyses. Environ. Microbiol. 10:446–459.

Chen, Y., M.G. Dumont, J.D. Neufeld, L. Bodrossy, N. Stralis-Pavese, N.P. McNamara, N. Ostle, M.J.I. Briones, and J.C. Murrell. 2008b. Revealing the uncultivated majority: Combining stable isotope probing, multiple displacement amplification and metagenomic analysis of uncultivated *Methylocystis* in acidic peatlands. Environ. Microbiol. 10:2609–2622.

Cheng, H.H., and L.T. Kurtz. 1963. Chemical distribution of added nitrogen in soil. Soil Sci. Soc. Am. Proc. 27:312–316.

Cupples, A.M., and G.K. Sims. 2007. Identification of in situ 2,4-dichlorophenoxyacetic acid-degrading soil microorganisms using DNA-stable isotope probing. Soil Biol. Biochem. 39:232–238.

Dedysh, S.N., S.E. Belova, P.L. Bodelier, K.V. Smirnova, V.N. Khmelenina, A. Chidthaisong, Y.A. Trotsenko, W. Liesack, and P.F. Dunfield. 2007. *Methylocystis heyeri* sp. nov., a novel type II methanotrophic bacterium possessing 'signature' fatty acids of type I methanotrophs. Int. J. Syst. Evol. Microbiol. 57:472–479.

DeRito, C.M., and E.L. Madsen. 2008. Stable isotope probing reveals *Trichosporon* yeast to be active *in situ* in soil phenol metabolism. ISME J. 3:477–485.

DeRito, C.M., G.M. Pumphrey, and E.L. Madsen. 2005. Use of field-based stable isotope probing to identify adapted populations and track carbon flow through a phenol-degrading soil microbial community. Appl. Environ. Microbiol. 71:7858–7865.

Dumont, M.G., and J.C. Murrell. 2005. Stable isotope probing—Linking microbial identity to function. Nat. Rev. Microbiol. 3:499–504.

Dumont, M.G., S.M. Radajewski, C.B. Miguez, I.R. McDonald, and J.C. Murrell. 2006. Identification of a complete methane monooxygenase operon from soil by combining stable isotope probing and metagenomics analysis. Environ. Microbiol. 8:1240–1250.

Dunfield, P.F., A. Yuryev, P. Senin, A.V. Smirnova, M.B. Stott, S. Hou, B. Ly et al. 2007. Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. Nature 450:879–882.

Evershed, R.P., Z.M. Crossman, I.D. Bull, H. Mottram, J.A.J. Dungait, P.J. Maxfield, and E.L. Brennand. 2006. ^{13}C-labelling of lipids to investigate microbial communities in the environment. Curr. Opin. Biotechnol. 17:72–82.

Friedrich, M.W. 2006. Stable-isotope probing of DNA: Insights into the function of uncultivated microorganisms from isotopically labeled metagenomes. Curr. Opin. Biotechnol. 17:59–66.

Haichar, F., W. Achouak, R. Christen, T. Heulin, C. Marol, M.F. Marais, C. Mougel, L. Ranjard, J. Balesdent, and O. Berge. 2007. Identification of cellulolytic bacteria in soil by stable isotope probing. Environ. Microbiol. 9:625–634.

Hamberger, A., M.A. Horn, M.G. Dumont, J.C. Murrell, and H.L. Drake. 2008. Anaerobic consumers of monosaccharides in a moderately acidic fen. Appl. Environ. Microbiol. 74:3112–3120.

Handelsman, J. 2005. Metagenomics: Application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. 68:669–685.

Handelsman, J., M.R. Rondon, S.F. Brady, J. Clardy, and R.M. Goodman. 1998. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. Chem. Biol. 5:245–249.

Hanson, J.R., J.L. Macalady, D. Harris, and K.M. Scow. 1999. Linking toluene degradation with specific microbial populations in soil. Appl. Environ. Microbiol. 65:5403–5408.

Hashsham, S.A., E. Gulari, and J.M. Tiedje. 2007. Microfluidic systems being adapted for microbial, molecular biological analyses. Microbe 2:531–536.

Héry, M., A.C. Singer, D. Kumaresan, L. Bodrossy, N. Stralis-Pavese, J.I. Prosser, I.P. Thompson, and J.C. Murrell. 2008. Effect of earthworms on the community structure of active methanotrophic bacteria in a landfill cover soil. ISME J. 2:92–104.

Hori, T., M. Noll, Y. Igarashi, M.W. Friedrich, and R. Conrad. 2007. Identification of acetate-assimilating microorganisms under methanogenic conditions in anoxic rice field soil by comparative stable isotope probing of RNA. Appl. Environ. Microbiol. 73:101–109.

Huang, W.E., R.I. Griffiths, I.P. Thompson, M.J. Bailey, and A.S. Whiteley. 2004. Raman microscopic analysis of single microbial cells. Anal. Chem. 76:4452–4458.

Huang, W.E., K. Stoecker, R. Griffiths, L. Newbold, H. Daims, A.S. Whiteley, and M. Wagner. 2007. Raman–FISH: Combining stable-isotope Raman spectroscopy and fluorescence *in situ* hybridisation for single cell analysis of identity and function. Environ. Microbiol. 9:1878–1889.

Isiriah, N.O., and D.R. Keeney. 1973. Nitrogen transformations in aerobic and waterlogged histosols. Soil Sci. 115:123–129.

Jehmlich, N., F. Schmidt, M. von Bergen, H.H. Richnow, and C. Vogt. 2008. Protein-based stable isotope probing (protein-SIP) reveals active species within anoxic mixed cultures. ISME J. 2:1122–1133.

Jones, M.D., D.R. Singleton, D.P. Carstensen, S.N. Powell, J.S. Swanson, F.K. Pfaender, and M.D. Aitken. 2008. Effect of incubation conditions on the enrichment of pyrene-degrading bacteria identified by stable-isotope probing in an aged, PAH-contaminated soil. Microb. Ecol. 56:341–349.

Kalyuzhnaya, M.G., A. Lapidus, N. Ivanova, A.C. Copeland, A.C. McHardy, E. Szeto, A. Salamov, et al. 2008. High-resolution metagenomics targets specific functional types in complex microbial communities. Nat. Biotechnol. 26:1029–1034.

Knief, C., A. Lipski, and P.F. Dunfield. 2003. Diversity and activity of methanotrophic bacteria in different upland soils. Appl. Environ. Microbiol. 69:6703–6714.

Kunapuli, U., T. Lueders, and R.U. Meckenstock. 2007. The use of stable isotope probing to identify key iron-reducing microorganisms involved in anaerobic benzene degradation. ISME J. 1:643–653.

Lasken, R.S. 2007. Single-cell genomic sequencing using multiple displacement amplification. Curr. Opin. Microbiol. 10:510–516.

Lechene, C., F. Hillion, G. McMahon, D. Benson, A.M. Kleinfeld, J.P. Kampf, D. Distel et al. 2006. High-resolution quantitative imaging of mammalian and bacterial cells using stable isotope mass spectrometry. J. Biol. 5:20.

Legg, J.O., F.W. Chichester, G. Stanford, and W.H. De Mar. 1971. Incorporation of ^{15}N tagged mineral nitrogen into stable forms of soil organic nitrogen. Soil Sci. Soc. Am. Proc. 35:273–276.

Leigh, M.B., V.H. Pellizari, O. Uhlík, R. Sutka, J. Rodrigues, N.E. Ostrom, J. Zhou, and J.M. Tiedje. 2007. Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). ISME J. 1:134–148.

Li, T., T.D. Wu, L. Mazéas, L. Toffin, J.L. Guerquin-Kern, G. Leblon, and T. Bouchez. 2008. Simultaneous analysis of microbial identity and function using nanoSIMS. Environ. Microbiol. 10:580–588.

Lu, Y., W.R. Abraham, and R. Conrad. 2007. Spatial variation of active microbiota in the rice rhizosphere revealed by *in situ* stable isotope probing of phospholipid fatty acids. Environ. Microbiol. 9:474–481.

Lu, Y., and R. Conrad. 2005. *In situ* stable isotope probing of methanogenic archaea in the rice rhizosphere. Science 309:1088–1090.

Lueders, T., R. Kindler, A. Miltner, M.W. Friedrich, and M. Kaestner. 2006. Identification of bacterial micropredators distinctively active in a soil microbial food web. Appl. Environ. Microbiol. 72:5342–5348.

Lueders, T., B. Wagner, P. Claus, and M.W. Friedrich. 2004. Stable isotope probing of rRNA and DNA reveals a dynamic methylotroph community and trophic interactions with fungi and protozoa in oxic rice field soil. Environ. Microbiol. 6:60–72.

Mahmood, S., G.I. Paton, and J.I. Prosser. 2005. Cultivation independent *in situ* molecular analysis of bacteria involved in degradation of pentachlorophenol in soil. Environ. Microbiol. 7:1349–1360.

Manefield, M., R.I. Groffotjs, M.J. Bailey, and A.S. Whiteley. 2006. Stable isotope probing: A critique of its role in linking phylogeny and function, p. 205–216. *In* P. Nannipieri and K. Smalla (eds.) Nucleic acids and proteins in soil. Soil Biology, Vol. 8. Springer, Berlin, Germany.

Manefield, M., A.S. Whiteley, R.I. Griffiths, and M.J. Bailey. 2002. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. Appl. Environ. Microbiol. 68:5367–5373.

Maxfield, P.J., E.R. Hornibrook, and R.O. Evershed. 2006. Estimating high-affinity methanotrophic bacterial biomass, growth, and turnover in soil by phospholipid fatty acid ^{13}C labeling. Appl. Environ. Microbiol. 72:3901–3907.

Maxfield, P.J., E.R. Hornibrook, and R.P. Evershed. 2008. Acute impact of agriculture on high-affinity methanotrophic bacterial populations. Environ. Microbiol. 10:1917–1924.

Middelburg, J.J., C. Barranguet, H.T.S. Boschker, P.M.J. Herman, T. Moens, and C.H.R. Heip. 2000. The fate of intertidal microphytobenthos carbon: An *in situ* ^{13}C-labeling study. Limnol. Oceanogr. 45:1224–1234.

Mohanty, S.R., P.L. Bodelier, V. Floris, and R. Conrad. 2006. Differential effects of nitrogenous fertilizers on methane-consuming microbes in rice field and forest soils. Appl. Environ. Microbiol. 72:1346–1354.

Murase, J., and P. Frenzel. 2007. A methane-driven microbial food web in a wetland rice soil. Environ. Microbiol. 9:3025–3034.

Nannipieri, P., and E.A. Paul. 2009. The chemical and functional characterization of soil N and its biotic components. Soil Biol. Biochem. 41:2357–2369.

Neufeld, J.D., Y. Chen, M.G. Dumont, and J.C. Murrell. 2008. Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. Environ. Microbiol. 10:1526–1535.

Neufeld, J.D., M.G. Dumont, J. Vohra, and J.C. Murrell. 2007. Methodological considerations for the use of stable isotope probing in microbial ecology. Microb. Ecol. 53:435–442.

Noll, M., P. Frenzel, and R. Conrad. 2008. Selective stimulation of type I methanotrophs in a rice paddy soil by urea fertilization revealed by RNA-based stable isotope probing. FEMS Microbiol. Ecol. 65:125–132.

Norman, A.G., and C.H. Werkman. 1943. The use of the nitrogen isotope ^{15}N in determining nitrogen recovery from plant materials decomposing in soil. J. Am. Soc. Agron. 35:1023–1025.

Ottesen, E.A., J.W. Hong, S.R. Quake, and J.R. Leadbetter. 2006. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. Science 314:1464–1467.

Prosser, J.I., J.I. Rangel-Castro, and K. Killham. 2006. Studying plant-microbe interactions using stable isotope technologies. Curr. Opin. Biotechnol. 17:98–102.

Pumphrey, G.M., and E.L. Madsen. 2008. Field-based stable isotope probing reveals the identities of benzoic acid-metabolizing microorganisms and their *in situ* growth in agricultural soil. Appl. Environ. Microbiol. 74:4111–4118.

Qiu, Q., M. Noll, W.R. Abraham, Y. Lu, and R. Conrad. 2008. Applying stable isotope probing of phospholipid fatty acids and rRNA in a Chinese rice field to study activity and composition of the methanotrophic bacterial communities *in situ*. ISME J. 2:602–614.

Radajewski, S., P. Ineson, N.R. Parekh, and J.C. Murrell. 2000. Stable-isotope probing as a tool in microbial ecology. Nature 403:646–649.

Radajewski, S., G. Webster, D.S. Reay, S.A. Morris, P. Ineson, D.B. Nedwell, J.I. Prosser, and J.C. Murrell. 2002. Identification of active methylotroph populations in an acidic forest soil by stable-isotope probing. Microbiology 148:2331–2342.

Raghoebarsing, A.A., A.J. Smolders, M.C. Schmid, W.I. Rijpstra, M. Wolters-Arts, J. Derksen, M.S.M. Jetten et al. 2005. Methanotrophic symbionts provide carbon for photosynthesis in peat bogs. Nature 436:1153–1156.

Rangel-Castro, J.I., K. Kilham, N. Ostle, G.W. Nicol, I.C. Anderson, C.M. Scrimgeour, P. Ineson, A. Meharg, and J.I. Prosser. 2005. Stable isotope probing analysis of the influence of liming on root exudates utilization by soil microorganisms. Environ. Microbiol. 7:828–838.

Schloss, P.D., and J. Handelsman. 2003. Biotechnological prospects from metagenomics. Curr. Opin. Biotechnol. 14:303–310.

Schmeisser, C., H. Steele, and W.R. Streit. 2007. Metagenomics, biotechnology with non-culturable microbes. Appl. Microbiol. Biotechnol. 75:955–962.

Schwartz, E. 2007. Characterization of growing microorganisms in soil through stable isotope probing with H$_2$^{18}O. Appl. Environ. Microbiol. 73:2541–2546.

Shrestha, M., W.R. Abraham, P.M. Shrestha, M. Noll, and R. Conrad. 2008. Activity and composition of methanotrophic bacterial communities in planted rice soil studied by flux measurements, analyses of *pmoA* gene and stable isotope probing of phospholipid fatty acids. Environ. Microbiol. 10:400–412.

Singh, B.K., and K. Tate. 2007. Biochemical and molecular characterization of methanotrophs in soil from a pristine New Zealand beech forest. FEMS Microbiol. Lett. 275:89–97.

Singleton, D.R., M. Hunt, S.N. Powell, R. Frontera-Suau, and M.D. Aitken. 2007. Stable-isotope probing with multiple growth substrates to determine substrate specificity of uncultivated bacteria. J. Microbiol. Methods 69:180–187.

Singleton, D.R., S.N. Powell, R. Sangaiah, A. Gold, L.M. Ball, and M.D. Aitken. 2005. Stable-isotope probing of bacteria capable of degrading salicylate, naphthalene, or phenanthrene in a bioreactor treating contaminated soil. Appl. Environ. Microbiol. 71:1202–1209.

Staddon, P.L. 2004. Carbon isotopes in functional soil ecology. Trends Ecol. Evol. 19:148–154.

Stewart, D.P.C., and A.K. Metherell. 1999. Carbon (^{13}C) uptake and allocation in pasture plants following field pulse-labeling. Plant Soil 210:61–73.

Treonis, A.M., N.J. Ostle, A.W. Stott, R. Primrose, S.J. Grayston, and P. Ineson. 2004. Identification of groups of metabolically-active rhizosphere microorganisms by stable isotope probing of PLFAs. Soil Biol. Biochem. 36:533–537.

van Vuuren, M.M.I., D. Robinson, C.M. Scrimgeour, J.A. Raven, and A.H. Fitter. 2000. Decomposition of ^{13}C-labeled wheat root systems following growth at different CO$_2$ concentrations. Soil Biol. Biochem. 32:403–413.

Wagner, M. 2004. Deciphering the function of uncultured microorganisms. ASM News 70:63–70.

Webster, G., L.C. Watt, J. Rinna, J.C. Fry, R.P. Evershed, R.J. Parkes, and A.J. Weightman. 2006. A comparison of stable-isotope probing of DNA and phospholipid fatty acids to study prokaryotic functional diversity in sulfate-reducing marine sediment enrichment slurries. Environ. Microbiol. 8:1575–1589.

Wellington, E.M.H., A. Berry, and M. Krsek. 2003. Resolving functional diversity in relation to microbial community structure in soil: Exploiting genomics and stable isotope probing. Curr. Opin. Microbiol. 6:295–301.

Williams, M.A., D.D. Myrold, and P.J. Bottomley. 2006. Carbon flow from ^{13}C-labeled straw and root residues into the phospholipid fatty acids of a soil microbial community under field conditions. Soil Biol. Biochem. 38:759–768.

Winsor, G.W. 1958. Mineralization and immobilization of nitrogen in soil. J. Sci. Food Agric. 12:792–801.

Yu, C.P., and K.H. Chu. 2005. A quantitative assay for linking microbial community function and structure of a naphthalene-degrading microbial consortium. Environ. Sci. Technol. 39:9611–9619.

V

Pedology

Larry T. West
United States Department of Agriculture

Larry P. Wilding
Texas A&M University

PEDOLOGY IS THE EARTH SCIENCE that quantifies the factors and processes of soil formation including quality, extent, distribution, and spatial variability of soils from microscopic to megascopic scales (Spositio and Reginato, 1992; Wilding et al., 1994). Spatial distribution and variability of soils in landforms is governed by the processes of soil formation which are, in turn, interactively conditioned by lithology, climate, biology, and relief through geologic time. Soils form a continuous drape across the landscape and are intimately linked to one another; processes and impacts on higher topographic surfaces directly affect adjacent lower lying surfaces. This is because transfer of energy and mass, the driving forces of pedogenesis, occur within, over, and among 3D soil landform components. Renewal vectors of biomass input, rainfall, and dust counter constituent losses via drainage waters, lateral interflow, and downslope migration of erosion products.

The development of open versus closed drainage patterns during landform evolution strongly governs energy and mass flux in the system. In the closed drainage network, dispersal of chemical and erosion products is distributed to adjacent local sinks and depressions. In contrast, dispersal is mostly external to source areas in open drainage systems. In this case, distribution occurs via upland drainage ways to tributary streams, rivers, lakes, and oceans. Differences in drainage network are paramount when one considers the effect of landscape on soil moisture and nutrient regimes, environmental contamination, recharge of groundwater aquifers, crop production potential, etc. Hence, to adequately comprehend, interpret, and transfer knowledge of soil resources from one area to the next, a systematic soil/geomorphic landscape model must be applied.

Through knowledge of soil/landform relationships, pedologists have verified the occurrence, configuration, depth, and pedogenic formation of root and water restrictive layers; documented the origin and distribution of cracking and fissuring patterns in soils and geologic materials that govern bypass flow of nutrients, chemicals solutes, and fluids; identified the scale, mode, and occurrence of systematic soil spatial distribution to aid the design efficiency and sampling of soil units; and used soil color patterns on a macro- and microscale to infer major periods of soil aeration/reduction and relative periods of excess, sufficiency, an deficiency of soil moisture contents for specific land uses. In most soil systems, reaction kinetics and diffusion rates rather than thermodynamic equilibria control chemical reactions, solute movement, and precipitation of chemicals and minerals. These and other impacts of geomorphology and pedogenic processes reflected in soil landscape patterns are covered in Chapters 29 and 30, respectively.

Pedology is an integrative and extrapolative science. It provides an organizational framework to catalog modes, mechanisms, and magnitudes of soil spatial variability, and to generalize this knowledge for synthesis and data population for models. Pedology provides a vehicle for extrapolation and scaling of spatial variability from components of soils (individual particles, aggregates, hand specimens, and horizons) to the populations of soils comprising the landscape as a whole (pedons, toposequences, watersheds, physiographic regions, and the pedosphere) (Figure 34.1). Figure 34.2 illustrates hierarchical levels in this continuum of soil organization relative to tools used to generalize information content at multiple scales of resolution.

Various taxonomic systems have been developed to accommodate this cataloging of soil attributes and their landscape distribution. *Soil Taxonomy* (Soil Survey Staff, 1998, 2010) is such a system developed using morphogenetic indicators (diagnostic

horizons and properties) as class criteria. Chapter 31 discusses the history leading to the development of *Soil Taxonomy*, an overview of class criteria, nomenclature, and interpretational inferences. Although *Soil Taxonomy* was developed to categorize soils from across the globe, the functions and objectives of it and other systems of soil classification systems, including the French, Canadian, Russian, and Chinese systems, often have a national bias. To enhance international understanding of and ability to communicate about the soil resource, FAO with support from the International Union of Soil Science (IUSS) is leading the development of the World Reference Base for Soil Resources (http://www.fao.org/docrep/w8594e/w8594e00.htm), and the IUSS is providing leadership for development of a universal soil classification system to enhance our ability to categorize, understand, and extrapolate knowledge of the global soil resource.

Spatial variability in soil systems belongs to two broad categories: systematic (structured) and random (unstructured and unknown causes). Systematic variability is a gradual or marked change in soil properties as a function of physiography, geomorphology, and the interactions of the soil forming factors (Wilding and Drees, 1983) and our ability to predict systematic soil change across the landscape is based on the soil–landscape paradigm (Hudson, 1991). Systematic variability permits pedologists to partition spatial variability in soils by subsets of properties that constitute soil survey map units corresponding to geomorphic landscape elements. The purpose of soil surveys is to partition the spatial variability of landforms into stratified subsets that are less variable than the landscape as a whole and to remove systematic error from the population of soils. Information gained from soil surveys on the quality and distribution of soils when correlated with their classification provides a power vehicle for transfer of knowledge of soil properties and appropriate technologies for sustainable management. Chapter 33 illustrates characteristics of soils in each of the 12 soil orders in *Soil Taxonomy*, their properties, processes, and distribution patterns on a global scale. Each of the 12 soil orders are addressed as a subset of this chapter.

Pedologists are being challenged to better integrate information on soil properties, processes, and distribution into assessments of food and fiber production, soil quality, environmental sustainability, ecosystem services, and risk avoidance. Land evaluation, the consideration of actual and potential land use as a function of land properties, can be realized with qualitative methods, but increasingly quantitative simulation models are used together with geographic information systems (GIS) as aids for land use decisions. These models and GIS systems need to be fed with adequate soil data, but use of databases with little attention to natural soil dynamics or landscape relationships may lead to unrepresentative results. Chapter 34 discusses both classical qualitative and modern simulation methods of land evaluation. This chapter illustrates land evaluations through case studies for both data-rich and data-poor scenarios, and outlines the role of pedology and soil survey in land use decisions for sustainable ecosystems.

Pedology is adapting to the demand for more data, better methods, and new approaches to address complex natural processes that strongly influence our ability to interpret soil behavior and design management systems that maintain or enhance soil function in an ecologically and environmentally sustainable manner. The remaining chapters in this section describe a selected set of these new areas of study and application that likely will be important pedological subdisciplines in the twenty-first century.

Hydropedology is a combination of pedology, hydrology, and soil physics disciplines that addresses soil and water interactions. It addresses multiscale basic and applied pedologic and hydrologic processes and their properties in the variably unsaturated zone (Lin, 2003). Hydropedology focuses on the synergistic integration of pedology and hydrology to enhance our understanding of soil–water interactions and landscape–soil–hydrology relationships across space and time as they relate to ecosystem functions. Chapter 35 provides an overview of the fundamentals and applications of hydropedology, including a review of principles and description of recent advances in soil architecture and preferential flow, soil hydromorphology, scaling, pedotransfer functions, and coupling biogeochemistry with hydropedology.

Subaqueous soils, those soils inundated with water almost continuously (positive water pressure at the soil surface for at least 21 h each day in all years), are found along oceanic coastal margins and in shallow water associated with lake margins. Historically, these subaqueous substrates were considered to be sediments instead of soils. Recent research in these important ecosystems, however, have shown that subaqueous soils are, in fact, soils with horizons and properties that are the result of pedogenic processes. Chapter 36 provides an overview of subaqueous soils including the history of their recognition as soils, descriptions of methods for characterization and mapping, subaqueous soil classification, genesis models, and examples of interpretation of properties and distribution for soil use and management.

The explosion in computation and information technology has motivated numerous initiatives around the world to build spatial data infrastructures including regional, continental, and worldwide soil databases. In soil science, the principal manifestation is soil resource assessment using GIS to produce digital soil property and class maps with the limited expensive fieldwork and laboratory analysis. Digital soil mapping refers to the production of spatial soil databases that are based on soil observations combined with environmental data through quantitative statistical relationships. It uses a range of technologies for more accurate and efficient prediction of soil properties through optimal sampling strategies, rapid analysis of soil properties, and rapid acquisition of environmental variables over large areas and can be thought of as a means for modernizing and systematizing traditional soil survey. Chapter 37 describes the development of and theory underlying digital soil mapping and reviews the various approaches with examples.

Pedologists have long recognized that soils change over time. With a few exceptions, however, this change has mostly been related to changes over the course of geologic time and landscape development. Recently, greater emphasis has been placed

on the concept of soil change as it relates to the domestication of soil over historic time; the transformation of soil into a cultural–historic–natural system. To understand soil change and apply that knowledge to sustainable land management, we must understand how human activities impact the soil itself and soil's interactions with the wider environment. Chapter 38 discusses soil change including its impacts on soil function and ecosystem services, approaches to evaluate soil change, factors and mechanisms affecting soil change, and speculates on the future science and management of soil change.

The demand for more and more comprehensive data on soils as they vary across the landscape led to development and evaluation of numerous methods to rapidly evaluate soil properties on samples in the laboratory, at specific sites in the field, continuously across the landscape, and for fields or larger-sized areas. Among these techniques are diffuse reflectance infrared spectroscopy, gamma-ray spectroscopy, x-ray fluorescence, electromagnetic induction, and ground-penetrating radar. Advances in technologies to allow instrument portability and enhanced data analysis have made these technologies more commonly available for pedological research and soil inventories. Chapter 39 discusses two of these technologies, ground-penetrating radar and electromagnetic induction including the theory behind the technologies, initial uses, and their application to pedological studies.

References

Hudson, B.D. 1991. Soil survey as a paradigm-based science. Soil Sci. Am. J. 56:836–841.

Lin, H.S. 2003. Hydropedology: Bridging disciplines, scales, and data. Vadose Zone J. 2:1–11.

Soil Survey Staff. 1998. Soil taxonomy. U.S. Dept. of Agric. Natural Resources Conservation Service, Washington, DC. http://soils.usda.gov/technical/classification/taxonomy/

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. U.S. Government Printing Office, Washington, DC.

Sposito, G., and R. Reginato. 1992. Opportunities in basic soil research. Soil Science Society of America, Madison, WI.

Wilding, L.P., J. Bouma, and D. Goss. 1994. Impact of spatial variability on modeling, p. 61–75. *In* R. Rryant and M.R. Hoosbeek (eds.) Quantitative modeling of soil forming processes. Special Publication No 39. Soil Sci. Soc. Am. J. SSSA, Madison, WI.

Wilding, L.P., and L.R. Drees. 1983. Spatial variability and pedology, p. 83–116. *In* L.P. Wilding et al. (eds.) Pedogenesis and soil taxonomy: I. Concepts and interactions. Elsevier Publ. Co., Amsterdam, the Netherlands.

29

Geomorphology of Soil Landscapes

Douglas A. Wysocki
United States Department of Agriculture

Philip J. Schoeneberger
United States Department of Agriculture

Daniel R. Hirmas
University of Kansas

Hannan E. LaGarry
Chadron State College

29.1 Introduction

Soils form a continuum across the earth's land surface, that is, the interface of atmospheric, biological, and geological processes. Soils are more than a veneer of surficial alteration on landscapes or sediments. Soils, landscapes, and surficial sediments or rocks together comprise 3D systems that coevolve through the interaction and balance of physical and chemical weathering versus erosion and deposition. One must comprehend the relationships among soils, landscapes, and surficial sediments to fully understand soil landscapes, successfully predict soil occurrence, and anticipate soil behavior. Soil geomorphology is the scientific study of the origin, distribution, and evolution of soils, landscapes, and surficial deposits and

the processes that create and alter them. As a science, soil geomorphology is directly linked to pedology, geology, hydrology, archaeology, geomorphology, physical geography, ecology, and geotechnical engineering (Figure 29.1). Soil geomorphology relies primarily, but not solely, upon geological principles and techniques (Daniels and Hammer, 1992). Geologic/geomorphic processes substantially, but not solely, determine the materials from which soils are derived via the nature and redistribution of sediments. Principles and techniques drawn from geology or from other sciences often have applications or expressions that are unique to soils and soil landscapes. This chapter summarizes the major geomorphic principles and techniques used to understand the relationships among soils, landscapes, surficial sediments (Hunt, 1986), and internal water movement.
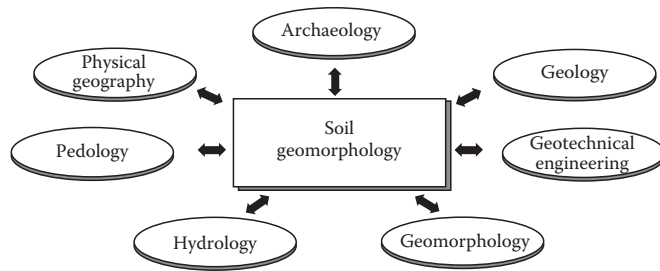
FIGURE 29.1 Conceptual diagram showing the relationships of soil geomorphology and other sciences.

29.1.1 Goals of Soil Geomorphology

Soil geomorphology serves two important capacities: (1) Geomorphic-based landscape models allow the segregation of the soil continuum into meaningful natural soil bodies that explain soil distribution from hillslope to continental scales. (2) It provides basic principles for understanding the geomorphic history of landscapes (e.g., spatial and time relationships among soils, landscapes, and surficial sediments).

29.2 Key Terminology

Before discussing the principles of soil geomorphology in detail, a few key definitions are needed. Although some terms may have alternate definitions, our discussion uses those following.

29.2.1 Depositional Surface

Depositional surface is a part of a geomorphic surface formed by the deposition of sediments (alluvial, colluvial, eolian, etc.) derived from an erosional surface or erosional processes (Ruhe, 1975). Water, wind, ice, and gravity are active agents that construct and shape depositional surfaces with water being the most universally prevalent agent.

29.2.2 Erosional Surface

An erosional surface is a part of a geomorphic surface formed by removal of rock, soil, or sediment under the wearing and transport action of water, wind, ice, and gravity (Ruhe, 1975). Running water is the dominant agent that forms and shapes most erosional surfaces (wind, ice, or gravity dominant agent in select settings).

29.2.3 Geomorphic Surface

A geomorphic surface is a definable land surface area that forms during a given time under a common set of erosional and depositional processes. A geomorphic surface may include multiple landforms, but it is a mappable spatial area based upon geomorphic techniques and field observations that has specific and identifiable borders (Ruhe, 1975).

29.2.4 Landform

A landform is any physical, recognizable feature on the earth's surface, having a characteristic shape and internal composition that is produced by natural processes (SSSA, 1997). Landforms are the individual features of the earth that together comprise the land surface (Ruhe, 1975).

29.2.5 Landscape

A landscape is a spatially adjacent collection or assemblage of landforms that can be observed in a single view or from a given vantage (Ruhe, 1975). The landforms that compose a single landscape may have the same age and origin or may vary in both age and origin. Another definition for landscape is the portion of the land surface the eye can comprehend in a single view (Ruhe, 1969).

29.2.6 Pedon

A pedon is a 3D soil body l with lateral dimensions large enough to permit the complete study of horizon shapes and relations. A pedon ranges from 1 to 10^2 m in land surface area and has a maximum defined depth of 2 m (SSSA, 1997).

29.2.7 Soil

Soil is a natural, 3D body with definable boundaries that occurs on the land surface composed of solids (mineral and organic), liquids (water), and gases. Soil is characterized by horizons distinguishable from the initial material as a result of additions, losses, transfers, and transformations in the surface environment and/or by the ability to support higher plant forms (Soil Survey Staff, 2010).

29.2.8 Soil Delineation

Soil delineation is an individual polygon shown on a soil map with a map unit symbol and name that defines a 3D soil body of specified area, shape, and location on the landscape (SSSA, 2010).

29.2.9 Soil Map Unit

A soil map unit is an aggregate of all soil delineations in a soil survey area that have a similar, established set of characteristics (Van Wambeke and Forbes, 1986). Each delineation of a soil map unit is identified by the same symbol and name in a soil survey.

29.3 Soil as a Landscape Unit or Body

Soil geomorphology is a field science that studies soils on landscapes. The soil continuum on a landscape, for ease of comprehension, is divided into discrete individuals. The pedon is the soil unit most commonly described, sampled, and classified by pedologists. The relative size or scale of a pedon is an intellectual
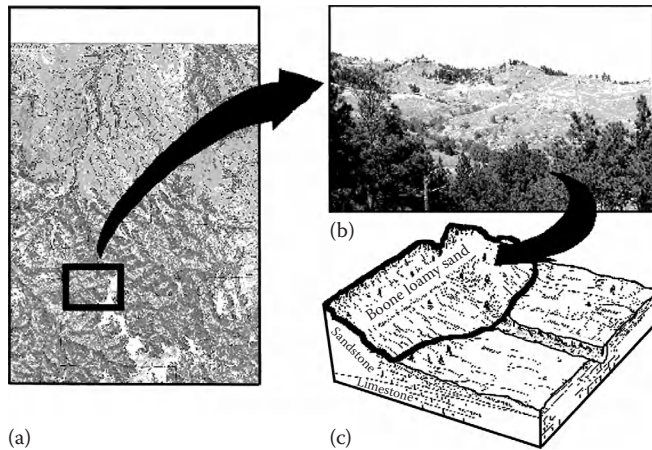
FIGURE 29.2 (a) Landscape image, (b) soil map image, and (c) block diagram showing soil as a landscape unit.

FIGURE 29.3 Conceptual diagram of Jenny's five soil-forming factors climate, biota, parent material, topography, and time.

construct useful and functionally necessary for description and classification. Pedons, however, lack distinct lateral boundaries and are, therefore, not natural landscape units. Geological processes, which are driven by the atmospheric agents of water, wind, ice, and gravity, impact landscapes as a continuum. Landscapes, however, possess natural internal boundaries that restrict or control mass and energy transfer. Examples of such boundaries are topographic divides, contacts between different rocks or sediment bodies, inflections in slope gradient or shape, and contacts between landforms of different age, origin, and internal structure.

29.3.1 Landscape Scale and Function

Soil delineations depicted on a soil map at or near a scale of 1:24,000 are more closely linked to geomorphic processes than are individual pedons. Accordingly, soil delineations (groupings of similar pedons) are better suited to geomorphic studies than individual pedons. Soil delineations depicted in a soil survey are landscape units (Figure 29.2). Boundaries between delineations are established by the soil surveyor, but not in an arbitrary fashion. Boundaries on a soil map delimit observable differences in soil morphology such as horizon type and thickness, and soil color, texture, and structure across the landscape. Changes in soil morphology across a landscape generally result from differences in the transfer of mass and energy driven by ecological, geomorphic, or atmospheric processes. Soil surveyors use observation, experience, and geomorphic landscape models to create soil maps.

29.4 Models of Soil Formation

Soils are complex systems that defy easy comprehension. Therefore, we use scientific models to help understand or explain soils. No single model provides complete understanding. Models of various form, function, and design are needed to understand
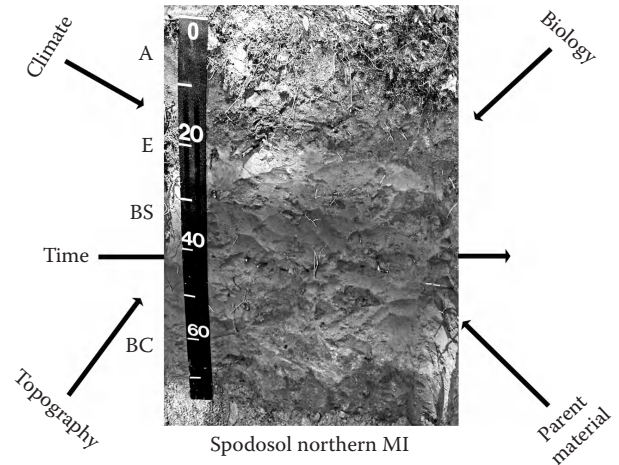
systems as complex as soils (Dijkerman, 1974). Two important and well-known soil models are those of Jenny (1941) and Simonson (1959).

29.4.1 Jenny's Factors of Soil Formation

The eloquent, conceptual model of Jenny (1941) describes soil as a function of climate, biological influences, topography, parent material, and time (Figure 29.3). Implicit in this model are the distinct relationships among ecosystems (biological factor), landscapes (topography), surficial sediments (parent material), and landscape evolution (time). Stratigraphy of sediments or bedrock and surface contours strongly influence the movement of water within and over the landscape. Topography and parent materials have a strong control on both local (e.g., soil water tables, water holding, nutrient capacity, salt content, soil temperature) and regional soil environments (e.g., rain shadows, elevational induced climate zones, adiabatic winds) and therefore impact ecosystem form and function. All five soil-forming factors are linked either directly or indirectly to landscapes, surficial sediments, and landscape evolution.

29.4.2 Simonson's Process Model

Simonson (1959) explained soil formation through the interaction of four processes: additions, removals, translocations, and transformations (Figure 29.4). This model is more helpful than Jenny's (1941) for understanding the spatial relationships within soil landscapes. Geologic or geomorphic processes cause additions, removals, translocations, and transformations on a landscape scale that create and modify landforms, sediments, and soils. For example, sediment eroded from the flank of a hillslope is deposited as colluvium at the base of the slope or as alluvium in nearby drainageways or floodplains. The sediment is incorporated into the uppermost horizons of an existing soil or becomes fresh parent material in which a new soil begins to form.

Additions
Water as precipitation, condensation and runoff
O, N, and CO_2, and dust from the atmosphere
Organic matter by the biosphere
Sediment by slope or water deposition
Energy by solar radiation.

Losses
Water by evapotranspiration
N by denitrification
CO_2 by organic matter oxidation
Sediment removal by erosion
Energy by radiational cooling

A Horizon

Transformations
Organic matter humification
Particle size reduction by weathering
Soil structure formation
Redox concentrations and depletions
Clay mineral alterations

Translocations
Clay, organics, sesquioxides by water
Plant nutrient uptake and cycling
Soluble salts in water
En masse by fauna

B Horizon

C Horizon

Losses
Water including chemicals in solution
and colloids in suspension

FIGURE 29.4 Simonson's (1959) process model of soil genesis showing the interactions of additions, removals, transfers, and transformations.

29.5 Soil Landscape Models

29.5.1 The Catena

The models of Simonson (1959) and Jenny (1941) provide an important conceptual framework for understanding soil formation. Neither model, however, establishes functional boundaries for segregating the soil continuum into natural landscape units. The catena is a fundamental concept that explains the soils pattern on hillslopes. Milne (1936a) coined the term catena to describe a repeating sequence of soils that occurs from the hillslope top to the adjacent valley bottom. Milne (1936a, 1936b) distinguished two catena types. The first type occurred on hillslopes developed over a single kind of parent rock. Despite the uniformity in parent rock, Milne observed a sequential change in soils along the slope gradient. Milne attributed the sequence of soils to variations in subsurface drainage, lateral transport of sediments, and the translocation of materials at or beneath the soil surface (Figure 29.5a).

In Milne's (1936a, 1936b) second example, the hillslope contained more than one type of parent rock (Figure 29.5b). An observable sequence of soils also occurred on this hillslope. Variations in drainage and lateral transport also produced this catena, but stratigraphic differences in the parent rock increased the complexity of the soil pattern. In this example, the surficial sediments form a drape on the landscape that is not coincident with the underlying rock strata. The catena concept includes both surficial stratigraphy and internal hillslope structure or lithology.

Furthermore, the catena concept is both a soil landscape model and a geomorphic model or system. Milne (1936a) recognized that lateral sorting contributed to the sequential soil variation down the hillslope. Erosion and sedimentation processes driven by relief and water movement redistribute sediments across hillslopes creating subtle lateral differences in soil parent materials (Kleiss, 1970). Recall that both parent material and topography are factors in Jenny's (1941) model of soil formation. The same erosional and depositional processes that drive landscape evolution influence the soil pattern on landscapes. The sequential change in soil morphology across a landscape is linked by process to landscape evolution on hillslopes both in time and space. In Jenny's model, landscape evolution means that parent material and topography are not independent variables, but rather are dependent variables that can covary in time.

A well-studied catena consists of the Clarion–Nicollet–Webster soils that occupy about 31,000 km^2 of the Des Moines Lobe in south-central Minnesota and north central Iowa (Figure 29.6). The Des Moines Lobe represents the last Late Wisconsinan glacial advance into Iowa about 14,000 YBP (Ruhe, 1969). Major topographic areas include hummocky, high-relief end moraines separated by undulating areas of low- to moderate-relief ground moraine (Ruhe and Scholtes, 1959; Ruhe, 1969; Kemmis et al., 1981). Closed, semiclosed, and linked depressions occur throughout the Des Moines Lobe, but closed depressions are most abundant in the end moraines (Kemmis, 1991). The Clarion–Nicollet–Webster soils are mollisols formed in a stratigraphic sequence composed of hillslope sediments, supraglacial sediments, and loam-textured till. The catenary relationships on the Des Moines Lobe landscape result from surficial sorting during deglaciation (Kemmis, 1991; Steinwand and Fenton, 1995), postglacial hillslope sorting (Walker, 1966; Burras and Scholtes, 1987; Steinwand and Fenton, 1995), and subsurface flow relationships (Steinwand and Fenton, 1995).
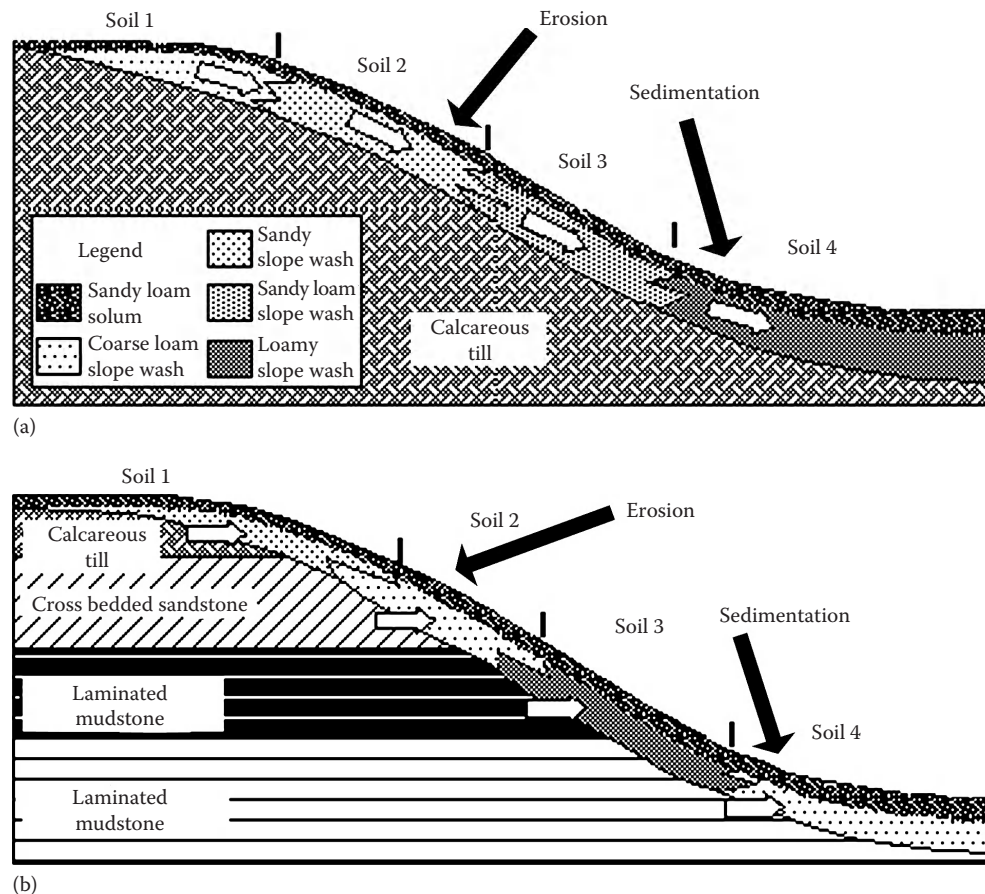
FIGURE 29.5 Two-dimensional diagram of Milne's (1936a, 1936b) catena showing idealized landscape relations.

Numerous studies (Dan and Yaalon, 1964; Blume and Schlichting, 1965; Blume, 1968; Walker and Ruhe, 1968; Huggett, 1975; Pennock and Vreeken, 1986; Pennock and Acton, 1989) have confirmed that catena relationships occur in various climates and landscapes. Conacher and Dalrymple (1977) and Dalrymple et al. (1968) provided a quantitative description of the catena. They defined the soil hillslope relationship as a 3D unit having arbitrary lateral dimensions extending from the hilltop to the valley bottom and from the soil surface to the base of the solum. They segmented the hillslope into nine land surface units (Figure 29.7) based on soil morphology, mobilization and transport of soil constituents, redeposition of soil constituents by overland and throughflow, or by gravity as mass movements.

29.5.2 The Toposequence

Bushnell (1942) studied morphological differences in soils across a hillslope gradient. The soils differed mainly in color. He attributed the changes in morphology to elevational position and local hydrology. This concept is commonly referred to as a toposequence. Unlike Milne, Bushnell did not recognize the influence of hillslope erosion and sedimentation. The terms catena and toposequence are often presently used as synonyms, but the original meanings are not identical.

29.5.3 The Valley Basin

Huggett (1975) expanded on the catena concept proposing that the basic 3D unit of the soil landscape is the first-order valley basin. The functional boundaries of this soil landscape are defined as the atmosphere–soil interface, the weathering front at the soil base, and the drainage basin divides. The topographic boundaries of a drainage basin define the physical limits and direction of overland flow and thus control geomorphic processes such as erosion, transport, and deposition. Also, groundwater divides are generally coincident with topographic divides that are boundaries for overland flow. Thus, the first-order basin forms the natural boundary conditions for chemical and colloidal transport and redistribution via water transport on most landscapes (Figure 29.8a). First-order basins connect to higher order basins forming a larger, linked system across landscapes.

The partitioning of water into surface and subsurface flow is an important component of both Huggett's (1976) watershed model and the catena. Surface runoff or overland flow is the mechanism that drives geomorphic processes. Water that enters the soil drives chemical and biological reactions and via throughflow transports soluble or colloidal materials. The movement of water and associated materials through the soil landscape occurs as both saturated and unsaturated flow. Transport
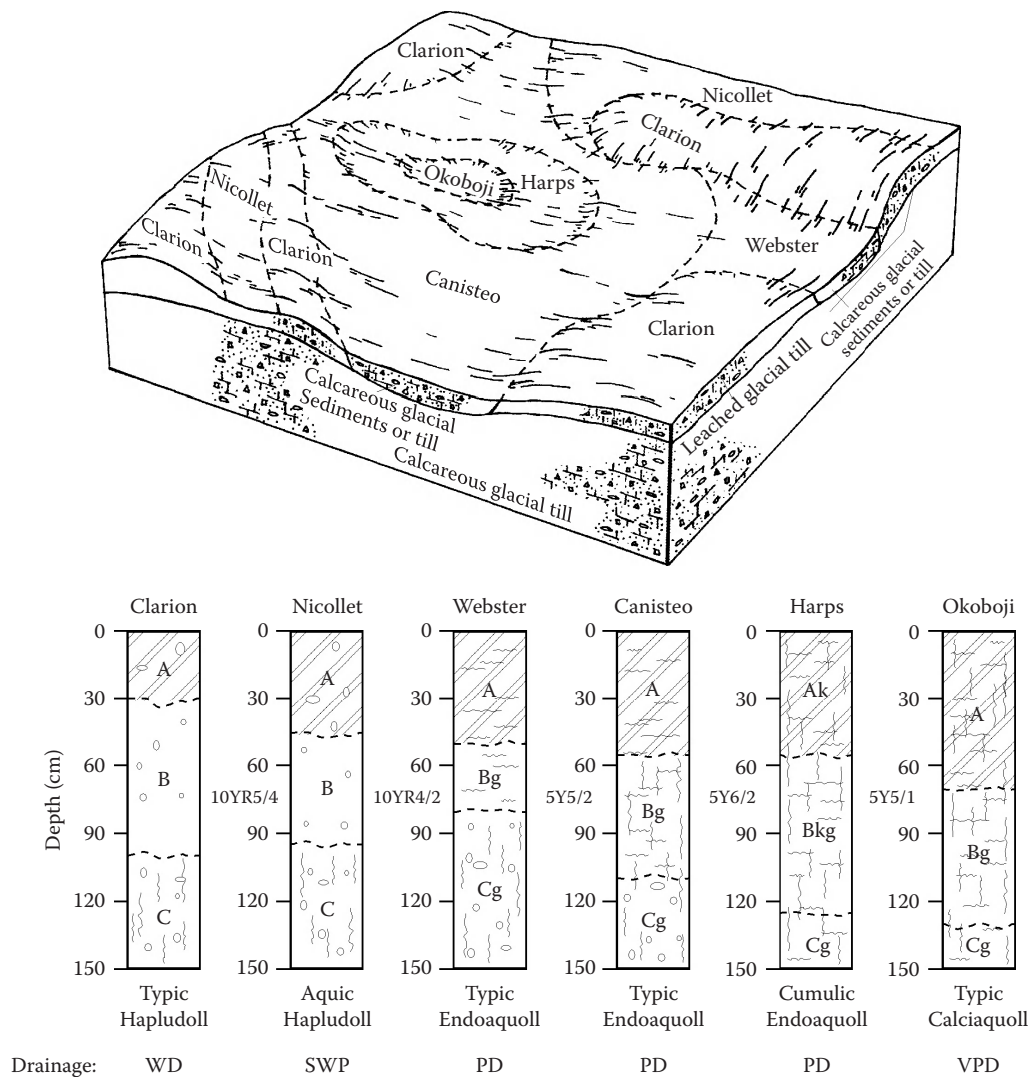
FIGURE 29.6   Two-dimensional diagram showing the landscape relations of the Clarion–Nicollet–Webster catena.

of material can occur between or within soil horizons, between soils at different landscape positions, or across an entire watershed. Deep percolation can transport material beyond the depth of a soil profile. A 3D approach must account for lateral, divergent, and convergent throughflow. Recent studies (Arndt and Richardson, 1989; Knuteson et al., 1989; Steinwand and Richardson, 1989), as well as earlier work (Glazovskaya, 1968; Cleaves et al., 1970; Crabtree and Burt, 1983), attest to the importance of throughflow and the geochemical link between soils, hydrology, and landscapes.

29.5.4 Open Basins

Drainage basins are either open or closed (Ruhe, 1969). An open drainage basin is confined by the head of the watershed and the perimeter divide. The mouth of the basin is open (Figure 29.8a). Surface water collected within the basin during precipitation events discharges through the outlet at the mouth. Eroded sediment can

be transported and redeposited either within the basin or removed entirely from it. Both mass and energy can enter and leave an open basin by surface flow or runoff. The sediment retained within an open basin represents only a portion of the total sediment produced by erosion. The sediment package in an open basin is an incomplete record of the geologic history. A study of the sediment stratigraphy, therefore, yields only a partial understanding of the geomorphic evolution of the basin.

29.5.5 Closed Basins

Closed drainage basins lack a surface outlet. The drainage pattern within the basin flow to a central area or point (Figure 29.8b). No surface water leaves the basin as runoff. Water from precipitation is lost by evaporation or transpiration or infiltrates and becomes subsurface water. Mass and energy transfers driven by surface flow or runoff occur only within the basin. Surface flow does not remove mass from the basin in the form of clastic sediment. Mass can only
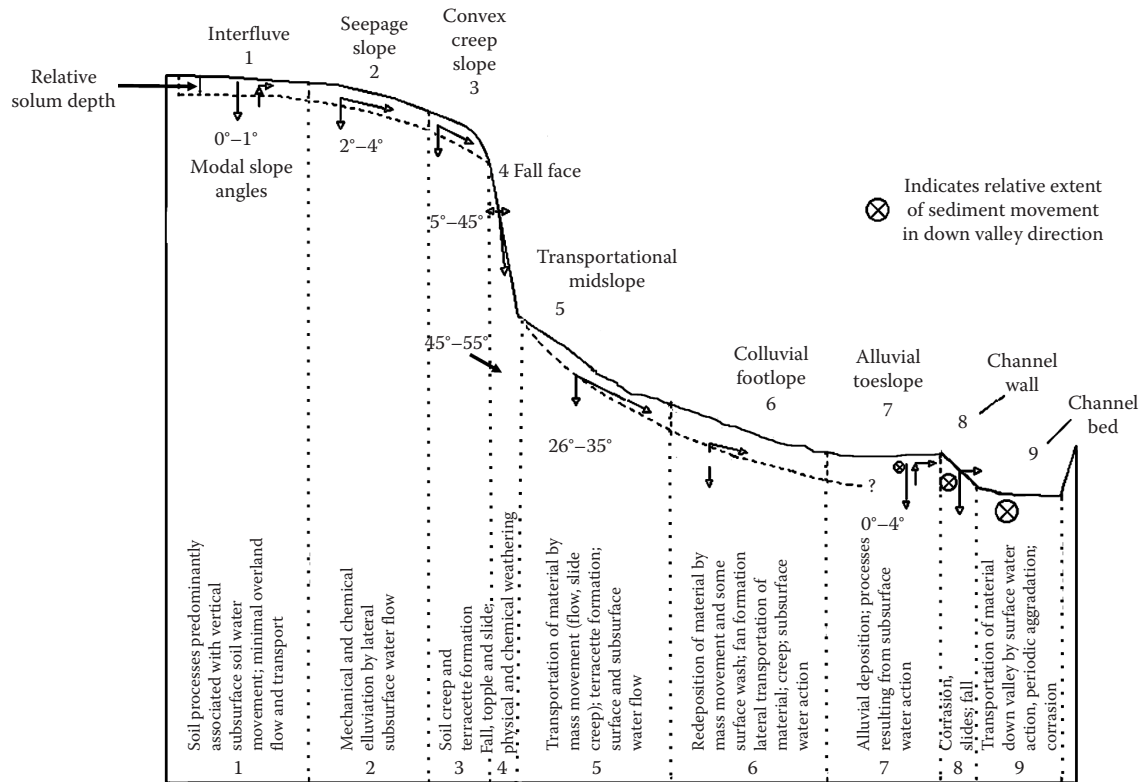
FIGURE 29.7 Two-dimensional nine unit soil and hillslope model (Dalrymple et al., 1968; Conacher and Dalrymple, 1977) showing soil hillslope relationships.
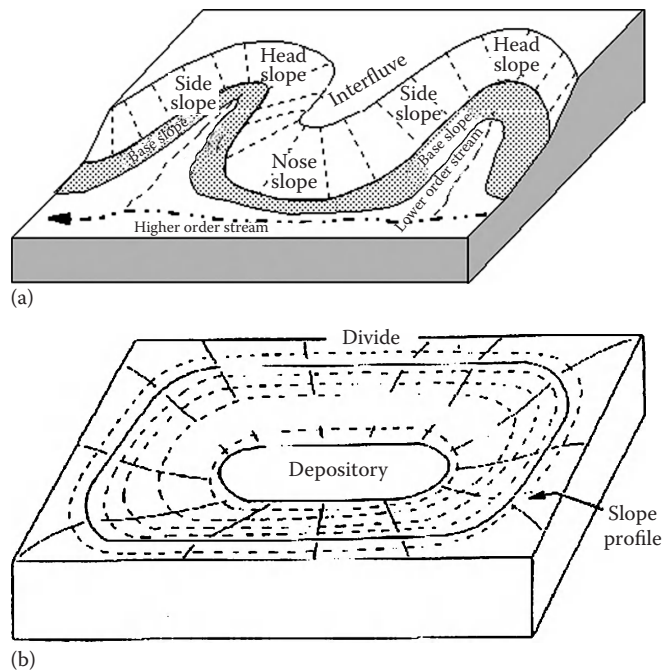
FIGURE 29.8 Three-dimensional diagram (a) open (Schoeneberger and Wysocki 1997) and (b) closed drainage basins as soil landscape models. (Modified from Walker, P.H., and R.V. Ruhe. 1968. Hillslope models and soil formation. I. Closed Systems. Trans. 9th Int. Congr. Soil Sci. 4:561–568.)

be lost as soluble or fine colloidal material in subsurface water. Mass can be removed or added to the basin by eolian processes, but the system is closed to surface runoff. The alluvial sediment contained in a closed basin forms a complete record of its erosional history.

The geomorphic concept of surface runoff that defines open versus closed basins is independent of scale. Open basins span the range from tiny, erosional rills on a hillslope to continental scale drainage basins like those of the Amazon or Mississippi rivers. Likewise, closed basins vary in size from small depressions like tree-tip pits to large tectonic basins such as Death Valley. The similarity is the lack of a surface outlet. The magnitude and complexity of the transfer processes within a basin increase with size. Surface flow never removes mass or energy from a closed basin regardless of size.

The lack of a surface outlet in closed basins precludes initial formation by water erosion. Closed basins originate through several mechanisms including deposition by ice or wind, dissolution of relatively soluble rock, subsidence due to structural failure of underlying rock often related to dissolution or groundwater fluctuations, and subsidence caused by faulting or tectonic downwarping. Closed basins are common in landscapes produced by these geomorphic processes. Examples include the "prairie potholes" and "chain lakes" of the glaciated midcontinent, depression lakes in dune topography such as the Sand Hills of Nebraska, karst topography created by dissolution of limestone, and the playa lakes of the southern Great Plains, which are partially related to wind deflation and deposition.
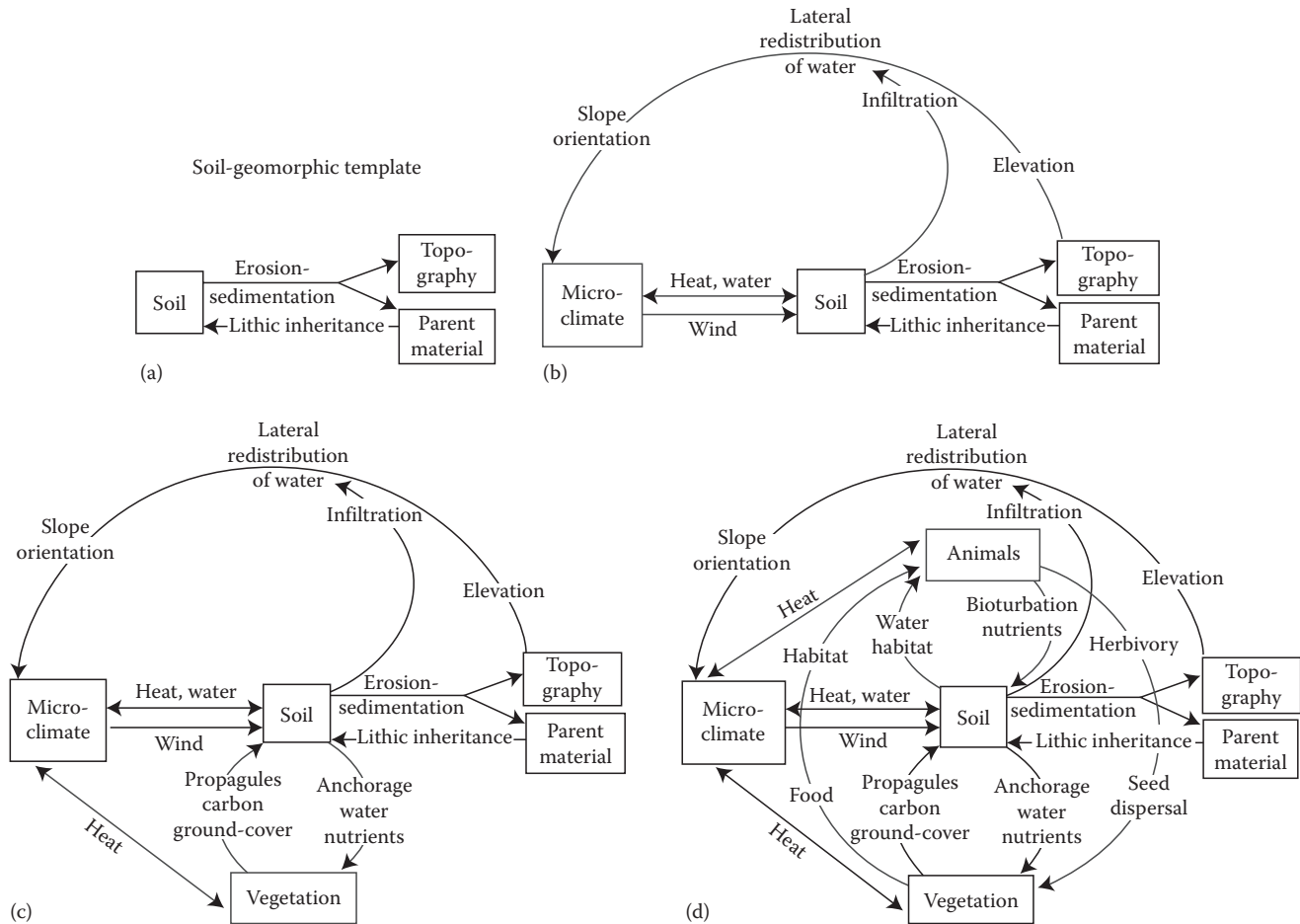
FIGURE 29.9 Linkages of the soil-forming factors in the soil-geomorphic template. (Modified from Monger, H.C., and B.T. Bestelmeyer. 2006. The soil-geomorphic template and biotic change in arid and semi-arid ecosystems. J. Arid Environ. 65:207–218.)

29.5.6 The Soil-Geomorphic Template

A more recent development in soil landscape models is the soil-geomorphic template introduced by Monger and Bestelmeyer (2006). It is a conceptual framework that depicts the interrelationships between the five soil-forming factors of Jenny (1941). Figure 29.9 illustrates these interrelationships. The soil-geomorphic template is defined as "…the soil, topography, and soil parent material" occurring on the landscape as a predictable pattern, which is reflected by the biotic community. Thus, it centers around the connection between topography and soil substrate on the one hand and ecological communities on the other. As such, it offers a framework to integrate various, often separated, disciplines to understand soil-geomorphic and ecological change across both short-term (centuries) and long-term (geologic) timescales.

The soil-geomorphic template has been incorporated in a proposed systematic procedure that can be used to unravel the complexity of landscapes (Peters et al., 2006; Figure 29.10). In this procedure, the researcher first identifies variation in the broad-scale environmental drivers and determines the spatial extent over which these drivers influence patterns on the landscape. Historical legacies are, then, identified and connections are made between these legacies and current landscape patterns.
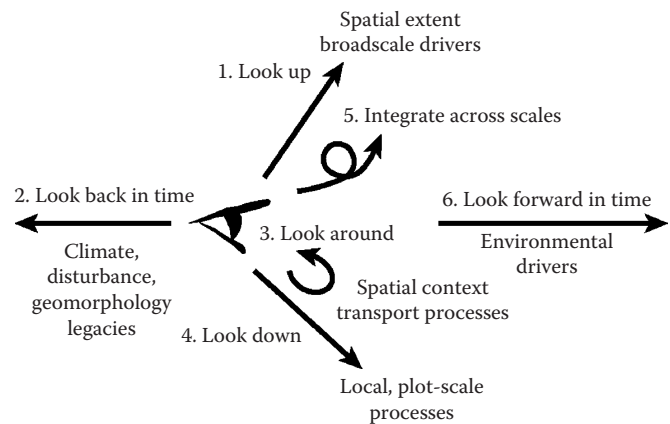
FIGURE 29.10 Six-step procedure to unravel landscape complexity. (Modified from Peters, D.P.C., B.T. Bestelmeyer, J.E. Herrick, E.L. Fredrickson, H.C. Monger, and K.M. Havstad. 2006. Disentangling complex landscapes: New insights into arid and semiarid system dynamics. Bioscience 56:491–501.)

Next, meso- and fine-scale spatial properties of the landscape units and the specific drivers connecting these units are identified. The information is integrated and key drivers and properties are identified at each spatial scale. Last, future effects from hypothesized changes in drivers and feedbacks are predicted.

29.6 Soil Hydrology

A major advancement in understanding soil systems has occurred in the last 20 years through insight into how water moves through landscapes: where the water goes, so goes soil development. Historically, the study of water flow in soils emphasized either overland flow, erosion, and sedimentation, or water movement within a pedon. Erosion and sedimentation control were foci of national soil programs and an integral part of agronomic applications of soil science. Efforts to understand and quantify water flow within a pedon or within fields (artificially delimited management areas versus naturally delimited bodies), such as for irrigation, have been an integral part of soil physics studies for over a century. However, water movement has not been pervasively integrated into the study of soil landscapes. To understand natural soil systems and thereby ecosystem behavior, conceptual and quantitative water-movement models must encompass water flow through soils, the vadose zone, and across and through landscapes.

Milne's concept of the catena (Milne 1936a, 1936b) recognized the existence of soils with different drainage classes across landscapes. His recognition of lateral transport as a contributing factor to landscape evolution has been de-emphasized over time. The main focus shifted to relatively static soil water conditions and vertical water flow (deep percolation) at a given position, hillslope or landscape position. Lateral transport was primarily recognized as overland flow with concomitant sediment sorting across the surface (Walker et al., 1966; Ruhe, 1975). The equally important processes of subsurface water flow and transport through soil landscapes were substantially ignored or forgotten. Soil hydrology studies in the last 20 years have expanded the catena concept to include water flow through landscapes (Arndt and Richardson, 1989; Steinwand and Fenton, 1995). There has been a simultaneous recovery from the historical emphasis on soil erosion caused by overland flow. The study of soil systems now includes the dynamic flow of water through as well as across soil and landscapes (Figure 29.11). This approach of dynamic water flow considers the entire vadose zone, not just the ground surface or the solum.

A variety of soil hydrologic terms have been developed or adopted, which partition water flow in soil systems. The terms can be placed in an idealized schematic to demonstrate relationships (Figure 29.12). These terms portray the potential fate of precipitation onto, into, and through soils. The emphasis is on water at or above the permanent groundwater table (i.e., the vadose zone). Water dynamics and aquifer conditions below the water table generally are the purview of groundwater hydrology. Water flow can be predominantly downward, lateral, or upward toward the ground surface, depending upon prevailing energy dynamics. This can be demonstrated by looking at prevailing water movement patterns in different climatic settings.
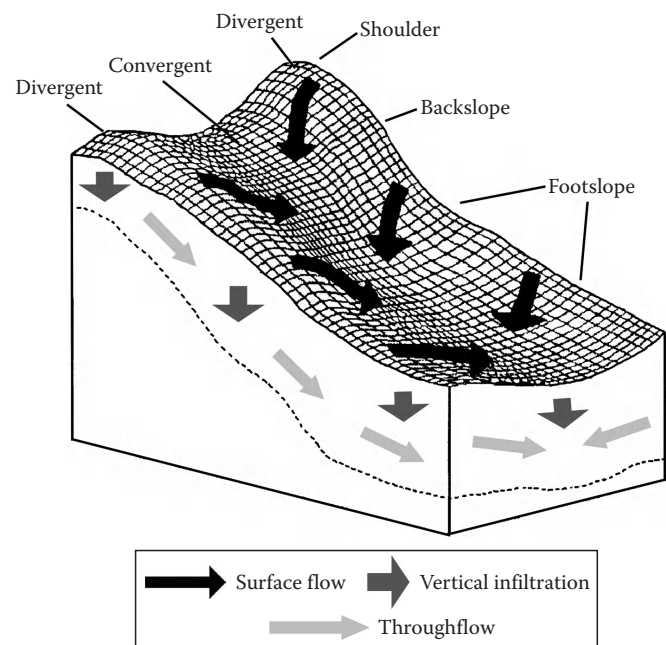
FIGURE 29.11 Three-dimensional hillslope with flow directions.
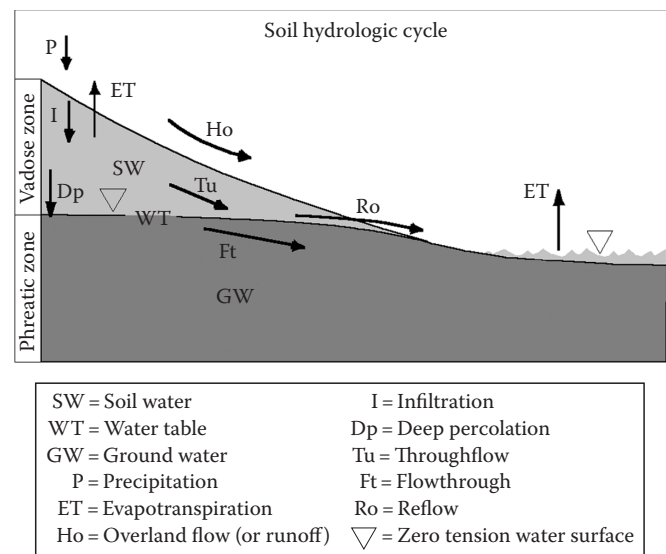
FIGURE 29.12 Soil hydrologic cycle with terminology.

29.6.1 Climate Influence

Regional climate conditions control precipitation levels and patterns and evapotranspiration dynamics (either directly via solar radiation or indirectly through vegetation populations). All else being equal, regional climate determines the dominant water flow direction through the vadose zone. Within a given temperature range, precipitation levels can vary with predictable, generalized results. In humid environments, water tables and flow paths tend to loosely mimic the local topography with water generally infiltrating (i.e., *recharge*) in topographically high areas, flowing to and discharging from lower-lying areas (Figure 29.13a). In arid to semiarid environments, conditions are reversed, with recharge sites dominantly in

low areas and subsurface water moving to and "discharging" from higher elevation areas (Figure 29.13b). Semiarid to subhumid environments behave as tension or transition zones between these two extremes. These transition zones temporarily follow either pattern depending upon short-term climatic conditions (e.g., seasonal climatic variations or annual cycles; Figure 29.13c).

Temperature also exerts a controlling climatic influence on soil hydrologic behavior both permanently and ephemerally. For example, permafrost effectively limits or precludes vertical water flow. If local conditions warm (human-induced or natural), permafrost melts and ceases to be a restriction to water flow. A seasonally ephemeral example occurs in areas that experience annual ground frost. The internal water flow behavior in soil is vastly different between winter, spring thaw, and summer (Emerson et al., 1990). Ground frost when present is an effective barrier to both

subsurface water flow and the infiltration of surface water. During several weeks in the spring, thawed soil layers immediately above a remnant frost layer can become saturated and highly erosive (e.g., Willamette Valley, Oregon). After the soil completely thaws, subsurface flow is greatly enhanced and the surface layer is no longer saturated. The transport and fate of contaminants can be radically different depending upon seasonal soil conditions. This is the basis for discouraging land application of manure on frozen soil.

29.6.2 Geostratigraphic Influence

Water restrictive (aquitard) or conductive (aquifer) sediments or rock strata can fundamentally alter subsurface water flow patterns. Tilted rock strata can preferentially redirect water such that infiltration in one location is shunted (laterally displaced) to an unlikely recharge or discharge site. Where bedrock strata approach or breach the ground surface, the vertical movement of subsurface water can be restricted, forcing water to flow back to the surface (i.e., *reflow*; Figure 29.12) and to form a spring, seep, or moist area. The composition of subsurface stratigraphic sequences and the extent to which they are connected to conductive materials also influence subsurface water flow. For example, a porous strata that is normally conductive will be nonconductive, if no outlets to other conductive material are available (e.g., a sand layer in a floodplain deposit confined by clay sediments; Figure 29.14). Similarly, a topographic position that precludes or minimizes water inputs may cause a strata that is normally conductive to be nonconductive.

29.6.3 Pedostratigraphic Influence

Some features unique to soil or derived from soil processes emulate geostratigraphic influences on water flow. Some pedologically derived layers such as an argillic layer (Schoeneberger et al., 1995), or duripan can restrict vertical water flow and enhance lateral flow (Figure 29.15). Water-repellent materials or layers (e.g., fire-induced hydrophobic layer in chaparral) can reduce infiltration of water into the soil or function as an aquitard and restrict vertical movement. Other pedogenic features or related phenomena such as pedogenic structure and biotic activity can greatly enhance water flow in upper soil horizons (Schoeneberger and Wysocki, 1996). Substantial differences in infiltration or internal water
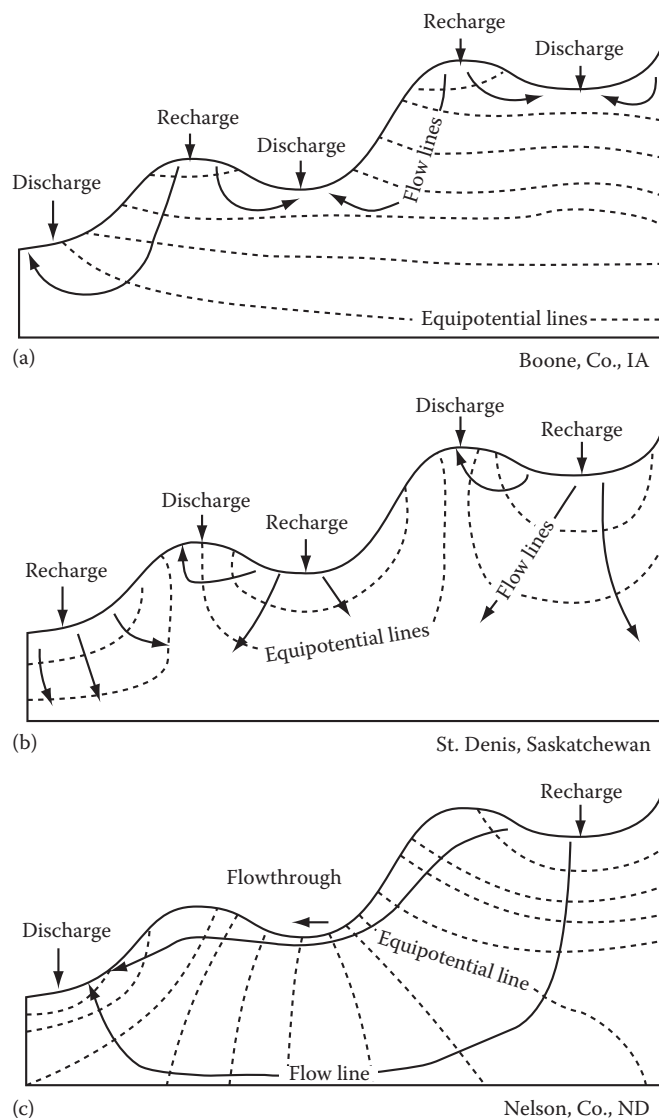
FIGURE 29.13 Two-dimensional landscape diagrams with recharge–discharge, flow lines, and equipotential lines in different climatic settings (a) humid, (b) arid to remiarid, (c) remiarid to subhumid.
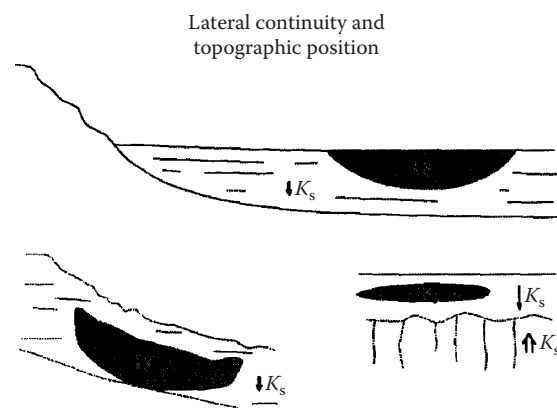
FIGURE 29.14 Stratigraphic isolation of normally conductive strata.
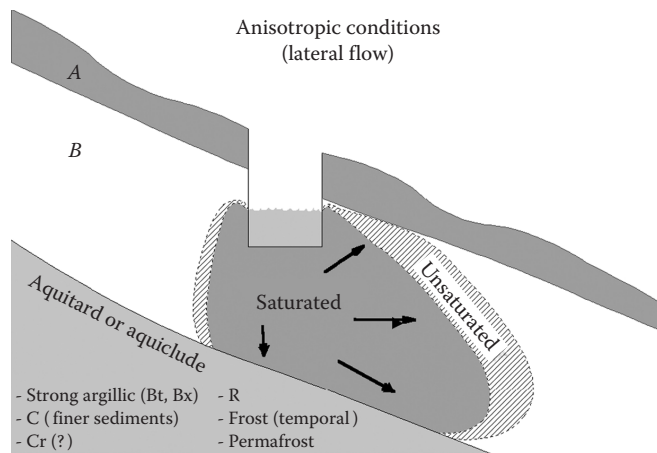
FIGURE 29.15   Pedogenic or geogenic acquitard restricting water flow through soil or strata. Note the A and B in the figure are soil horizon symbols. (From McAuliffe, J.R., E.P. Hamerlynck, and M.C. Eppes. 2007. Landscape dynamics fostering the development and persistence of long-lived creosote bush (*Larrea tridentata*) clones in the Mojave Desert. J. Arid Environ. 69:96–126. With permission from Academic Press.)

movement in a soil can result directly from different management practices (Franks et al., 1993, 1995).

29.6.4 Vegetative Pumping

Biotic activities, particularly plant respiration, can directly impact water flow patterns in soils. Plant communities can consume substantial volumes of water during transpiration. Water consumption may be sufficient to change local flow dynamics. Consider the case of phreatophytes fringing a marshes or riparian system (Figure 29.16). Minor differences in elevation (e.g., 10 cm) can promote plant establishment resulting in a disproportionately large reduction in water table, and subsequent differences in soils (e.g., the soil drainage class).

29.6.5 The "Soil Sponge"

An important but often overlooked function of soil is its role as a hydrologic "sponge." The porous character of soil allows
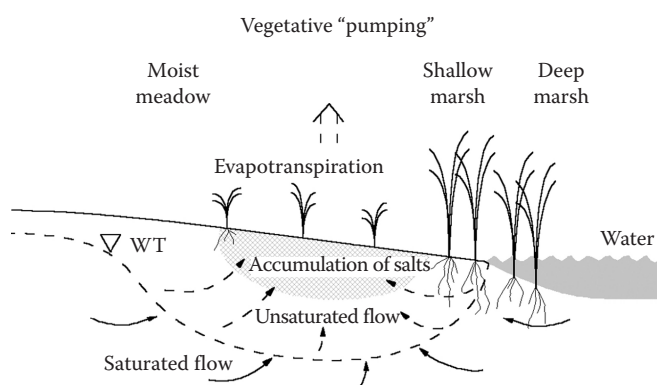
FIGURE 29.16   Diagram of vegetative pumping. (Drawn after Arndt and Richardson, 1989.)

much of the local precipitation to infiltrate. Obviously, the rate of water flow through the porous media of soil is less than that above ground. Infiltrated water is gradually transmitted vertically to recharge groundwater or laterally to be released as discharge to surface waters via throughflow. The hydrologic result is a delayed and dampened streamflow hydrograph. The importance of streamflow response can be seen from a practical, human perspective. The streamflow response for areas from which substantial amounts of soil are removed or sealed over (e.g., construction sites in urban areas) typically results in much higher and quicker peak flow compared to original conditions. This pattern can have a profound financial impact if structures are underdesigned (culverts, bridges, etc.) or if the new hydrologic dynamics initiates a new cycle of stream incisement or substantial change in sedimentation patterns (Hammer, 1995).

29.6.6 Soil Morphology, Landscapes, and Water

Water is the dominant catalyst, mediator, and transport agent in most natural systems on earth. This is true for soils. Flora, fauna, and geologic substrate all affect soil and landscape development; water, however, is the controlling force providing both chemical and potential energy. The presence and flow of water in a landscape can be observed directly by excavation or piezometers. This is a tedious and time-consuming task given the seasonal or transient presence of water. The presence of saturated conditions and the movement of water in landscapes can be predicted from soil morphology (e.g., redoximorphic features). Soil morphology forms in response to long-term prevailing water state conditions. Soil patterns on a landscape result from differing water conditions and the prevailing water flow dynamics. Soil patterns on landscapes can explain where the water is and how it flows through landscapes. Soil morphology and soilscape models can explain and/or predict soil water dynamics and subsequent soil geography.

29.6.7 Eolian Influences

In addition to, or in lieu of water, wind can serve as a dominant mediator or transport agent on landscapes. Eolian processes directly affect soils and landscapes through the addition, redistribution, or loss of clastic materials, soil nutrients, carbonates, and soluble salts. The wind-blown sediments entrained through eolian processes can significantly alter soils and landscapes from individual landforms to entire regions. The effects of these sediments vary depending upon (among other factors) climate and localized landform characteristics and include the following: supplying essential nutrients to plants, influencing the character of near-surface horizons that directly affect surface hydrology, providing materials necessary for the development of certain morphologic horizons, and providing the parent material in which thick, well-developed soils have formed (Simonson, 1995). Examples of landscapes that are dominantly affected by eolian processes are those in which soils have developed in thick loess mantles, areas inundated with volcanic ash, dune systems, and

regions with arid and semiarid climates. The impacts of loess, volcanic ash, and eolian sand (dunes) are generally geomorphically recognized. The influence of dust is less recognizable.

29.6.8 Arid and Semiarid Dust

Two profound ways in which dust can affect soil landscapes in arid and semiarid areas are as follows: the evolution of accretionary desert pavements and the formation of coppice dunes. Desert pavements are naturally occurring surficial features that consist of closely packed coarse gravel to cobble-sized rock fragments embedded into the underlying soil (Wood et al., 2005). These surfaces are often associated with silt-enriched horizons containing vesicular pores that directly underlie these pavements. Vesicular horizons are important because they significantly decrease infiltration rates into soils of these landforms thereby influencing the redistribution of water on these landscapes (Wells et al., 1985; Young et al., 2004). Desert pavements are the complex result of cumulic soil development whereby eolian fines are trapped by the surface, moved between peds into the soil, and into ped interiors through lateral conduits (Anderson et al., 2002). Both the surface armoring of the pavement and the low infiltration rate of the vesicular horizon limit the effective moisture in soils capped by desert pavement resulting in abundant quantities of near-surface nitrates and other soluble salts (Wood et al., 2005; Graham et al., 2008).

Coppice dunes form as a result of the trapping of entrained eolian silts and fine sands by shrubs. In the southwestern United States, these shrubs are commonly mesquite (*Prosopis pubescens*), greasewood (*Sarcobatus vermiculatus*), creosote bush (*Larrea tridentata*), or saltbush (*Atriplex* spp.) (Stuart et al., 1971; Rango et al., 2000; McAuliffe et al., 2007). Figure 29.17 illustrates the formation of coppice dunes. These dunes codevelop with the vegetation as the deposition of coarse silt and fine sand increases infiltration rates and enhances plant–water relationships. The dunes also provide ideal sites for burrowing fauna due to the lack
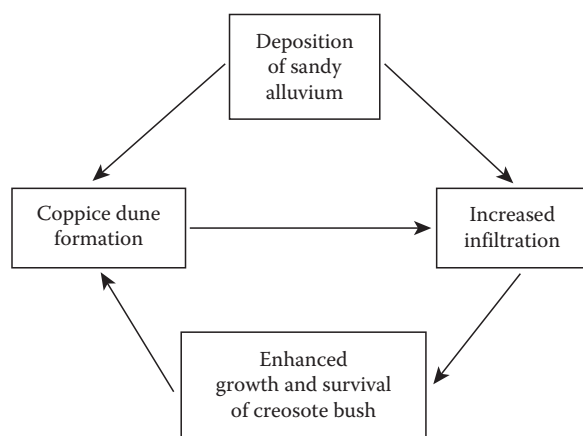
of stones, abundant coarse material, and shading by the shrub. The burrowing activity further increases water infiltration into these areas. As plant-available water is increased in these areas, shrub survival and growth is enhanced so that more sediment is trapped causing the dune to further enlarge.

29.7 Geomorphic Description of Landscapes

29.7.1 Hillslopes

Soil patterns or sequences on hillslopes will in general follow a catenary relationship. The soil and landscape relationship inherent in the catena can be used to predict and describe soil occurrence. A fairly simple set of geometric or morphometric descriptors can be used to define hillslopes and therefore associated soil patterns. These descriptors include slope gradient, slope aspect, slope shape, slope complexity, slope position, and geomorphic component (position).

29.7.2 Slope Gradient

Slope gradient, which is measured along the vertical profile of a slope, is the angle of inclination of the ground surface from the horizontal plane. Commonly, slope gradient is expressed in degrees from the horizontal plane (e.g., 45°) or as a percent, which for a 45° slope is one unit of drop or rise per one unit of distance (slope gradient equals 100%). Slope gradient is a proxy measure of potential energy that drives mass movements and the erosive force of surface runoff on a slope.

Soil map units in a soil survey include a typical range for the slope gradient (e.g., Chemawa loam, 8%–15% slopes; Van Wambeke and Forbes, 1986). Inflection points or changes in slope gradient that repeat on a landscape are readily discernible and generally correspond to differences in the internal structure (underlying lithology), past erosion events (stream incision), or contacts between landforms or sediment bodies (Figure 29.18). Inflection points on slopes, therefore, often mark natural boundaries between soils. Soil surveyors use slope inflections as visible clues to changes in soils on landscapes.

FIGURE 29.17 Diagram illustrating linkages in the codevelopment of creosote bush and coppice dunes in the Mojave Desert. (Redrawn from McAuliffe, J.R., E.P. Hamerlynck, and M.C. Eppes. 2007. Landscape dynamics fostering the development and persistence of long-lived creosote bush (*Larrea tridentata*) clones in the Mojave Desert. J. Arid Environ. 69:96–126.)
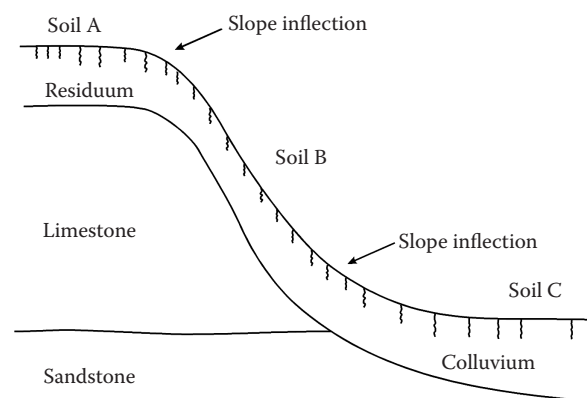
FIGURE 29.18 Diagram of natural landscape boundaries.

29.7.3 Slope Aspect

Aspect is the direction that a slope faces. Slope aspect is usually expressed as a compass azimuth (e.g., 215°) or as a cardinal direction (e.g., SW). Soil microclimates are strongly influenced by the amount of direct solar radiation, which is a function of slope aspect. In the northern hemisphere, north- and east-facing slopes are cooler and moister than south and west facing slopes. Distinct differences in soils occur as a result of aspect (Lotspeich and Smith, 1953; Finney et al., 1962; Franzmeier et al., 1969). The effects of slope aspect are more pronounced in mountainous or high-relief terrain than in low-relief areas. The influence of slope aspect is also more pronounced in temperate latitudes than in equatorial latitudes (Buol et al., 1989).

29.7.4 Slope Shape

Slope shape is the 3D geometry of a slope. The geometric form is obtained by combining the shapes of both the vertical profile (up and downslope) and the elevation contours (across slope). A 2D shape is either linear or curved. If curved, the shape can be convex or concave. This yields nine possible geometric forms to describe all slopes (Figure 29.19).

Slope shape is a property of hillslopes that strongly influences the movement of water both as overland flow and throughflow. For example, a slope that is linear in both vertical profile and contour shape creates parallel, lateral flow (Figure 29.19a). A slope that is convex in both profile and contour causes divergent flow (Figure 29.19d), and a slope concave in both profile and contour causes convergent flow (Figure 29.19h). Slope shape redistributes moisture received by precipitation, creating distinct microenvironments on the landscape. Areas of convergent flow are moister than the local climate whereas areas of divergent flow are dryer than the local climate. The influences of slope shape on water movement and soil moisture on landscapes in turn controls soil formation and vegetation.

In addition to its role in the redistribution of runoff, slope shape is a visual clue to the internal structure of a hillslope. Hillslopes that have a steeply convex, vertical profile are inherently resistant to erosion. Convex slopes occur predominately in landscapes where erosion is controlled by resistant bedrock. Soils on these slopes are shallow and usually display limited horizon development. In contrast, linear or concave slopes generally occur on unconsolidated material or weakly resistant rocks. Soils on these slopes are thicker and display greater horizon development.

Concave slopes or concave portions of slopes denote a decrease in slope gradient. Both potential energy and the velocity of overland flow decease at the inflection of a concave slope. An accumulation of colluvium or sediment derived from slope wash often occurs on concave slopes. Soils on these slopes often form by cumulic processes and usually have thick, poorly developed horizons compared to soils on adjacent uplands.

29.7.5 Slope Complexity

Slope complexity is a simplistic description of the ground surface in respect to the downslope path encountered by overland flow (Figure 29.20). Simple slopes are relatively smooth, with few obstructions to surface flow or decreases in slope gradient and thus nominal opportunities for sediment deposition. Complex slopes contain substantial irregularities in slope conditions.

(a)

(b)
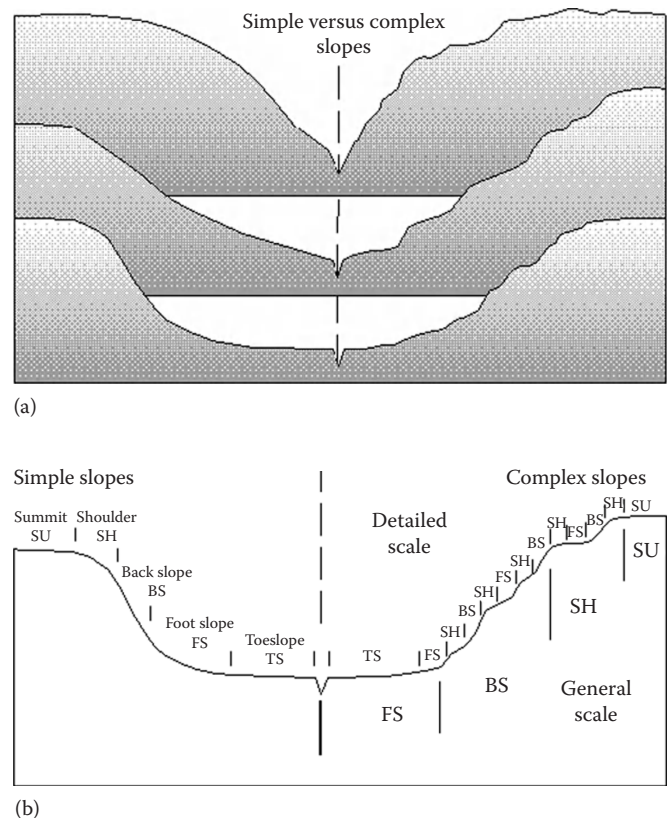
FIGURE 29.20 Two-dimensional diagram of simple versus complex slopes.

FIGURE 29.19 Three-dimensional diagram of the nine hillslopes shapes with flow lines shown: L = linear, V = convex, C = concave. (Schoeneberger and Wysocki, 1997).

Surface water flow is apt to be interrupted and largely nonparallel with considerable changes in flow velocity and subsequent erosion and sediment transport capacity.

29.7.6 Slope Position

Slopes can be divided into segments or elements along a 2D, cross-sectional profile based on slope shape, the degree of erosion or deposition, the presence or absence of sediment, and the nature of the sediment. Wood (1942) identified four segments across a "fully developed slope" for a bedrock-controlled landscape. These elements are the waxing slope, free face, debris slope, and pediment. King (1957) elaborated on the geomorphic processes active on these four slope elements. The waxing slope is the convex crest of a hillslope dominated by chemical weathering rather than erosional removal. The free face is an outcrop of bare bedrock on the upper reach of a hill. Erosion is most active on this element. The debris slope occurs below the free face and is composed of material eroded and transported from the free face. Below the debris slope is the pediment, which is an inclined ramp extending from the hillslope base to an alluvial basin.

Ruhe (1960) modified Wood's 2D hillslope elements and applied them to the study of soil landscapes. Ruhe's hillslope profile elements include the summit, shoulder, backslope, footslope, and toeslope. These elements can be distinguished by inflections in slope gradient and line-segment shape (Figure 29.20). The *summit* is the relatively level, uppermost portion of a hillslope profile. It is the most geomorphically stable and least erosive part of a hillslope. The main vector of water flow is downward through the soil and erosional transport is minimal. Soils in this position display the greatest degree of profile development. The *shoulder* is the convex portion of the hillslope below the summit. The break between summit and shoulder is identified by an increase in slope gradient. The shoulder is subject to a greater degree of erosion, and greater lateral flow of water compared to the summit. Soils tend to be similar to, but thinner than those on the summit, and may appear to be vertically compressed or truncated. The shoulder descends to the steepest and more linear portion of the slope, the *backslope*, where surface runoff and erosional transport are greatest. The vector of water flow may be more lateral than vertical depending on the slope gradient. Soils may reflect inputs of less weathered parent material than soils on the summit and shoulder. On long slopes some degree of lateral sorting may be evident. The backslope descends to the concave portion of a hillslope, the *footslope*. The decrease in slope gradient reduces the carrying capacity of flowing water and increases sediment accumulation. Water flow vectors may be primarily lateral, but the position is characterized by the concentration of water from upslope. The footslope merges downslope with the *toeslope*. The toeslope is predominantly linear or slightly concave. The comparatively low slope gradient and low-lying position at the toeslope allows alluvial processes to dominate, given sufficient surface water from upslope or adjacent streams.

Soils tend to be deep, comparatively moist, and composed of or strongly influenced by alluvial sediments. Toeslope sediments from lower order streams reflect short distance transport and a lower degree of fluvial modification (sorting, rounding, stratification, etc.) than toeslope sediments derived from higher order streams.

29.8 Geomorphic Components

A major focus of soil geomorphology is the movement of water through soil bodies and the way that a landscape sheds or concentrates water (surface flow). This focus is the result of the impact of surficial fluvial processes on major agricultural areas and population centers. Consequently, terms have evolved to describe portions of the earth's surface that share a common location, form, and geomorphic process and the practical emphasis on surficial water movement.

In addition to the 2D hillslope elements (Figures 29.20 and 29.21; Ruhe, 1960), Ruhe refined and popularized geomorphic descriptors for 3D pieces of landforms (Ruhe, 1969, 1975). These area descriptors, called *geomorphic components*, are based in part upon the convergent, linear (or parallel), or divergent nature of overland water flow (Figure 29.19) and associated sediment transport. The most widely used suite of geomorphic components was developed for and is most appropriately applied to hills. Historically, this set of terms has been applied to most landscapes, including mountains, due more to the lack of alternatives than to their utility. Other geomorphic settings (mountains, terraces, flat plains) have unique dynamics or complexity and warrant different area descriptors.

29.8.1 Geomorphic Components: Hills

The *interfluve* is the uppermost area of a hill, and represents the oldest, most stable part of the landscape, typically with the most developed soils. A noted exception is where opposing hillslopes have narrowed an interfluve or merged (e.g., a saddle or crest) to the extent that erosion begins to lower the crest. Hillsides are areas characterized by active backwearing (erosion) in general, and at the heads of streams in particular.

Hillsides can be divided into discrete parts (Walker and Ruhe, 1968) based on the dominant behavior of overland water
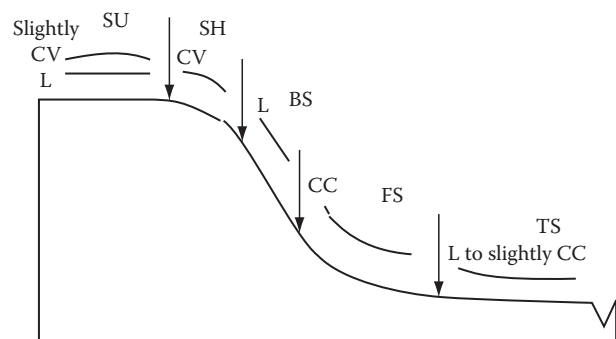
FIGURE 29.21 Hillslope profile terms and associated slope shapes.

flow (Figure 29.8a): converging (*head slope*), linear or parallel (*side slope*), or diverging flow (*nose slope*). Soils on hillsides are dominated by colluvial sediments in gradual transport down the slope and modified to varying degrees by slope wash (nonchannel, overland flow). Unless masked by changes in parent materials, some degree of lateral sorting downslope is usually present. Soil profiles on hillsides can range from thin to thick depending on the rate of erosion (high or low, respectively), the magnitude of the slope gradient, and the extent to which the bedrock is resistant to weathering and erosion. In many landscapes, the traditional assumption of a prevalence of shallow soils over hard rock on hillsides and mountainsides is erroneous (Knox, 1982; Graham, 1986). The *base slope* is commonly an apron or wedge of colluvium at the bottom of hillslopes (Schoeneberger and Wysocki, 1996). This drape of transported material can range from coarse debris to finer sediments that have been winnowed or sorted by slope-wash processes. The base slope does not typically include sorted and stratified alluvium associated with channel deposition. Distal base slope sediments commonly grade into or interfinger with alluvial fills.

29.8.2 Geomorphic Components: Mountains

Mountains represent a unique geomorphic setting due to their scale and slope complexity. Mountainsides commonly have long, complex backslopes up to thousands of meters long, steep slope gradients, highly diverse sediment mantles and complex, near-surface hydrology. Mass movement processes and features are more prevalent than in hills. Consequently, new area descriptors have been developed (Schoeneberger and Wysocki, 1996; Hirmas, 2008) that effectively identify and name geomorphic components for mountains.

The *mountaintop* (Figure 29.22) is the summit or crest of a mountain and is commonly characterized by comparatively short, simple slopes composed of bare rock, residuum, or short-transport (angular) colluvial sediments. In humid climates, soils on mountaintops can be quite thick (Oliver et al., 1997), whereas in arid climates, these soils may be capped by well-developed desert pavements and accumulate considerable quantities of carbonate, nitrate, and other soluble salts (Graham et al., 2008). The side of a mountain, the *mountainflank*, is characterized by comparatively long, complex slopes dominated by mantles of long transport (subangular) colluvium. Residuum, if present, is usually buried by 1–2 m of colluvium. Incised drainageways, structural benches, and mass movement features can be common. Rock outcrops, while visually prominent, are not a major portion of the land surface. The *mountainbase* is an apron of colluvium at the bottom of a mountain slope and analogous to the *base slope* in hills. It is marked by a substantial decrease in slope gradient compared to the mountainflank. The mountainbase is characterized by a thick mantle or wedge of colluvium and commonly contains a comparatively high percentage of coarse rock fragments. The colluvium can extend out onto more level land surfaces, and ultimately interfingers with or is buried by alluvium, or thins and joins re-emergent residuum. In desert environments, windward sides of these mountainflanks and mountainbases can trap significant amounts of dust within which, thick soils may develop (Blank et al., 1996; Hirmas, 2008). Hirmas (2008) described a feedback process in these landscapes where sediment is added to lower-lying landforms through alluvial and debris-flow processes while, concurrently, material entrained from these landforms is carried back to the mountains by wind to serve as parent material for mountain soils (Figure 29.23).
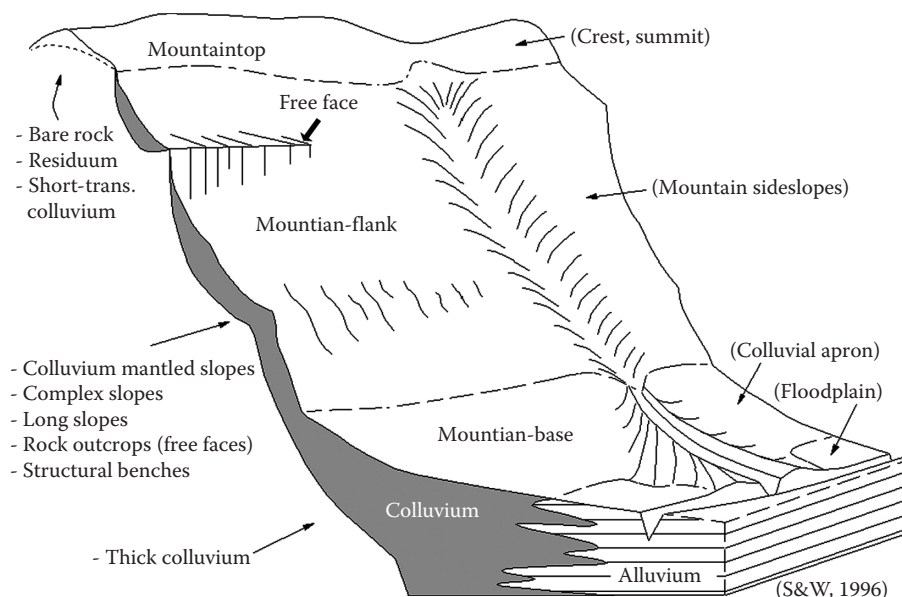
FIGURE 29.22 Geomorphic components for mountain landscapes. (From Schoeneberger, P.J., and D.A. Wysocki. 1996. Geomorphic descriptors for landforms and geomorphic components: Effective models, weaknesses and gaps. (Abstract). American Society of Agronomy, Annual Meetings. ASA, Indianapolis, IN.)
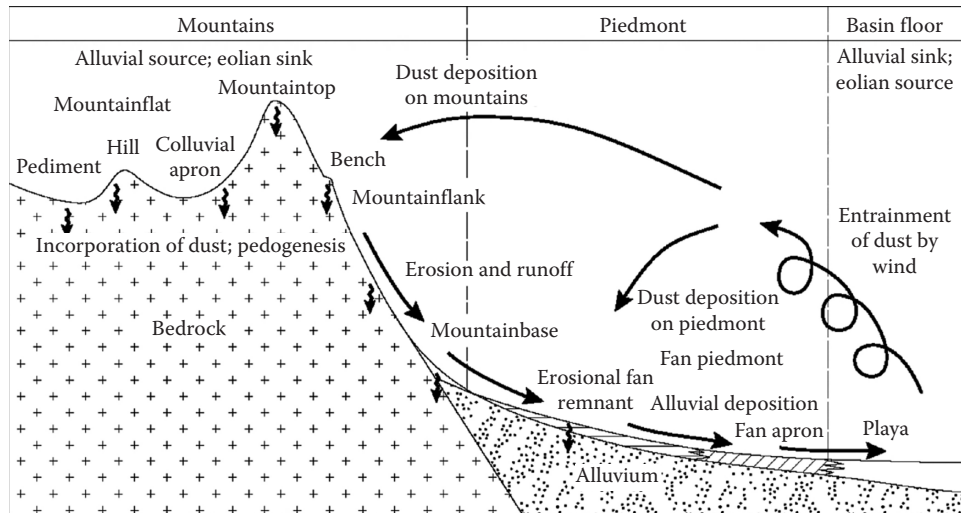
FIGURE 29.23 Idealized geomorphic model of desert mountain–piedmont–basin floor interactions. (After Hirmas, D.R. 2008. Surface processes, pedology, and soil-landscape modeling of the southern Fry Mountain Bolson, Mojave Desert, California. Ph.D. Dissertation. University of California. Riverside, CA.)

29.8.3 Geomorphic Components: Terraces

Terraces are landforms that form a relatively level or gently inclined surface, a constructional strip, or plain that borders a stream, lake, or sea. Terraces are a unique geomorphic setting and have dynamics quite different from those of hills or mountains. Stream terraces and floodplain steps (Figure 29.24) are originally developed by alluvial processes and sediments, rather than the dynamic slope processes and colluvial sediments that characterize hills and mountains or the localized, low energy processes typical of flat plains.

The *tread* is the comparatively broad, generally level part of a terrace or floodplain step. Treads can extend laterally for many kilometers (Gamble, 1993; Saucier, 1994). Treads are level or gently inclined (low gradients) and underlain by alluvial, lacustrine, or marine sediments. Strath terraces are similar in form to other stream terraces, but are erosional landforms characterized by thin alluvial sediments over an eroded bedrock bench or platform.

The *riser* is an escarpment that separates terrace or floodplain levels. The riser commonly consists of a short, steep, planar slope cut into the sediments that underlie the adjacent tread. The areal extent of a riser rarely exceeds tens of meters and therefore is typically depicted on a soil map with a spot symbol rather than a delineation. Geomorphically, risers represent an abrupt change to a lower hydrologic base level, which suppresses the water table along the edge of the adjacent, higher surface. This directly affects soil processes and subsequent soil geography. Soils above and adjacent to a riser tend to be better drained than those farther away from the escarpment (Figure 29.25). Daniels and Gamble, 1967 called this relationship the "red edge effect." Soils at the base of a riser tend to be as wet as or wetter than those farther from the scarp. This is especially true in floodplains.
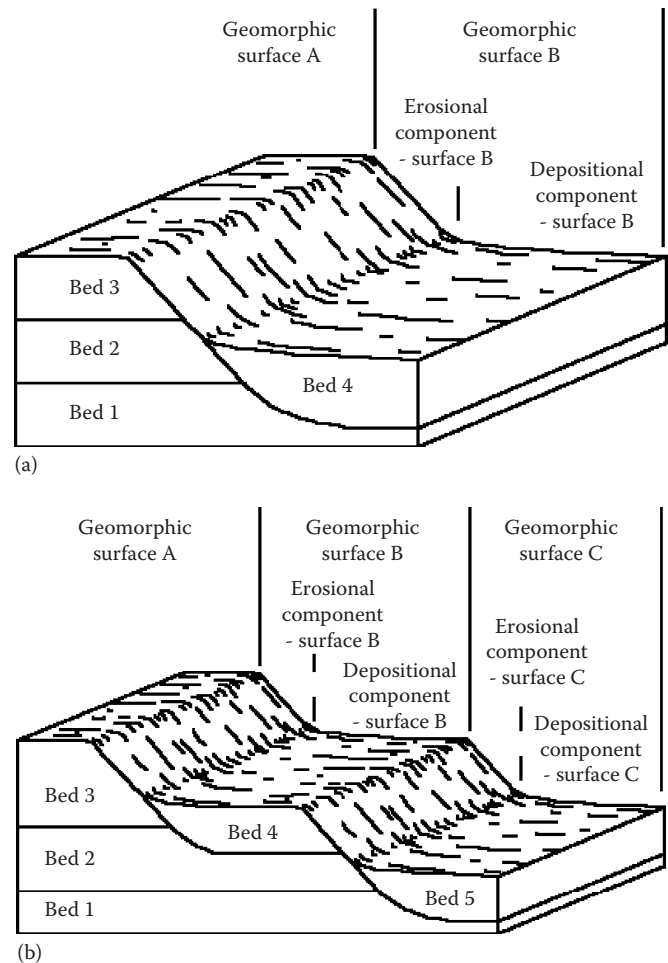
FIGURE 29.24 Geomorphic components for terraces and floodplains.
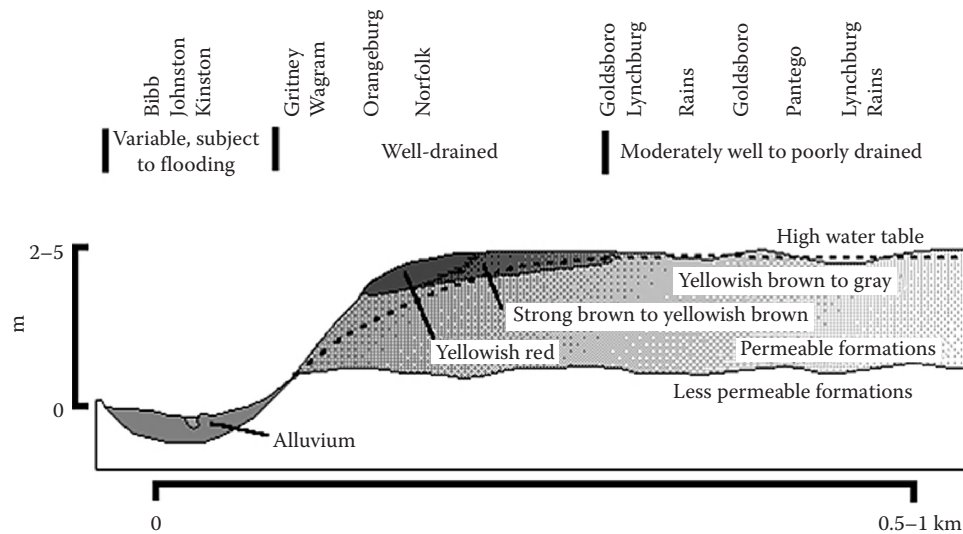
FIGURE 29.25 Two-dimensional landscape diagram showing drainage relationships on low-relief divides.

29.8.4 Stream Terraces versus Floodplain Steps

Fluvial landforms include the active channel, surfaces that are intermittently inundated, and higher areas that no longer flood or receive alluvial sediments. The distinction between stream terraces and floodplain steps is an arbitrary break in a natural continuum. There are practical pedological reasons for separating areas that are within the active floodplain from areas that flood or receive sediments in rare instances. For most land uses (e.g., home site), it is valuable to know the relative occurrence of flooding. Terraces are defined as a part of the fluvial system that no longer actively experiences fluvial modification (Ruhe, 1975), that is, no longer aggrading (receiving additional alluvial sediments), no longer experience significant flooding. Floodplain steps are morphometrically similar to terraces, except that they occur within the fluvially active portion of a floodplain and are subject to relatively frequent modification (experience relatively regular or significant flooding and alluvial sediment inputs). Pedologically, terraces commonly have more extensively developed soils (e.g., alfisols, ultisols) compared to lesser-developed soils (e.g., inceptisols, entisols) of floodplain steps. The practical separation of terraces from floodplain steps varies. Stream terraces that have not flooded during recorded history could be inundated during rare, catastrophic events (e.g., 500 year flood). Soil development on terraces is usually controlled more by moisture derived from precipitation than from the adjacent river. Terrace soils in cold or dry climates, despite a greater age and lack of new sediment, may closely resemble soils on adjacent floodplain steps that periodically receive fresh sediment. Conversely, soils on floodplain steps in hot, humid settings may exhibit extensive development and resemble soils on the older, adjacent terraces. In temperate North America, a practical separation between terraces and floodplain steps (Figure 29.24) is the 100 year flood stage (Soil Survey Staff, 1998).

29.8.5 Geomorphic Components: Flat Plains

Broad, flat plains such as proglacial lake plains, low-lying coastal plains, and low-gradient glacial till plains are a geomorphic setting distinct from hills, mountains, or terraces. The primary differences are geomorphic origin (nonfluvial) and the low potential energy to drive water flow. Historically, generic terms (e.g., flat, broad interstream divide, rise) or terms developed for other settings have been applied to flat plains with unsatisfactory results. Recent interest and recognized value of wetlands and hydric soils (e.g., Mausbach and Richardson, 1994) highlights the need for unique descriptors for flat areas. Small differences in elevation (e.g., 15 cm or less) on a landscape can mark the change from moderately well drained to very poorly drained soils (Richardson, 1997). The search is presently on for better, unique descriptors to express the geomorphic components of flat plains. The following terms are proposed as provisional contenders.

Flat plains are subdued landscapes dominated by broad areas with low slope gradients (e.g., 0%–1%) called *talfs* (Figure 29.26), and closed depressions, with only nominal changes in local relief (i.e., microhighs, microlows). A *rise* is a slightly elevated area (i.e., microhigh), which tends to be broad with low slope gradients (e.g., 1%–3%). The low gradients of flat plains result in minimal erosion by running water, especially if vegetated. Consequently, fluvial drainage networks tend to be poorly developed: nonintegrated, incipient, or deranged. Precipitation tends to pond locally and lateral transport is slow both above and below ground. These conditions favor accumulation of organic matter (e.g., pocosins, upland bogs) and retention of water and fine earth (<2 mm diameter) sediments.

29.8.6 Microrelief

As used here, microrelief refers to slight variations in the height of a land surface that are too small to delineate on a
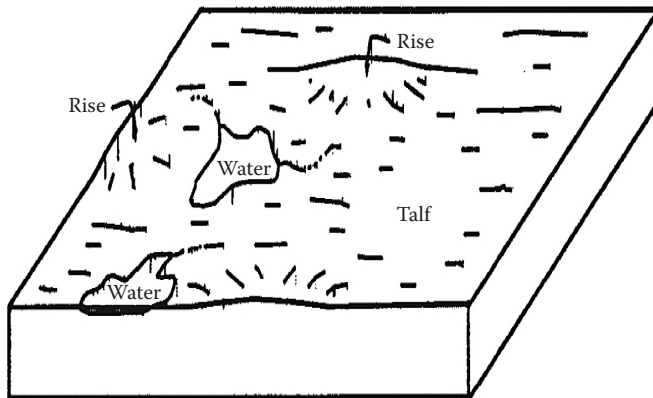
- Very low gradients (e.g., slope 0%–1%)

- Deranged, non integrated, or incipient drainage network

- "High areas" are broad and low (e.g., slope 1%–3%)

- Sediments commonly lacustrine, alluvial, or till

FIGURE 29.26   Geomorphic components for flat low-relief plains.

topographic or soils map at commonly used scales (e.g., 1:24,000 and 1:10,000). Examples of microrelief include microhighs and microlows. These minor elevational differences can have a surprisingly large impact on vegetation distribution, soil hydrologic dynamics (Hopkins, 1996), and sedimentation patterns. This approach to microrelief is strictly morphometric. We make a distinction between small elevational differences (*microrelief*) and small features (*microfeature*) with which these elevational differences are associated. An example is gilgai (a kind of microfeature), with the associated microrelief (mound areas = microhighs, bowl areas = microlows). It is an unfortunate yet common practice to confuse the two items and describe both as "microrelief."

29.9 Landscapes, Landforms, Microfeatures, and Anthropogenic Features

Armed with the ability to describe the most detailed morphometric nuances of the land surface, one is inevitably faced with some variation of the question: "Taken all together, what is this feature that we see and what should it be called?" A potpourri of names has been applied over time that connotes internal composition, form, arrangement, collective relationships, and origin. These terms range in scale from small, human scale features to continental-scale assemblages. Confronted with such complexity, some organization of terminology is reasonable and in fact, necessary. Various schema have been developed to achieve this, typically from a particular geographical perspective: Developed schema include physiography (Fenneman, 1931, 1938, 1946; Thornbury, 1965, 1969; Hunt, 1967), landuse (USDA-SCS, 1981), and ecology (Omernik, 1987, 1995; Bailey et al., 1994; USDA-EPA, 1996). A different approach is to de-emphasize the geographical context and focus on geomorphic

process (Peterson, 1981). One pseudohierarchy of terms has been assembled specifically from and for soil survey and geographic applications (Schoeneberger and Wysocki, 1997). In addition to morphometric terminology and physiographic location, this system loosely arrays land surface features in a progression of scale in the following fashion.

A *microfeature* is a small, local, natural form (feature) on the land surface that is too small to delineate on a topographic or soils map at commonly used map scales (e.g., 1:24,000 to 1:10,000). *Note*: The conventional use of microrelief commonly encompasses some of the terms or features that in this system are contained within microfeature (see below).

A *landform* is any physical, recognizable form or feature on the earth's surface, having a characteristic shape and range in composition, and produced by natural causes; it can span a wide range in size (e.g., *dune* encompass both *parabolic dune*, which can be several tens-of-meters across, as well as *seif dune*, which can be up to 100 km long). Landforms provide an empirical description of similar portions of the earth's surface.

A *landscape* is a collection of spatially related, natural landforms, usually the collective land surface that the eye can comprehend in a single view (Soil Survey Staff, 1998).

An *anthropogenic feature* is an artificial feature on the land surface, having a characteristic shape and range in composition, composed of unconsolidated earthy or organic materials, artificial materials, or bedrock, that is the direct result of human manipulation or activities. Historically, landforms and the like have been defined as, and therefore restricted to, natural features. There are relatively consistent formational processes with common compositional, structural, stratigraphic results within a population of natural features (e.g., sand dunes). Human-made features may have a common theme in intent (e.g., water impoundment), but can have an almost limitless variety of formational processes, material composition, and internal structure. This simplifies the study and interpretation of the land's surface but ignores human-made features, which progressively apply to ever more of the earth.

29.10 Age Assessment of Soil Landscapes

Age assessment of landscapes is the single most important contribution that geomorphology makes to the understanding of soils. The law of superposition and the concept of geomorphic surfaces are the key principles for defining relative age relationships in a landscape (Daniels et al., 1971; Hall, 1983). Radiometric, isotopic, or paleontological dating combined with field studies using geomorphic and stratigraphic principles can establish the absolute age of a deposit, soil, or a landscape.

The longer a landscape is exposed to subaerial weathering, the greater the potential for soil development. The terms "weathered," "well developed," and "old" are relative indicators of time in soil formation, but none are accurate descriptors of age. *Weathered* refers to the relative stability stage (Goldich, 1938;

Jackson, 1968) of the minerals contained in a soil. The *weathering stage* is a function of age, weathering intensity, and mineral stage of the parent material. A chronologically young soil may be composed of highly weathered minerals if the parent material is preweathered.

Development is a relative measure of soil formation based on the type and degree of horizonation. Soil development is a function of both age and the rate of horizon formation. Rate of horizon formation depends on the intensity of soil-forming processes (see Chapter 30) and composition and resistance of the parent material to change. Soils of similar morphology can differ substantially in age. For example, Burges and Drover (1953) reported that spodic horizons in Australia formed in sandy beach deposits within 1000–3000 years. Franzmeier and Whiteside (1963) found that spodic horizons took 8000 years to form on dunes in northern Michigan. Holzhey et al. (1975) suggested 21,000–28,000 years for development of thick spodic horizons in North Carolina. Well-developed, weathered soils are described as *old*. In this context, *old* is a proxy for age based on weathering and degree of soil formation. Geomorphic principles are a means to assess soil and landscape age independent of horizon development or weathering stage.

The law of superposition specifies the sequence of deposition of sedimentary rocks and unconsolidated sedimentary deposits. Sediment can only be deposited atop preexisting materials. Younger beds or strata invariably overlie older beds, if not overturned. Bed 2 (Figure 29.27a) is younger than Bed 1. Once deposition ceases, the top of Bed 2 represents a depositional surface (surface A) equal in age to the youngest strata in the bed. Stabilization of the surface allows a soil to form in the sediment (Figure 29.27b). This is time zero for subaerial weathering and soil formation. The surface and soil age are the same. The passage of time is recorded by an increasing degree of soil development. Soil formation requires a period of nondeposition and therefore indicates an unconformity in the sedimentary sequence. A buried

soil represents an unconformity in the sedimentary sequence. For example, deposition of Bed 3 buries the older units and the soil formed on them (Figure 29.27c). Bed 3 is the youngest deposit. The top of Bed 3 is a new, younger depositional surface B. The stratigraphic record contains three depositional beds and a buried soil.

29.10.1 Geomorphic Surfaces

In a landscape, an erosion cycle or event (e.g., stream incision, hillslope retreat, pedimentation) that cuts existing beds creates a geomorphic surface of erosional origin (Figure 29.28). This erosion surface is younger than any strata, soil, or surface that it cuts. An erosional surface can only cut preexisting materials, strata, or surfaces. As you ascend a landscape geomorphic surfaces increase in age. This is the principle of ascendancy. Surface E1 is younger than surface A and younger than beds 1, 2, and 3. The cutting of an erosional surface produces sediment that is deposited downslope or down valley. The erosional surface descends or grades to a corresponding depositional surface and the underlying alluvium (Figure 29.28). The depositional

(a)

(b)
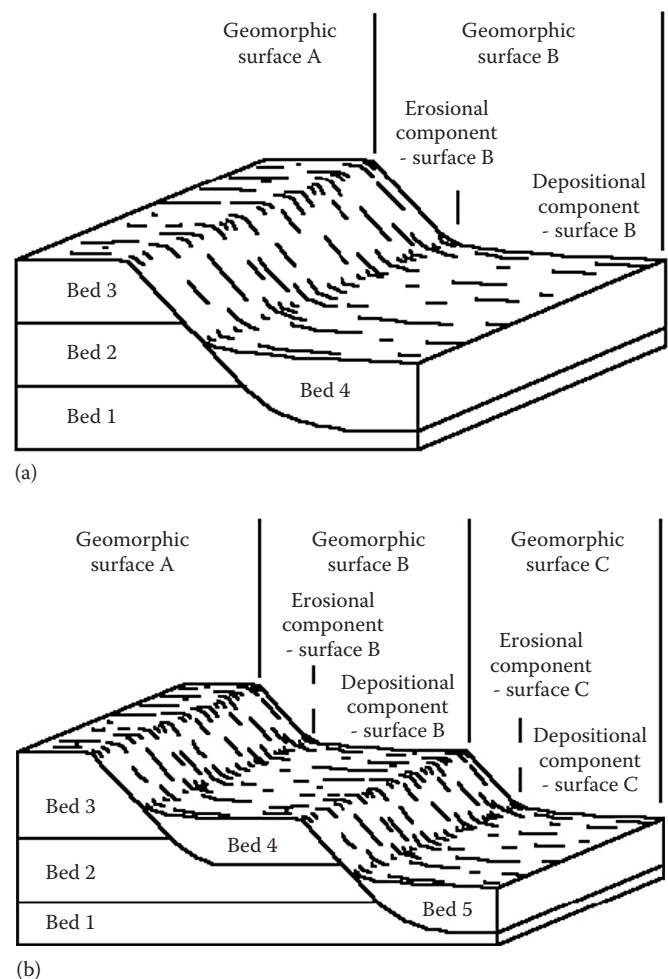
FIGURE 29.28 Three-dimensional block diagram showing geomorphic surfaces and associated time relationships.

(a)   (b)

(c)

FIGURE 29.27 Three-dimensional block diagram depicting the law of superposition.
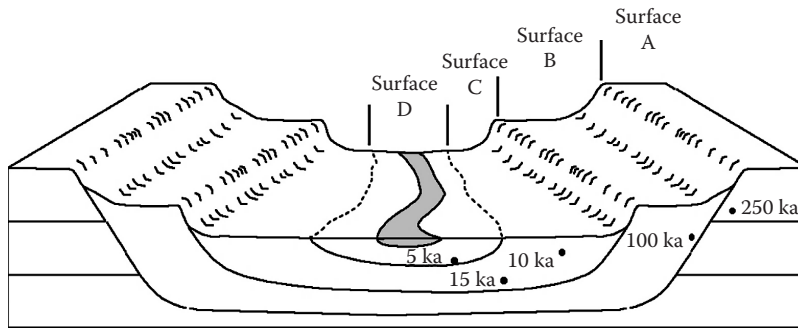
FIGURE 29.29   Three-dimensional block diagram showing geomorphic surfaces and dating of sediments.

and erosional surfaces and the top of the alluvium are all the same age. An erosional surface is the same age as the alluvium to which it descends. This is the principle of descendancy. Relative age of an erosional surface in the field can be determined by observing what it slopes downward to in a smooth, concave profile (Daniels et al., 1971).

29.10.2 Stepped Surfaces

Landscape evolution commonly proceeds episodically through time, which creates a series of stepped or faceted geomorphic surfaces along interfluves (Ruhe et al., 1968; Gamble, 1993; Figure 29.29). The age relationships of stepped surfaces follow the principles of ascendancy and descendancy. Surface A on the interfluve summit is the oldest, surfaces become successively younger as you descend the landscape. The age differences between surfaces depend on the timing of the erosional processes that produced them.

29.10.3 Absolute Age Assessment

Dating techniques along with the geomorphic relationships provide a framework for establishing a landscape chronology. For example, ages shown in Figure 29.29, which could be obtained by C14, luminescence techniques, or fossil analysis, establish the periods for deposition. Surface D is underlain by sediment with an age of 5 ka at the base. This unit is inset into an alluvium that has an age of 15 ka at the base and 10 ka near the top. Stream incision and back filling therefore occurred between 5 and 10 ka that deposited the alluvium under surface D. Surface C includes both an erosional depositional element. Cutting of the erosional surface occurred between 10 and 15 ka based on the alluvium to which it descends. The alluvium is a time transgressive deposit as is the cutting of the erosional surface. Alluvium under surface B has an age of 100 ka and the erosional element ascends to a surface of 250 ka. The erosional processes that created surface B began later than 250 ka and ended by 100 ka. Given this scenario, the absolute age differences between the surfaces and the soils on them can be established.

29.10.4 Geomorphic History

The concepts as displayed in Figures 29.27 through 29.29 are straightforward. Application to soil landscapes requires detailed stratigraphic control obtainable only by field observations from natural (e.g., stream banks) or man-made (e.g., road embankments) exposures or drill cores and the ability to think in three dimensions. The geomorphic history of a region must sometimes be established through detailed stratigraphic investigation before soil patterns are completely understood (Ruhe et al., 1967; Daniels et al., 1970).

Consider the geomorphic history of a hypothetical landscape in Figure 29.30. A landscape is composed of soil s1 on geomorphic surface A, which is formed in a single deposit (Figure 29.30a). The deposit is buried by a younger material (e.g., alluvium, glacial till, loess) with subsequent development of soil s2 on surface B (Figure 29.30b). The landscape now consists of two stacked stratigraphic units separated by a buried soil s1. Stream incision cuts an erosion surface across both units (Figure 29.30c). The landscape becomes a two stepped sequence comprising surfaces B and C. Soil s3 forms in surface C (Figure 29.30d). Two soils s2 and s3 of different age make up the landscape. Valley alluviation buries surface C and reduces relief in the landscape (Figure 29.30e). Soil s4 forms in surface D on the alluvium (Figure 29.30f). The soil landscape at this point consists of two soils s2 and s4, which again differ considerably in age. Stream incision and back filling inset a deposit into the preexisting alluvium (Figure 29.30g). The inset is denoted by surface E and soil s5. The landscape now includes three surface soils s2, s4, and s5, all of different age. Soil distribution on the landscape, soil age, and the geomorphic history of this landscape can only be established from multiple field observations and drill cores at critical locations that identify all the stratigraphic units and buried soils.

29.10.5 Soils and Geomorphic Surfaces

Despite the strong association between soils and geomorphic surfaces, soil patterns and morphology cannot be used to initially define or recognize geomorphic surfaces. Geomorphic surfaces must first be defined using geomorphic principles
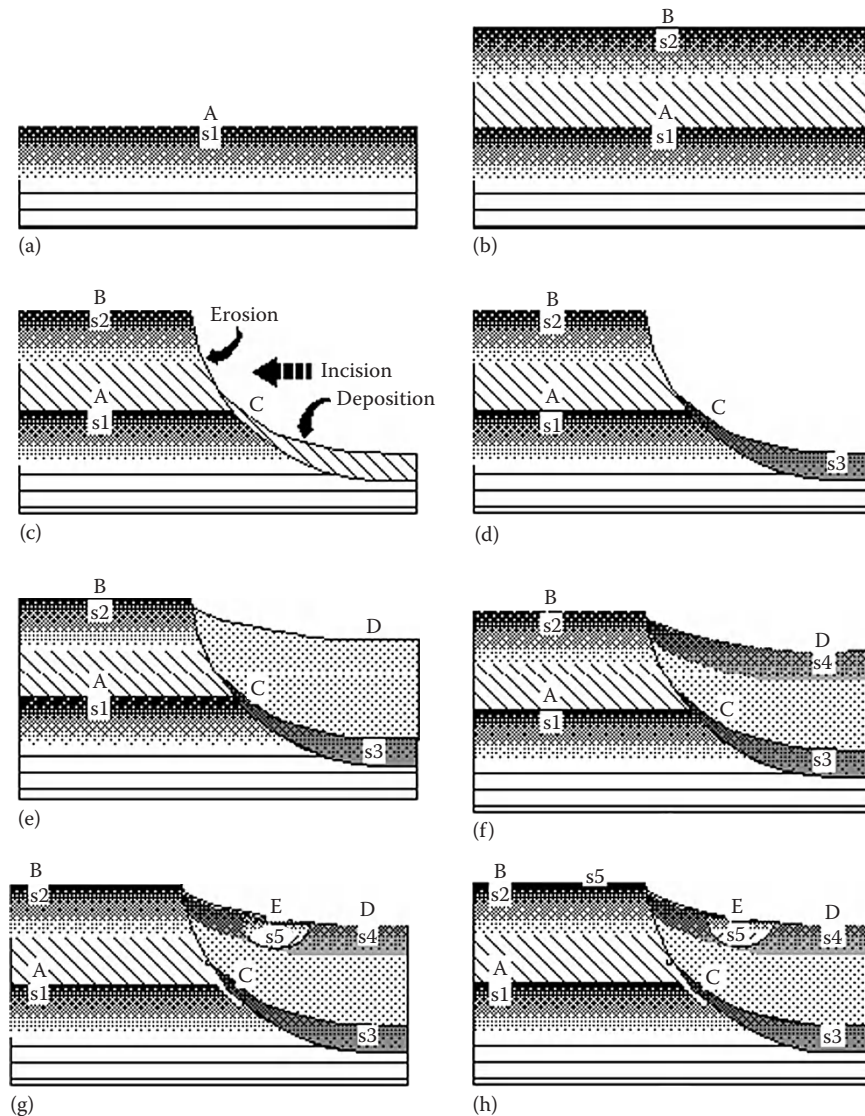
FIGURE 29.30 Two-dimensional diagram explaining hypothetical landscape evolution.

and then the linkages made to soil patterns. A soil or suites of soils may occur on more than one geomorphic surface and soil boundaries do not always correspond to geomorphic boundaries on a landscape. Consider Figure 29.31, which shows a stratigraphic sequence and soil horizons cut by an erosional surface. Soils 2, 3, and 4 on erosional surface B form over beds of varying composition and age and from preexisting soil horizons. Soils 2, 3, and 4, however, are all the same age (the age of surface B). In this case, three different soils occur on the erosional surface. The sequence of soils is determined by the stratigraphic units present and erosional sorting along the surface. The same erosional surface across its geographic occurrence can cut different strata or sediments. The soils in that case would differ.

The boundary between soil 1 and soil 2 (Figure 29.31c) lies below the slope break that denotes the erosional surface. Erosion has not removed enough of the original profile to cause a change

in the soil. The boundary between the surfaces and the boundary between soils are not coincident. A soil map and a geomorphic surface of the same landscape would not have coincident boundaries.

29.11 Paleosols, Geosols, and Climate Interpretation

29.11.1 Paleosols

Geologic processes that create and destroy soils and landscapes vary in time and space. Continental-scale processes like glaciation can create or destroy entire landscapes. More commonly landscapes evolve by erosion and deposition on only parts of them. Erosion removes soils and sediments from the active parts of landscapes, while soils continue to form on the stable parts. Sediment from
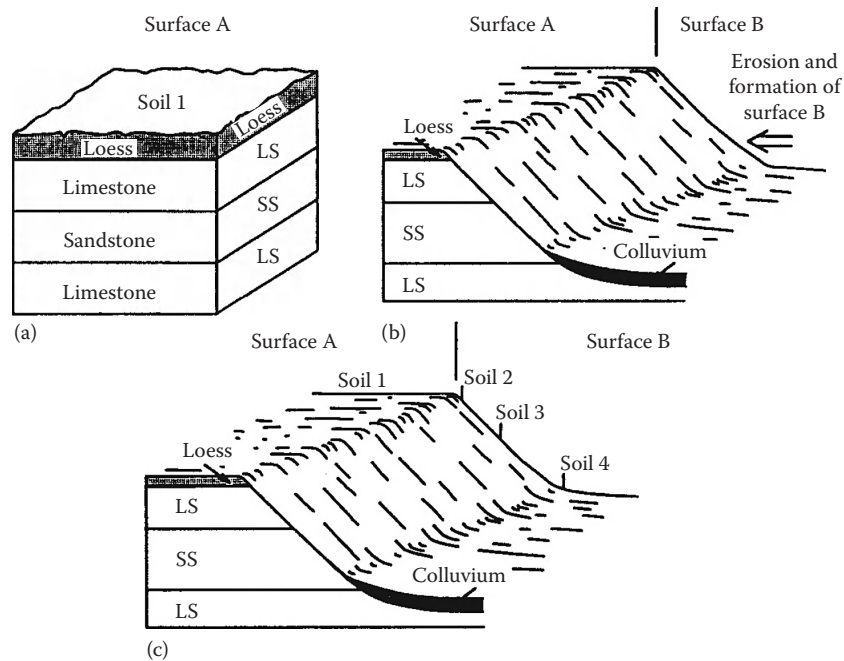
FIGURE 29.31   Two-dimensional diagram relating soils and geomorphic surfaces.

erosion buries preexisting surfaces, soils, and sediments. Buried soils and soils that endure on stable landscapes for long periods become part of the geologic record and are valuable for interpreting earth history. *Paleosols* are soils that formed on landscapes of the past (Yaalon, 1971; Valentine and Dalrymple, 1976). Age alone, however, does not denote a paleosol. The process or processes responsible for the soil morphology must no longer operate due to a change in climate, local environment, or because of burial.

Ruhe (1965) recognized three types of paleosols buried, exhumed, and relict. *Buried paleosols* form at the land surface, but later are covered by sediment, which removes the paleosol from the dominant soil-forming zone. Burial must be deep enough to suspend the soil-forming processes, and rapid enough to prevent formation of a cumulic soil profile. Postburial and diagenetic changes in soil properties and horizons are common (Yaalon, 1971; Olson and Nettleton, 1999) and must be considered when studying paleosol composition.

Exhumed paleosols form at the land surface, are buried, and later re-exposed on younger landscapes by erosion of the covering sediment (Ruhe, 1969). Erosion exposes both the paleosol and the former land surface. Exhumed paleosols rarely contain completely intact paleosolums. Former A horizons are usually truncated or mixed with the burial sediment and difficult to recognize. In contrast, B horizons are the best preserved and most easily discerned feature of exhumed paleosols. An exhumed paleosol differs from a buried one in that the outcrop can be traced across a significant area of the present land surface. An exhumed paleosol can be out of balance with both the existing climate and the adjacent soils on the landscape. Landscapes that include exhumed paleosols may not conform to the expected catena relationship in a geographic area (Figure 29.32; Ruhe, 1969).
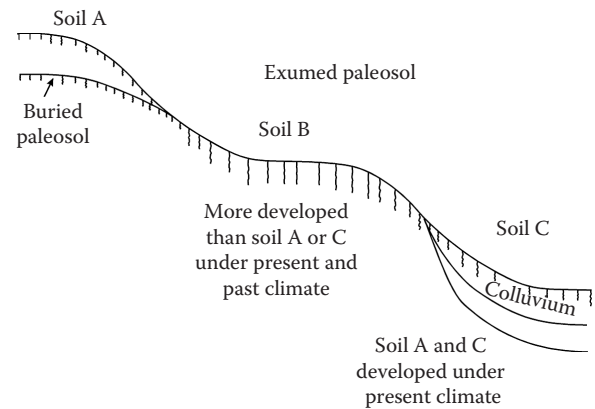
FIGURE 29.32   Landscape diagram showing buried and exhumed, paleosols.

Relict paleosols are soils formed on preexisting landscapes and were never buried (Ruhe, 1965). Therefore, identification of relict paleosols is somewhat problematic. A relict paleosol must have endured one or more shifts in regional climate or local environment (e.g., lowering of a water table by stream incision or baselevel change) such that it is no longer is in balance with the existing conditions. This concept requires that the age of the landscape be known or inferred and that we have a precise understanding of the processes responsible for soil morphology (see Chapter 30 for discussion). Soils on old, easily distinguishable geomorphic surfaces such as terraces (Figure 29.33) are the most commonly recognized relict paleosols.

Our present knowledge of soil processes is inadequate to determine relict versus nonrelict soil conditions in all situations.
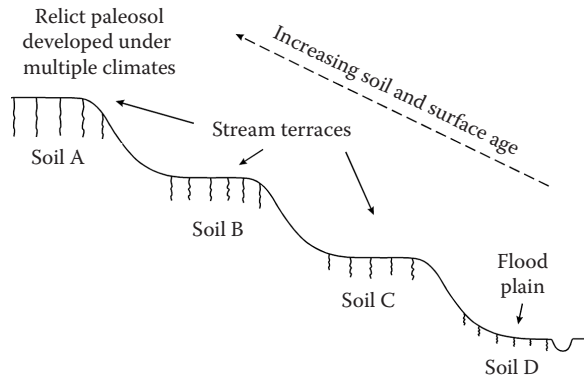
FIGURE 29.33 Landscape diagram depicting relict paleosol.

Soil-forming processes can be reversible self-terminating, steady state or metastable, or irreversible self-terminating (Yaalon, 1971). This means that all morphological features do not have equal utility as indicators of climate change. For example, the formation of silica-cemented soil horizons, or duripans, occurs in arid or semiarid climates (Soil Survey Staff, 1996). Once formed, a duripan is resistant to destruction even in humid conditions. Thus, the presence of a duripan in a humid climate is strong evidence of a relict condition. In contrast, some morphological features such as A horizon thickness and color, which have a strong correlation to present climatic regimes, can adjust rapidly to a shift in climate. Despite the difficulties in defining and recognizing relict paleosols, this concept is valuable because it focuses on the relationships of landscape age and soil processes.

29.11.2 Geosols

Buried paleosols are integral components of the stratigraphic record. Formation of a soil requires a period of landscape stability with exposure to subaerial weathering. A buried paleosol, therefore, (1) represents an unconformity in a sedimentary sequence and (2) marks a consistent stratigraphic position at the top of a sedimentary body or bodies. Buried paleosols that are laterally traceable with 3D form and have mappable distribution at a scale of 1:24,000 can be formally defined as pedostratigraphic units (Morrison, 1967) called *geosols* (NACSN, 1983). Geosols include only pedogenic horizons formed in preexisting rock or sediment bodies. Buried O horizons and some C horizons of paleosols are excluded from geosols. A geosol may form in parent materials of diverse ages and composition. Formal naming and establishing of geosols follows the same criteria of other stratigraphic units (NACSN, 1983). A geosol is named after a geographic feature in the type area.

29.11.3 Climate Change and Paleosols

Jenny's (1941) model can be solved for an individual soil-forming factor. For example, the equation

$$\text{Soil} = f'(c)_{p,r,b,t} \qquad (29.1)$$

ascertains the climate during soil formation, where the subscripts denote a constancy in the factors of parent material, relief, biota, and time. There are problems with this approach, however, Equation 29.1 defines the soil-forming factors as independent variables, but in reality the factors covary. It is also difficult to assign quantitative values to some factors, particularly parent material and the biota. Nonetheless, this approach allows useful, semiquantitative comparisons between soil morphology and individual soil-forming factors.

Processes of soil formation occur over time scales of tens to thousands of years. Soil morphology forms a physical record of the climate, vegetation, and/or environment during the time of formation. Recent emphasis on global climate change has renewed interest in paleosols as indicators of past climates. Numerous studies have used soil morphology to interpret past climates or environments (Busacca, 1989; Driese et al., 2005; Retallack, 2005; Brock and Buck, 2009). The interpretation of climate based on paleosols requires one or more of the following: (1) that the paleosol existed on a stable landscape and had reached a steady state within a particular climate, (2) identification of the ancient soil environment (e.g., floodplain), and/or (3) identification of the catenary relationship or position of the paleosol on the ancient landscape.

For example, a climatic interpretation based solely on the properties of the Clarion profile (Figure 29.6), such as A horizon thickness, organic carbon content, depth to inorganic carbonate, and B horizon color are indicative of a well drained, prairie soil formed in the present humid climate of central Iowa. If only the Harps soil (Figure 29.6) is considered, which has a thicker, darker A horizon, a gray B horizon, and inorganic carbonates are at or near the surface, a climate more humid than the Clarion morphology is indicated. Carbonates near the surface suggest a climate more arid than Clarion and the A horizon thickness suggests a cooler climate. Despite occurring only tens of meters from Clarion soils on the landscape, the morphology of Harps soils leads to an ambiguous and a potentially erroneous interpretation of the regional climate. Climatic interpretations based on paleosol morphology must include paleocatenary relationships.

References

Anderson, K., S. Wells, and R. Graham. 2002. Pedogenesis of vesicular horizons, Cima volcanic field, Mojave Desert, California. Soil Sci. Soc. Am. J. 66:878–887.

Arndt, J.L., and J.L. Richardson. 1989. Geochemistry of hydric soil salinity in a recharge-throughflow-discharge prairie-pothole wetland system. Soil Sci. Soc. Am. J. 53:848–855.

Bailey, R.G., P.E. Avers, T. King, and W.H. McNab. 1994. Ecoregions and subregions of the United States (map) (supplementary table of map unit descriptions compiled and edited by McNab, W.H., and Bailey, R.G.): Scale 1:7,500,000. USDA Forest Service, Washington, DC.

Blank, R.R., J.A. Young, and T. Lugaski. 1996. Pedogenesis on talus slopes, the Buckskin range, Nevada, USA. Geoderma 71:121–142.

Blume, H. 1968. Die pedogenetische Deutung einer Catena durch die Untersuchung der Bodendynamik. Trans. 9th Int. Congr. Soil Sci. 4:441–449.

Blume, H.P., and E. Schlichting. 1965. The relationship between historical and experimental pedology. *In* E.G. Hallsworth and D.V. Crawford (eds.) Experimental pedology. Butterworths, London, U.K.

Brock, A.L., and B.J. Buck. 2009. Polygenetic development of the Mormon Mesa, NV petrocalcic horizons: Geomorphic and paleoenvironmental interpretations. Catena 77:65–75.

Buol, S.W., F.D. Hole, and R.J. McCracken. 1989. Soil genesis and classification. Iowa State University Press, Ames, IA.

Burges, A., and D.P. Drover. 1953. The rate of podzol development in sands of the Woy Woy district, New South Wales. Aust. J. Bot. 1:83–94.

Burras, C.L., and W.H. Scholtes. 1987. Basin properties and post-glacial erosion rates of minor moraines in Iowa. Soil Sci. Soc. Am. J. 51:1541–1547.

Busacca, A.J. 1989. Long quaternary record in eastern Washington, U.S.A., interpreted from multiple buried paleosols in loess. Geoderma 45:105–122.

Bushnell, T.M. 1942. Some aspects of the soil catena concept. Soil Sci. Soc. Am. Proc. 7:466–476.

Cleaves, E.T., A.E. Godfrey, and O.P. Bricker. 1970. Geochemical balance of a small watershed and its geomorphic implications. Geol. Soc. Am. Bull. 81:3015–3032.

Conacher, A.J., and J.B. Dalrymple. 1977. The nine unit land surface model: An approach to pedogeomorphic research. Geoderma 18:1–154.

Crabtree, R.W., and T.P. Burt. 1983. Spatial variation in solutional denudation and soil moisture over a hillslope hollow. Earth Surf. Processes Landforms 8:151–160.

Dalrymple, J.B., R.J. Blong, and A.J. Conacher. 1968. A hypothetical nine-unit landsurface model. Z. Geomorphol. 12:60–76.

Dan, J., and D.H. Yaalon. 1964. The application of the catena concept of pedogenesis in Mediterranean and desert fringe regions, p. 751–758. Trans. 8th Int. Congr. Soil Sci. Bucharest, Romania.

Daniels, R.B., and E.E. Gamble. 1967. The edge effect in some ultisols in the North Carolina coastal plain. Geoderma 1:117–124.

Daniels, R.B., E.E. Gamble, and J.G. Cady. 1970. Some relations among coastal plain soils and geomorphic surfaces North Carolina. Soil Sci. Soc. Am. J. 34:648–653.

Daniels, R.B., E.E. Gamble, and J.G. Cady. 1971. The relation between geomorphology and soil morphology and genesis. Adv. Agron. 23:51–88.

Daniels, R.B., and R.D. Hammer. 1992. Soil geomorphology. John Wiley & Sons, New York.

Dijkerman, J.C. 1974. Pedology as science: The role of data models and theories in the study of natural systems. Geoderma 11:73–93.

Driese, S.G., L.C. Nordt, W.C. Lynn, C.A. Stiles, C.I. Mora, and L.P. Wilding. 2005. Distinguishing climate in the soil record using chemical trends in a Vertisol climosequence from the Texas Coast Prairie, and application to interpreting Paleozoic paleosols in the Appalachian basin, U.S.A. J. Sediment. Res. 75:339–349.

Emerson, D.G., M.D. Sweeney, V.M. Dressler, and S.W. Norbeck. 1990. Instrumentation and data for a study of seasonally frozen soil in southeastern North Dakota. Open File Report 90-107. U.S. Geological Survey.

Fenneman, N.M. 1931. Physiography of the western United States. McGraw-Hill Co., New York.

Fenneman, N.M. 1938. Physiography of the eastern United States. McGraw-Hill Co., New York.

Fenneman, N.M. 1946 (reprinted 1957). Physical divisions of the United States. U.S. Geological Survey, 1:7,000,000. U.S. Government Printing Office, Washington, DC.

Finney, J.R., N. Holowaychuk, and M.R. Heddleson. 1962. The influence of micro-climate on the morphology of certain soils of the Allegheny Plateau of Ohio. Soil Sci. Soc Am. Proc. 26:287–292.

Franks, C.D., L.C. Brockman, P.M. Whited, and S.W. Waltman. 1993. Using the benchmark site selection process to enhance modeling. USDA–NRCS Internal Report. National Soil Survey Center, Lincoln, NE.

Franzmeier, D.P., E.J. Pedersen, T.J. Longwell, J.G. Byrne, and C.K. Loshe. 1969. Properties of some soils in the Cumberland Plateau related to slope aspect and position. Soil Sci. Soc. Am. Proc. 33:755–761.

Franzmeier, D.P., and E.P. Whiteside. 1963. A chronosequence of podzols in northern Michigan. I. Ecology and description of pedons. II. Physical and chemical properties. Mich. State Univ. Agric. Exp. Stn. Quart. Bull. 46:1–36.

Gamble, E.E. 1993. Geomorphic study in the upper Gasconade River Basin, Laclede and Texas counties, Missouri. Soil Survey Investigations Report No 43.

Glazovskaya, M.A. 1968. Geochemical landscapes and types of geochemical soil sequences. Trans. 9th Int. Congr. Soil Sci. 4:303–312.

Goldich, S.S. 1938. A study in rock weathering. J. Geol. 46:17–58.

Graham, R.C. 1986. Geomorphology, mineral weathering, and pedology in an area of the Blue Ridge Front, North Carolina. Ph.D. Dissertation. North Carolina State University. Raleigh, NC.

Graham, R.C., D.R. Hirmas, Y.A. Wood, and C. Amrhein. 2008. Large near-surface nitrate pools in soils capped by desert pavement in the Mojave Desert, California. Geology 36:259–262.

Hall, G.F. 1983. Pedology and geomorphology, p. 117–140. *In* L.P. Wilding et al. (eds.) Pedogenesis and soil taxonomy. I. Concepts and interactions. Elsevier, New York.

Hirmas, D.R. 2008. Surface processes, pedology, and soil-landscape modeling of the southern Fry Mountain Bolson, Mojave Desert, California. Ph.D. Dissertation. University of California. Riverside, CA.

Holzhey, C.S., R.B. Daniels, and E.E. Gamble. 1975. Thick Bh horizons in the North Carolina coastal plain: II. Physical and chemical properties and rates of organic additions from surface sources. Soil Sci. Soc. Am. Proc. 39:1182–1187.

Hopkins, D.H. 1996. Hydrologic and abiotic constraints of soil genesis and natural vegetation patterns in the sandhills of North Dakota. Ph.D. Dissertation. North Dakota State University. Fargo, ND.

Huggett, R.J. 1975. Soil landscape systems: A model of soil genesis. Geoderma 13:1–22.

Huggett, R.J. 1976. Lateral translocation of soil plasma through a small valley in the northern Great Wood Hertfordshire. Earth Surf. Processes 1:99–109.

Hunt, C.B. 1986. Surficial deposits of the United States. Van Nos Reinhold Co. Inc., New York.

Jackson, M.L. 1968. Weathering of primary and secondary minerals in soils. Trans. 9th Int. Congr. Soil Sci. 4:281–292.

Jenny, H.J. 1941. Factors of soil formation. McGraw-Hill, New York.

Kemmis, T.J. 1991. Glacial landforms, sedimentology, and depositional environments of the Des Moines Lobe northern Iowa. Ph.D. Dissertation. University of Iowa, Iowa City, IA.

Kemmis, T.J., G.R. Hallberg, and A.J. Lutenegger. 1981. Depositional environments of glacial sediments and landforms on the Des Moines Lobe. Iowa Geological Survey Guidebook Series No. 6. Iowa Department of Natural Resource, Des Moines, IA.

King, L.C. 1957. The uniformitarian nature of hillslopes. Trans. Edinburgh Geol. Soc. 17:81–102.

Kleiss, H.J. 1970. Hillslope sedimentation and soil formation in northeastern Iowa. Soil Sci. Soc. Am. Proc. 34:287–290.

Knox, J.C. 1982. Quaternary history of the Kickapoo and lower Wisconsin River Valleys, Wisconsin. *In* J.C. Knox, L. Clayton, and D.M. Mickelson (eds.) Quaternary history of the driftless area, Wisconsin Geological and Natural History Survey, field trip guidebook No. 5, Madison, WI.

Knuteson, J.A., J.L. Richardson, D.D. Patterson, and L. Prunty. 1989. Pedogenic carbonates in a Calciaquoll associated with a recharge wetland. Soil Sci. Soc. Am. J. 53:495–499.

Lotspeich, F.B., and H.W. Smith. 1953. Soils of the Palouse loess. I: The Palouse catena. Soil Sci. 76:467–480.

Mausbach, M.J., and J.L. Richardson. 1994. Biogeochemical processes in hydric soil formation, p. 68–127. *In* Current topics in wetland biogeochemistry. Vol. 1. Wetland Biogeochemistry Institute, Louisiana State University, Baton Rouge, LA.

McAuliffe, J.R., E.P. Hamerlynck, and M.C. Eppes. 2007. Landscape dynamics fostering the development and persistence of long-lived creosote bush (*Larrea tridentata*) clones in the Mojave Desert. J. Arid Environ. 69:96–126.

Milne, G. 1936a. A provisional soil map of East Africa. *In* East African agriculture research station. Amani Memoirs 34p, Tanganyika Territory.

Milne, G. 1936b. Normal erosion as a factor in soil profile development. Nature 138:548–541.

Monger, H.C., and B.T. Bestelmeyer. 2006. The soil-geomorphic template and biotic change in arid and semi-arid ecosystems. J. Arid Environ. 65:207–218.

Morrison, R.B. 1967. Principles of quaternary soil stratigraphy, p. 1–69. *In* R.B. Morrison and R.E. Wright (eds.) Quaternary soils. International Association for Quaternary Research (INQUA), VII Congr. Proc. Vol. 9. Center for Water Resources Research, Desert Research Institute, University of Nevada, Reno, NV.

NACSN (North American Commission on Stratigraphic Nomenclature). 1983. North American stratigraphic code. Am. Assoc. Pet. Geol. Bull. 67:841–875.

Oliver, M.C. 1997. An investigation of two landscapes on the Holston limestone formation in McMinn County, TN. M.S. Thesis. University of Tennessee, Knoxville, TN.

Olson, C.G., and W.D. Nettelton. 1999. Paleosols and the effects of alteration. Quat. Int. 51–52:185–194.

Omernik, J.M. 1987. Ecoregions of the conterminous US (map supplement) (scale 1:7500000). Ann. Assoc. Am. Geogr. 77(1):118–125.

Omernik, J.M. 1995. Ecoregions—A framework for environmental management, p. 49–62. *In* W.S. Davis and T.P. Simon (eds.) Biological assessment and criteria—Tools for water resource planning and decision making. Lewis Publishers, Boca Raton, FL.

Pennock, D.J., and D.F. Acton. 1989. Hydrological and sedimentological influences on Boroll catenas, Central Saskatchewan. Soil Sci. Soc. Am. J. 53:904–910.

Pennock, D.J., and W.J. Vreeken. 1986. Soil-geomorphic evolution of a Boroll catena in southwestern Alberta. Soil Sci. Soc. Am. J. 50:1520–1526.

Peters, D.P.C., B.T. Bestelmeyer, J.E. Herrick, E.L. Fredrickson, H.C. Monger, and K.M. Havstad. 2006. Disentangling complex landscapes: New insights into arid and semiarid system dynamics. Bioscience 56:491–501.

Peterson, F.F. 1981. Landforms of the Basin and Range Province for soil survey. NV Argic. Exp. Stn. Tech. Bull. 28.

Rango, A., M. Chopping, J. Ritchie, K. Havstad, W. Kustas, and T. Schmugge. 2000. Morphological characteristics of shrub coppice dunes in desert grasslands of southern New Mexico derived from scanning LIDAR. Remote Sens. Environ. 74:26–44.

Retallack, G.J. 2005. Pedogenic carbonate proxies for amount and seasonality of precipitation in paleosols. Geology 33:333–336.

Richardson, J.L. 1997. Soil development and morphology in relation to shallow ground water—An interpretation tool, p. 229–233. *In* K.W. Watson and A. Zaporozec (eds.) Advances in ground-water hydrology: A decade of progress. American Institute of Hydrology, St. Paul, MN.

Ruhe, R.V. 1960. Elements of the soil landscape. Trans. 7th Int. Congr. Soil Sci. 23:165–169.

Ruhe, R.V. 1965. Quaternary paleopedology, p. 755–764. *In* H.E. Wright and D.E. Frey (eds.) The Quaternary of the United States. Princeton University Press, Princeton, NJ.

Ruhe, R.V. 1969. Quaternary landscapes in Iowa. Iowa State University Press, Ames, IA.

Ruhe, R.V. 1975. Geomorphology. Houghton Mifflin, Boston, MA.

Ruhe, R.V., R.B. Daniels, and J.G. Cady. 1967. Landscape evolution and soil formation in southwestern Iowa. USDA Technical Bulletin No. 1349.

Ruhe, R.V., W.P. Dietz, T.E. Fenton, and G.F. Hall. 1968. The Iowan problem northeastern Iowa. Report of Investigation No. 7. Iowa Geological Survey, Iowa.

Ruhe, R.V., and W.H. Scholtes. 1959. Important elements in the classification of the Wisconsin glacial stage. J. Geol. 67:585–593.

Ruhe, R.V., and P.H. Walker. 1968. Hillslope models and soil formation. I. Open systems. Trans. 9th Int. Congr. Soil Sci. 4:551–560.

Saucier, R.T. 1994. Geomorphology and Quaternary geologic history of the lower Mississippi Valley. U.S. Army Engineer Waterways Experiment Station. U.S. Army Corps of Engineers, Vicksburg, MS.

Schoeneberger, P.J., A. Amoozegar, and S.W. Buol. 1995. Physical properties of a soil and saprolite continuum at three geomorphic positions. Soil Sci. Soc. Am. J. 59:1389–1397.

Schoeneberger, P.J., and D.A. Wysocki. 1996. Geomorphic descriptors for landforms and geomorphic components: Effective models, weaknesses and gaps. (Abstract). American Society of Agronomy, Annual Meetings. ASA, Indianapolis, IN.

Schoeneberger, P.J., and D.A. Wysocki. 1997. Geomorphic description system. *In* P.J. Schoeneberger, D.A. Wysocki, E.C. Benham, and W.D. Broderson (eds.) Field book for describing and sampling soils. USDA–NRCS, Lincoln, NE.

Simonson, R.W. 1959. Outline of a generalized theory of soil genesis. Soil Sci. Soc. Am. Proc. 23:152–156.

Simonson, R.W. 1995. Airborne dust and its significance to soils. Geoderma 65:1–43.

Soil Survey Staff. 1996. Keys to soil taxonomy. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1998. National soil survey handbook. USDA Natural Resources Conservation Service, Lincoln, NE.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. U.S. Government Printing Office, Washington, DC.

SSSA (Soil Science Society of America). 1997. Glossary of soil science terms. SSSA, Madison, WI.

SSSA (Soil Science Society of America). 2010. Glossary of soil science terms. SSSA, Madison, WI. Available online at Web site: https://www.soils.org/publications/soils-glossary

Steinwand, A.L., and T.E. Fenton. 1995. Landscape evolution and shallow groundwater hydrology of a till landscape in central Iowa. Soil Sci. Soc. Am. J. 59:1370–1377.

Steinwand, A.L., and J.L. Richardson. 1989. Gypsum occurrence in soils on the margin of semipermanent prairie pothole wetlands. Soil Sci. Soc. Am. J. 54:836–842.

Stuart, D.M., G.E. Schuman, and A.S. Dylla. 1971. Chemical characteristics of the coppice dune soils in Paradise Valley, Nevada. Soil Sci. Soc. Am. J. 35:607–611.

Thornbury, W.D. 1965. Regional geomorphology of the United States. John Wiley & Sons, Inc., New York.

Thornbury, W.D. 1969. Principles of geomorphology. 2nd edn. John Wiley & Sons, Inc., New York.

USDA-EPA. 1996. Developing a spatial framework of ecological units for the United States. p. 46–54. *In* GIS/LIS 1996 Ann. Conf. Expo. Proc., Denver. CO. American Society for Photogrammetry and Remote Sensing, Bethesda, MD.

USDA-SCS (United States Department of Agriculture-Soil Conservation Service). 1981. Land resource regions and major land resource areas of the United States: Agricultural handbook 296. USDA, Washington, DC.

Valentine, K.W.G., and J.B. Dalrymple. 1976. Quaternary buried paleosols: A critical review. Quat. Res. 6:209–222.

Van Wambeke, A., and T. Forbes. 1986. Guidelines for using taxonomy in the names of soil map units. Soil Management Support Services. Technical Monograph No. 10. SCS, Washington, DC.

Walker, P.H. 1966. Postglacial environments in relation to landscape and soils on the Cary drift, Iowa. Iowa Agric. Exp. Stn. Res. Bull. 549:838–875.

Walker, P.H., and R.V. Ruhe. 1968. Hillslope models and soil formation. I. Closed systems. Trans. 9th Int. Congr. Soil Sci. 4:561–568.

Wells, S.G., J.C. Dohrenwend, L.D. McFadden, B.D. Turrin, and K.D. Mahrer. 1985. Late Cenozoic landscape evolution on lava flow surfaces of the Cima volcanic field, Mojave Desert, California. Geol. Soc. Am. Bull. 96:1518–1529.

Wood, A. 1942. The development of hillside slopes. Geol. Assoc. Proc. 53:128–138.

Wood, Y.A., R.C. Graham, and S.G. Wells. 2005. Surface control of desert pavement pedologic process and landscape function, Cima volcanic field, Mojave Desert, California. Catena 59:205–230.

Yaalon, D.H. 1971. Soil-forming processes in time and space, p. 29–40. *In* D.H. Yaalon (ed.) Paleopedology. Israel University Press, Jerusalem, Israel.

Young, M.H., E.V. McDonald, T.G. Caldwell, S.G. Benner, and D.G. Meadows. 2004. Hydraulic properties of a desert soil chronosequence in the Mojave Desert, USA. Vadose Zone J. 3:956–963.

Pedogenic Processes

Judith K. Turk
University of California

Oliver A. Chadwick
University of California

Robert C. Graham
University of California

30.1 Introduction

Soil is a mixture of geological parent material, living organisms, and the colloidal residue of their interaction. The colloidal fraction is particularly important in imparting soil with its unique characteristics. The nature of soil exposes a large surface area that gives rise to adsorption of water and chemicals, ion exchange, adhesion and capillarity, swelling and shrinking, and dispersion and flocculation. Soil colloids include inorganic (clay minerals) and organic (humus) components. The nature and amount of these colloids vary in response to environmental stimuli and organic input. Their variation is responsible for lateral differentiation in soil properties. Soils are separated vertically into horizons that indicate differences in the internal soil environment, which, in turn, determines the amount and character of soil colloid accumulation. This chapter provides an overview of the way in which soil properties vary as soils develop under varying environmental stimuli.

Soil formation is powered by gradients in chemical and physical potential as Earth's atmosphere and biosphere interact with rocks and minerals. Rock minerals are formed at high temperature and pressure, which makes them susceptible to attack by water and the biospherically derived acids existing in soils. The fundamental mechanisms of physical, chemical, and biological processes in soils are discussed in Parts I, II, and IV. Here the focus will be on the matrix of processes that produce 3D soil bodies called pedons, which have differentiated horizons and lateral variability.

30.2 Environmental Factors That Drive Pedogenesis

The concept of soil-forming factors is one of the earliest and most important of soil science. It defines soil as a component of ecosystems that must be characterized in terms of both geological substrate and biological input (Jenny, 1941; Amundson and Jenny, 1997). It provides a far-reaching description of the controls on soil processes that allows prediction of general soil properties based on the interaction of a relatively small number of driving variables. In the words of Hans Jenny (1980), "Pedogenic order in a landscape is unraveled by stratified random sampling along vectors of the state factors." Knowledge of the soil properties produced by the interaction of environmental variables allows prediction of soil characteristics in detail, locally, or in general, regionally and globally. The fact that characteristics of soil properties can be predicted based on environmental variables is critical to soil resource mapping because it is impossible to observe soil profiles everywhere.

The interaction of environmental variables establishes the soil-forming processes whose actions on parent material manifest themselves in characteristic soil morphologies, which may, in turn, alter the nature of the ongoing processes. Environmental variables are broadly grouped as geological, climatological, biological, and topographical. The geological characteristics constitute a site factor that sets the initial condition for soil formation, whereas the climatological and biological factors represent energy input that drives soil development. The energy factors share a strong spatial covariance because of the dependence of organisms on climate. Within any of these groupings, local differences in topography modify the activity of the more broadly defined variables. Human manipulation of edaphic environments is significant enough to warrant close consideration of their effect on soil properties (Amundson and Jenny, 1991). Dynamic soil properties (e.g., soil organic carbon, bulk density, pH, salinity, aggregate stability) that change on timescales from decades to centuries are most affected by human land use (Tugel et al., 2005).

30.2.1 Geologic Foundation: Lithology, Mineralogy, and Topography

Geological processes ranging from plate tectonics and volcanism to deep ocean carbonate sedimentation to glaciations create conditions that eventually provide substrate for soil

development. Plate tectonics drive global processes that differentiate lighter, silica-rich rocks from the mantle into the crust and create relief that produces erosion, accumulation of alluvium, and formation of sedimentary rocks. Continental areas that are actively uplifted, such as the Tibetan plateau and the cordillera of South and North America, support thin, young soils. In contrast, those areas on stable cratons, such as central West Africa, Brazil, and Australia, have deep soils that are superimposed on thick, highly weathered saprolite zones (Paton et al., 1995). In these ancient cratons, depth to unweathered rock can exceed 100 m. Volcanism brings locally melted silica-rich rock in subduction zones to the surface in areas such as Japan, Indonesia, and Washington state. Hot spots, such as Hawaii, the Canary Islands, or Yellowstone, provide a direct conduit from mantle to surface for Fe- and Mg-rich rocks. A dramatic example of mantle-fed volcanism is exhibited by flood basalts on the Columbia and Deccan plateaus. On ocean floors, far from continental inflow, deposition of carbonate produces limestone that, when subaerially exposed, has distinct parent material properties when compared with igneous rock.

In polar and temperate zones, glacial advances deposit moraines and outwash composed of the mix of lithologies crossed by the glacier. The more finely ground glacial drift is lofted into the atmosphere and deposited as loess. Numerous Pleistocene glaciations have left distinctly younger soils than the deeply weathered soils of stable craton areas in the tropics. Even within temperate zones, there are strong differences in soil properties between recently glaciated areas and those that have not been overrun. Globally, much of the variation in soil properties can be attributed directly to differences in parent material mineralogy and the dynamics of Earth's history. Locally they are important as well. For instance, stream channel and associated hillslope erosion may cut through an extensive surficial geological unit into an underlying lithology that has different mineralogical and chemical properties, which affect soil properties of soils that form on that unit (Chapter 29).

Parent material lithology determines the physical and mineralogical nature of soil. For example, soil formed in marine clays or shale inherits a large quantity of colloids that lead to poor water infiltration, relatively low plant-available water, and substantial shrink/swell behavior. In contrast, soils formed on quartz sandstone or weathered granite are likely to be sandy with little inherited clay. Commonly, they are excessively drained and often have low fertility and water-holding capacity. Fracture patterns in the underlying bedrock provide preferential lines of water flow, which enhance weathering and soil formation. The stratigraphic juxtaposition of soft and hard lithologies provides gentler and steeper relief and less or more rock fragments in the overlying soil.

Parent material mineralogy determines ecological soil properties, such as nutrient supply and retention, and water movement. For example, mafic mantle-derived rocks typically weather to a smectite and Fe (hydr)oxide-rich colloidal fraction with the simultaneous release of Ca and Mg for use by plants and soil organisms. The smectites provide high cation exchange capacity (CEC), which can easily retain nutrient cations. Felsic crust-derived rocks tend to weather to a kaolin- and gibbsite-rich colloidal fraction with release of Ca and K for use by the biota. Kaolin and gibbsite provide only a small amount of cation exchange capability, leading to relatively rapid leaching of nutrient cations. The presence of quartz in the felsic rocks provides a skeletal framework of resistant sand size grains that tends to promote water percolation. In contrast, the smectites derived from mafic rocks tend to form clay-rich horizons. The clays tend to swell when wet, which can reduce water flow by closing off macropores. Often igneous rocks contain a broad suite of minerals that weather to provide a mixture of nutrients and a diverse array of secondary minerals. In contrast, limestone is quite pure calcite, which should only release Ca and H_2CO_3 when it reacts with water. Soils forming on limestone often have inherited their colloidal fraction from weathering of impurities in the rock or from eolian addition. Similarly, soils forming on quartz sandstone or quartzite must inherit their colloidal fraction from eolian input.

30.2.2 Energy: Water, Temperature, and Biology

Soil development is strongly dependent on energy that is ultimately provided by the sun and Earth's gravitational field. Soil is heated directly by energy absorption at the surface and indirectly by transfer from the atmosphere. These processes set up superimposed diurnal, seasonal, and annual waves of energy moving through soil profiles. Soil is cooled by latent heat transfers or through long-wavelength emissions. Soil provides short-term storage of water in the global hydrological cycle, which moves water from oceans to the atmosphere and back. Rain and snowmelt provide water to soils through downward infiltration under the influence of gravity and adhesive forces. Within soil profiles, water is held by adhesive and cohesive forces; however, once the water-holding capacity is exceeded, water can move downward to the groundwater table or laterally to river systems. Evaporation and transpiration remove water directly from the soil surface and from plant leaves. Plant roots collect water from much greater depths, thus extending greatly the role of latent heat flux in driving land–atmosphere hydrologic transfers.

During Earth's history, evolution of photosynthesis and establishment of rooted plants on land initiated a powerful source of energy to drive pedogenesis. Photosynthesis produces complex organic molecules, which are shed to the soil either directly or after they have cycled through animals. In the soil, organic matter is metabolized by microorganisms; this process (often called soil respiration) releases the CO_2 and water that was utilized in photosynthesis. The paired photosynthesis and soil respiration reactions set up a giant planetary reduction–oxidation cycle involving carbon and oxygen. Carbon in CO_2 is reduced to organic compounds; O in water is oxidized to molecular form. In turn, microbial breakdown of organic compounds oxidizes C back to CO_2, transferring electrons to O_2, which is reduced to

water. Photosynthesis captures massive amounts of solar energy and much of the ensuing organic matter is transferred to the soil, where its breakdown by soil's vast microbial populations provides most of the energy that drives pedogenesis. Without water, nutrients, and an appropriate temperature range, biological reactions cannot function, but once these environmental conditions are met, it is the C–O redox cycle that releases organic acid into soil, which, in turn, drives specific pedological transformations.

Organic acids lower the pH of soil solution thereby enhancing mineral weathering, but the release of cations during weathering acts to neutralize the acidity. The amount of water percolation and cation removal determines the extent to which weathering buffers soil pH against acidification and determines the nature of secondary mineral synthesis. In well-drained humid environments, dilute soil solution leads to strong chemical gradients, intense weathering, efficient leaching of soluble components (basic cations and silicic acid), and secondary mineral synthesis involving only the least soluble components (Fe and Al (hydr) oxides and kaolin minerals). In poorly drained sites, soil solution accumulates highly soluble ions, as well as sparingly soluble ions, leading to synthesis of smectite (in environments with high Si and Mg activity) and other minerals. In these soils, Fe and Mn may be reduced and slowly leached away. In contrast, arid environments often have soil solutions that contain large concentrations of ions, leading to minor chemical gradients, short leaching distances for only the most soluble ions, and secondary mineral synthesis that is dominated by ionic salts (calcite, gypsum, halite, etc.). Thus, long-term average soil solution concentrations determine most pedochemical properties. Furthermore, they help to condition the terrestrial portion of many biogeochemical cycles because elements released by weathering and not utilized biologically or assimilated into pedogenic minerals are rapidly leached from soil into groundwater, rivers, lakes, and oceans (Berner and Berner, 1996).

The annual amount and seasonal distribution of precipitation and solar energy define broad life zones or biomes based on ecosystem composition and productivity (Aber and Melillo, 1991). Up to some maximum, increasing rainfall leads to greater net primary production of C, which, in turn, leads to greater production of organic matter and more rapid rates of nutrient cycling (Schlesinger, 1997). Highly productive ecosystems often have high levels of CO_2 and enhanced organic acid production, which lead to low soil solution pH values and high weathering rates that release more nutrients until their store of primary minerals is exhausted. Temperature controls the rate of biological processes and mineral weathering/synthesis. At 0°C, and below, little water is available to mediate chemical reactions. In extremely cold environments, physical disintegration of rock material often predominates over chemical breakdown of minerals. For each 10°C increase, biologically mediated reaction rates double until the upper limit of enzymatic functioning is reached. This implies that, all other things being equal, soils in tropical environments will receive far more yearly heat energy and will be more chemically weathered than those in arctic environments (Ugolini and Spaltenstein, 1992; van Wambeke, 1992). In hilly and mountainous terrain, precipitation and solar energy inputs are modified by local topographic position. Soil moisture varies along hillslopes because of redistribution from convex to concave positions, with consequent impact on vegetation type and productivity. Locally, sites are drier or moister than are predicted by average climatic parameters. Similarly, local relief modifies the angle and duration of solar illumination and hence temperature.

In addition to capturing solar energy by photosynthesis, plants have more direct influences on soil formation. The locus of deposition of dead plant material is a factor determining the nature of near-surface horizons. In forests, periodic litter fall deposits organic matter on the soil surface resulting in O horizons. Litter partly decomposes in the O horizon, releasing residues that are carried to the mineral soil in solution as colloidal flow or mixed into the mineral soil by soil fauna. Downward movement of organic acids provides chelating power, which can produce horizons of depletion in the mineral surface soil (E horizon), where one might expect to find organic accumulation (A horizon). In contrast, grasslands tend to be dominated by fire, which releases nutrients as ash at the soil surface, and they shed more organic matter as dead roots and root exudates in the mineral surface soil horizons (producing A horizons). When trees fall, soil and underlying geologic material are often pulled up with the root mass, acting to mix and destroy existing horizons. Roots break up geological substrate through physical expansion and enhance chemical weathering by exuding acidic molecules. In this way, they produce preferential flow paths where pedogenic processes are intensified. Roots fuel nutrient cycling by providing a direct conduit for nutrient movement from soil to aboveground plant parts, which are eventually returned to the upper soil horizons. Mycorrhizal fungi associated with the roots of many plant species are particularly good at decomposing mineral substrate to release P (Chapter 27).

Roots are an important source of food for animals that burrow for food or shelter. Fossorial mammals are effective at mixing parent material fabric and soil horizons. They can sort particle sizes and create macropores that direct preferential water flow. Smaller animals are equally effective at driving processes that enhance or retard specific vectors of soil formation. Ants, termites, springtails, and earthworms move organic matter from the surface downward, mix soil horizons, and create macropores. On hillslopes, animals that deposit soil material on the surface can significantly enhance downslope redistribution of mass through their diggings.

30.3 Pedogenic Processes

In combination, environmental variables drive processes that act on geological or preexisting soil substrate to effect pedogenic alteration at scales ranging from the microscopic parts of soil fabric to watershed soil mantles. Commonly, there will be many processes acting partly in conjunction and partly in opposition. In their totality, they produce soil bodies with recognizable horizons.

In some cases, continued enhancement of horizon properties can, itself, modify ongoing processes, changing the vector of soil formation, and initiating horizon deterioration.

Understanding soil formation requires measurement of present processes, prediction of future trends based on these processes, and interpretation of past processes based on present, relict morphology. A large body of pedological research informs this chapter. The authors have not referenced most of it, choosing instead to focus on essential concepts and specific examples.

30.3.1 Conceptual Process Model

A given body of in situ soil material, regardless of size, can be altered by additions, losses, transformations, or translocations (Simonson, 1959). These processes interact differently depending on the depth from the soil surface and the mix of environmental variables at a specific location. Some of these processes lend themselves to mass balance calculations, others to use of specific tracers or extractions, but, in general, there is no single, universal approach that will provide knowledge of these four processes.

30.3.1.1 Additions

Soil is formed when C (and N) is added from the atmosphere by biological fixation processes. Enzymes in plants and associated microorganisms reduce atmospheric components into biologically useful forms. Dead plants and animals feed a host of microorganisms, which utilize their C and N for growth and energy. Residues from these processes accumulate in soils in complex forms collectively known as soil organic matter (SOM). Since addition of organic residues is greatest near the surface, soil functioning can sometimes be conceptualized into two compartments: an upper part where organic acids and microbial dynamics hold sway and a lower part where inorganic processes dominate, though still influenced by organic processes. It is the mixture of SOM with the porous matrix of inherited and newly synthesized minerals that fundamentally defines Earth's soils.

Solutes and particulate matter are added to soils either directly from the atmosphere (Simonson, 1995; Derry and Chadwick, 2007) or by movement from topographically higher points in the soil landscape (Paton et al., 1995). Although these added components are often incorporated with little outward evidence of their influence, their role can be significant, such as when ecosystems growing on highly weathered soils are sustained by nutrients added from atmospheric sources. Atmospheric additions provide nutrients to rainforests that have inherently low fertility, sometimes providing the majority of nutrient inputs to the ecosystem (Chadwick et al., 1999; Okin et al., 2004; Pett-Ridge, 2009). When additions are rapid and large, existing soils may be buried and new soils developed on the added material. An intermediate result is formation of cumulic soils where soil formation continues as new matter is added at the surface, thickening the soil, as may occur along the lower portions of hillslopes where matter eroded form higher positions accumulates or in areas receiving moderate amounts of airborne dust (Chapter 29).

30.3.1.2 Losses

Mass is lost from soils through erosion by wind and water at the surface and by leaching of solutes and colloids. Eroded soils often lie on the upper slope portions of hillslopes. Microbes break down SOM and minerals weather, releasing ions to the soil solution. In the organically dominated soil compartment, organic chelates facilitate movement of ions deeper into the profile. When excess water moves through soil pores, some elements are carried into streams or groundwater. Typical cations lost through leaching are Na^+, Ca^{2+}, K^+, Mg^{2+}, and as soil acidity increases, smaller amounts of Al are leached in the form of variously charged hydrolysis products. Except at pH values >9, $Si(OH)_4$ is leached as an uncharged compound. Anions leached from soil are dominated by HCO_3^-, which leads to significant amounts of atmospheric carbon sequestration on geological timescales (Chadwick et al., 1994). Lesser amounts of NO_3^- and SO_4^{2-} are leached from soils, but those receiving acid rain may show large losses of these anions (Sposito, 1989). Under aerobic conditions, little Fe is leached from soils and both Fe and Al tend to accumulate by residual enrichment because of leaching of other elements. Iron and Mn can be leached from soils under anaerobic conditions because their reduced forms are soluble.

30.3.1.3 Transformations

Organic matter added to soil is modified by microbial respiration leading to accumulation of more complex SOM or organic compounds that bond with inorganic colloids. Weathering of parent material minerals provides not only nutrients for biological activity but also building blocks for synthesis of secondary minerals. Compared with mantle- and crust-derived minerals that form soil parent material, secondary minerals are enriched in stable elements such as Al and Fe and depleted in easily solubilized elements such as base cations and Si. Under humid, well-drained conditions, where leaching is maximized, secondary mineral formation is characterized by synthesis of aluminosilicate clays and Al and Fe (hydr)oxides. In arid environments, leaching is negligible and even basic cations accumulate in the lower parts of soil profiles as soluble and semisoluble salts. These chemical transformations also lead to physical transformations. For example, wetting and drying results in swelling and shrinking that transform soil fabric, producing a structure reflective of its mineralogical, textural, and organic matter status. Smectite, vermiculite, and short-range order clays hold much more hydroscopic water than do the primary minerals from which they form. Chemical and mineralogical transformations can not only release nutrients for biological use but also sequester nutrients. Nitrogen can be immobilized by NH_4^+ fixation in clay minerals and P can be fixed by sorption onto Fe and Al (hydr)oxides at low to neutral pH values and precipitation with or adsorption to $CaCO_3$ at high pH values.

30.3.1.4 Translocations

Colloids, composed of low-molecular-weight organic matter and small clay particles, and dissolved constituents can move through soil pores. Commonly, the location of mobilization and deposition are controlled by complex chemical interactions between soil solution and the colloidal fraction. Translocation in humid environments usually is effected by downward flow of water. In arid environments, it is often downward as well, but salts can also be carried upward by capillary flow of water in response to evaporation. These processes lead to formation of zones of colloid and solute depletion, such as E horizons, and zones of colloid enrichment and solute precipitation, such as Bhs (≈spodic), Bt (≈argillic), Bk (≈calcic), By (≈gypsic), and Bqm (≈duripan) horizons. Another class of translocations includes mixing processes, which move bulk soil material upward, downward, or laterally in the soil. Mixing can be caused by activity of soil organisms, freeze–thaw cycles, or shrinking and swelling of expansive clays. These processes counterbalance colloid redistribution because they stir soil profiles, preventing or slowing concentration of translocated materials in subsurface horizons.

30.3.2 Dominant Processes

Locally, soil processes reflect a balance among gains, losses, internal redistribution, and chemical and physical changes. The resulting soil properties represent the long-term effect of these processes acting on a 3D reaction column that is open to exchange of matter and energy with the environment. Important soil-forming processes are described later, beginning with two universal aspects of soil formation: organic matter accumulation and formation of soil structure. Subsequent sections are organized to cover processes that are driven by water flux and to some degree, reflect the influence of progressively greater leaching: accumulation of soluble salts and gypsum, accumulation of $CaCO_3$, accumulation of opaline silica, redistribution of clay, complexation and redistribution of Fe and Al, leaching of Si, and concentration of resistant (hydr)oxides. Next, the processes in soil environments where the presence of water is so prevalent that it greatly limits the supply of O are discussed. In these soils, microbially induced reduction controls important aspects of the chemistry, mineral transformation, and morphology. Finally, biological and physical processes of soil mixing are considered. Soil mixing occurs to some degree in most soil-forming environments but is especially influential under some conditions, such as in the seasonally thawed layer above a permafrost table.

30.3.2.1 Organic Matter Accumulation and Alteration

The primary source of organic matter in soils is vegetation, with leaves, stems, and floral parts added to the surface as they drop from the plant, and roots added directly into the soil itself as they grow. Animals and microorganisms (Part IV) feed on the vegetable matter, and each other, decomposing the organic matter parent material to yield gases (e.g., CO_2, CH_4) and humus. The influence of SOM is out of proportion to its weight fraction,

a reality that is recognized in the designation of soil materials as organic when they contain a minimum of only 12%–20% organic C. The organic fraction of soils can be extremely important in determining many aspects of soil processes. For example, even in soils with as little as 0.5%–1% organic C, SOM controls pesticide adsorption.

30.3.2.1.1 Organic Matter Decomposition

Soil organisms use organic tissues as a source of energy and C. Each type of organism has its own role in the decomposition process, from primary decomposers, such as beetles, ants, termites, earthworms, and fungi, to decomposers that feed on the feces or tissue of the primary decomposers, such as bacteria, fungi, and various meso- and macrofauna. A succession of populations operates, each using the altered material from the previously active population and further altering the organic substrate itself. The result is a successive depletion of chemical energy sources, increased resistance to microbial attack, and a lowering of the C:N ratio of the SOM.

The rate of decomposition is controlled by a number of factors related to organic matter quality, the interaction of organic matter with soil minerals, and environmental conditions that affect biological activity. The amount and kind of the various compounds in the organic matter are critical (Melillo et al., 1989). Simple organic acids, sugars, and starches are very quickly utilized by microorganisms (on the order of days to months). Protein, chitin, cellulose, and hemicellulose are used less rapidly and are listed in order of increasing resistance to decomposition. Lignin, fats, and waxes are most resistant to decomposition.

Association with soil minerals can be an important factor in stabilizing organic matter against decomposition. In a study of organic matter in subsoil horizons of acidic forest soils, 73% of stable organic matter was found to be mineral protected (Mikutta et al., 2006). Furthermore, most variability in the size of the mineral-protected organic matter pool was explained by the concentration of poorly crystalline minerals. Soils dominated by short-range order minerals have slowly decomposing organic carbon pools that are larger than those of other mineral soils (Rasmussen et al., 2006). The high surface area of these minerals adsorbs organic carbon, decreasing its availability for decomposition, resulting in relatively slow decomposition rates.

Biological activity in soils is greatest when the soil pH is between 6 and 8, soil temperature is 20°C–40°C, soil water potential is between field capacity and about −1.5 MPa, and there are no deficiencies in essential nutrients. Rates of decomposition are especially slow when the soil is saturated for long enough to become anaerobic and when the temperature is below biological zero (≈5°C). Under these conditions, thick layers of organic matter (>40 cm) may accumulate above the mineral soil. On the other hand, there are microbes adapted for virtually every condition that may come about naturally in soils (Atlas and Bartha, 1981).

30.3.2.1.2 Characteristics of Soil Organic Matter

Soils may contain organic matter ranging across the entire spectrum of decomposition products, particularly if a mixing

mechanism, such as burrowing animals or surface cracks, operates to incorporate fresh organic material into the soil (Quideau et al., 1998). In the absence of aggressive mixing, much decomposition occurs within a litter layer. In general, SOM is increasingly decomposed with depth. The final product is humus: the dark brown, complex, microbially resistant, colloidal compounds that are highly modified original materials or materials synthesized by microbes. Humus has high surface area and chemical reactivity (CEC ranges from 100 to 300 $cmol_c$ kg^{-1}), is very hydroscopic, and readily forms complexes with the inorganic fraction (Chapter 11). Humus is so complex that its composition is usually defined using the following operational criteria (Oades, 1989): fulvic acid fraction, which is soluble in both alkali and acid, humic acid fraction, which is soluble in alkali but insoluble in acid, and the humin fraction, which is insoluble in both alkali and acid. Fulvic acid is composed of the lowest molecular weight compounds and has the greatest number of acid groups and the highest CEC. In contrast, humin is composed of the highest molecular weight compounds and has relatively few acid groups and the lowest CEC. In soils of pyrogenic ecosystems, charcoal may be a significant component of the carbon pool (MacKenzie et al., 2008). Carbon in the form of charcoal is highly resistant to decomposition and influences nutrient availability.

30.3.2.1.3 Relationship of Organic Matter to Morphology and Function

The accumulation of humified organic matter is key to formation of an A horizon, which, if it becomes thick enough, may evolve to qualify as a mollic or umbric epipedon. A horizons are an important link between soils and plants, since most plant roots exist within the A horizon and most of the nutrients are extracted from there. Decomposition of organic matter in the A horizon, including the roots themselves, biocycles essential nutrients from vegetation back into plant-available pools. High levels of organic acids lead to maximum rates of primary mineral decomposition in A horizons as well. Addition of organic matter also influences the physical properties of soils. As the A horizon develops in young soils, there is a decrease in bulk density, and the soil color becomes darker and redder (Turk et al., 2008). Organic matter tends to promote development of soil structure and porosity, thereby enhancing water infiltration and water-holding capacity.

30.3.2.2 Development of Soil Structure

An essential component of soil development is the substitution of soil structure for the geologic structure of parent material. While geologic structure is inherited from jointing cracks and bedding planes, soil structure forms as a function of shrinking and swelling induced by change in moisture content and by the downward movement of water and colloids along preferential flow lines. Soil structural units, known as peds, become more strongly expressed as their faces are differentiated from their interiors. The surface/interior differentiation comes about through several mechanisms. Wetting-induced expansion forms pressure faces or stress argillans by orienting platy silt and clay particles along

contact planes between peds. Subsequent drying causes contraction, which opens voids between peds. Upon rewetting, water flows preferentially through the contraction voids depositing clays and organic matter on the ped faces, further differentiating the outer from the inner parts of the peds. Other compounds, including calcite, silica, and Fe and Mn (hydr)oxides can also be deposited along ped faces. Conversely, under some conditions, outer margins of peds become depleted by preferential leaching of clay and soluble compounds. Roots follow paths of least resistance and concentrate in voids between structural units, enriching ped surfaces with organic matter and enhancing preferential drying of ped exteriors relative to interiors.

30.3.2.2.1 Size, Degree of Expression, and Reinforcement of Peds

A number of factors influence the size and degree of expression of soil structural units. The amount of volume change associated with wetting and drying cycles (swelling and shrinking) is very important. Fine-textured soils and those with smectitic mineralogy typically have relatively large volume changes and, therefore, develop the most strongly expressed structural units. Fine-textured soils also have less resistance to shear and so tend to have more cracks and finer structural units. Frequent shrink/swell cycles, which can be related to cyclic drying/wetting, tend to reinforce the expression of structural units (Southard and Buol, 1988). This is reflected in the development of soil structure in shallow agricultural drainage ditches that undergo cyclic drying/wetting but not in deeper ditches that are continually saturated (Vaughan et al., 2008). Slow drying allows a relatively uniform shrinkage and contraction of the soil mass, producing larger peds. Crystallization of salts in saline soils yields a finer structure than is found in similar nonsaline soils (Reid et al., 1993).

30.3.2.2.2 Granular Structure in Surface Horizons

The structure of surface horizons is often most strongly influenced by biological factors, including organic matter, microbes, plant roots, and soil fauna (Figure 30.1). Fecal casts of soil macrofauna, such as earthworms (Lee and Foster, 1991) and snails (Anderson et al., 1975), often form compact, stable aggregates, which dominate in some A horizons. Dense networks of fine grass roots and fungal hyphae cause uniform fine-scale shrinkage cracks and enmesh soil materials into granular or crumblike units (Oades, 1993). At the most fundamental level, soil aggregates in surface horizons are usually held together by SOM–mineral complexes (Chapters 3 and 11).

30.3.2.2.3 Platy Structure

Platy structure typically forms in response to unidirectional compressional forces (Figure 30.1). It is most often produced in surface soils through compaction by animals or machines, or by raindrop impact. Vertical compaction yields horizontally oriented, platy structural units. Gravity can contribute to the necessary vertical force for platy structure development in the silty vesicular horizons that are common in arid and semiarid regions. In these surface horizons, vesicles are formed by escaping gases as the soil material is wetted. Repeated wetting and

Type of structure	Soil environment	Formation processes
Granular	A horizon	Aggregation by biological agents: organic compounds, fungal hyphae, fine roots, fecal casts
Platy	A and other surface horizons Petrocalcic horizons Duripans	Vertical compression from compaction
Prismatic	B horizons with uniform texture, fine roots, few coarse fragments, relatively slow drying	Horizontal compression from shrink-swell
Blocky	B horizons with coarse fragments, heterogeneous root sizes	Shrink-swell interrupted by nonhomogeneous matrix

FIGURE 30.1 Types of soil structure, their occurrence within soils, and the processes responsible for their formation.

drying cycles enlarge the vesicles until they can no longer support the weight of the overlying material and collapse, forming planes that bound platy structural units (Miller, 1971). Platy structure is also common in highly developed petrocalcic horizons and duripans. The origin of platy structure in these subsoil horizons is less clear but may be caused by vertical compression generated as precipitation of secondary calcite and opal increases subsoil volume. In pergelic soils (Gelisols), platy structure is associated with the formation of lenticular ice (vein ice) (Section 33.6)

30.3.2.2.4 Prismatic and Blocky Structure

Shrinkage of moist soil material as it dries results in multidirectional compression centered around numerous loci within the mass. When shrinkage forces are largely resolved in lateral directions, the soil material may contract uniformly toward more or less equally spaced centers forming prismatic or columnar peds bounded by vertical cracks (Figure 30.1). In the formation of prismatic structure, uniform shrinkage is important because it develops the consistently spaced and arranged cracks that form ped boundaries; uniform shrinkage is favored by slow drying and homogeneous soil materials. It is often associated with uniformly fine-textured sediments in poorly drained, low-lying landscape positions. Prismatic or columnar structure is common in natric horizons because of enhanced swelling and poor drainage associated with Na-saturated clays. In contrast to the elongated character of prismatic structure, blocky structure forms as a result of combined lateral and vertical shrinkage forces (Figure 30.1), with neither predominating. Conditions favoring blocky structure development are rapid drying and factors that break continuity of soil fabric, such as woody roots and coarse fragments. Also, bedded moderately fine and fine-textured fluvial or marine deposits tend to form blocky structure in B horizons because horizontal planes of weakness, inherited from parent materials, are bisected by vertical shrinkage cracks. In permafrost soils, blocky structure is associated with ice nets, a cryostructure consisting of 2–5 cm diameter polygons separated by vertical or diagonal desiccation cracks (Ping et al., 2008).

30.3.2.2.5 Vertic Features

Vertisols and other clayey, expansive soils develop characteristic structural features as a result of their dynamic shrink/swell behavior (Section 33.7). Upon wetting, the fabric of these soils swells, commonly increasing its volume by about 10%–15%. In surface horizons, this increase in volume is accommodated by upward movement, but in the subsoil, overburden pressures confine vertical movement so that lateral swelling pressures are at least four times as great as the vertical pressures and exceed the soil shear strength (Wilding and Tessier, 1996). Under these conditions, shear failure takes place at angles between 10° and 60° from horizontal. The shear planes present slick, grooved surfaces, known as slickensides that bound characteristic wedge-shaped aggregates.

30.3.2.2.6 Aggregation by Fe and Al (Hydr)oxides

Highly weathered soils, such as Oxisols, have a fine (<2 mm diameter) granular structure that is the result of aggregation by Fe and Al (hydr)oxides (van Wambeke et al., 1983) (Section 33.13). Aggregation is affected by very small positively charged Fe and Al (hydr)oxide particles and surface coatings that form bridges between negatively charged clay particles to build up small aggregates (Hsu, 1989; Schwertmann and Taylor, 1989).

30.3.2.3 Accumulation and Redistribution of Salts

Gypsum and more soluble salts accumulate in poorly leached soils and soils where evapotranspiration (ET) greatly exceeds leaching. The composition and concentration of soluble salts

influence the physical behavior of soil particles because of the effect that dissolved salts have on dispersion and flocculation. Ecosystems with salt-affected soils are subject to many plant physiological stresses, related to surface crusting, low osmotic potential of the soil solution, and specific ion toxicities.

30.3.2.3.1 Characteristics of Soluble Salts

Sulfate, carbonate, chloride, and nitrate salts occur in soils of arid environments (Table 30.1). The concentration and type of salts influence the physical behavior and chemical properties of the soils in which they occur. Salt-affected soils have major impacts on the flocculation or dispersion of clay minerals and organic matter (Chapters 17 and 18 of *Handbook of Soil Sciences: Resource Management and Environmental Impacts*). In general, high electrolyte concentrations keep soils flocculated, whereas colloidal dispersion occurs when the electrolyte concentration is low and Na predominates on exchange sites. Thus, sodic soils [exchangeable sodium percentage (ESP) > 15, electrical conductivity (EC) < 4 dS m^{-1}] are dispersed, resulting in clogged pores and very low infiltration rates, whereas saline (ESP < 15, EC > 4 dS m^{-1}) and saline-sodic (ESP > 15, EC > 4 dS m^{-1}) soils are flocculated, promoting porosity and higher infiltration rates.

Salts that occur in arid-land soils vary in their solubilities (Table 30.1). Salts of Na, Mg, and K are more than 100 times more soluble in water than calcite and gypsum and move readily with saturated and unsaturated water flow in soil. As a result of the high solubility of Na salts and the production of hydroxide ions when carbonate ions dissolve, soils containing Na_2CO_3 have high pH values (9–12):

$$Na_2CO_3 + H_2O \rightleftharpoons 2Na^+ + HCO_3^- + OH^- \qquad (30.1)$$
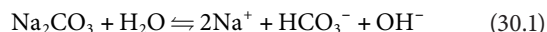
Sodium carbonate can form in sodic soils through several mechanisms. Weathering of Na aluminosilicates releases Na, which reacts with HCO_3^- in water to form $NaHCO_3$, which concentrates

TABLE 30.1 Some Common Salts in Soils and Their Solubilities in Water at 25°C

Name	Formula	Solubility (mol L^{-1})
Soda niter	$NaNO_3$	10.69
Niter	KNO_3	3.53
Halite	$NaCl$	6.15
Thenardite	Na_2SO_4	1.97
Mirabilite	$Na_2SO_4 \cdot 10H_2O$	2.74
Nahocolite	$NaHCO_3$	1.22
Soda	$Na_2CO_3 \cdot 10H_2O$	2.77
Trona	$Na_3H(CO_3)_2 \cdot 2H_2O$	2.56
Bloedite	$Na_2Mg(SO_4)_2 \cdot 4H_2O$	2.31
Hexihydrite	$MgSO_4 \cdot 6H_2O$	3.17
Epsomite	$MgSO_4 \cdot 7H_2O$	3.03
Gypsum	$CaSO_4 \cdot 2H_2O$	0.005
Calcite	$CaCO_3$	0.0006 (pH 8 and $CO_2 = 10^{-4}$ MPa)

to Na_2CO_3 under the influence of evaporation. Neutral Na salts can react with $CaCO_3$ to yield Na_2CO_3:

$$Na_2SO_4 + CaCO_3 \rightleftharpoons Na_2CO_3 + CaSO_4 \qquad (30.2)$$

and Na can be replaced from the exchange by H or Ca from carbonates:

$$[CLAY] - 2Na + CaCO_3 \rightleftharpoons [CLAY] - Ca + Na_2CO_3 \qquad (30.3)$$

30.3.2.3.2 Natural Sources of Soil Salinity

The ultimate source of salts is weathering of primary minerals (Figure 30.2), which produces dilute soil solution. Atmospheric deposition contributes appreciable quantities of salts to soils, particularly as dry fall (dust) in arid regions and as sea spray along coastal margins. Most saline soils receive salts from sources where salts have been concentrated by geological processes. For instance, dissolution of marine or closed basin lake sediments provides preconcentrated salts to near-surface waters. Isotopic studies in the Atacama Desert documented the transition from coastal soils receiving salts from marine sources to inland soils, in which salts have isotopic signatures similar to local surface and ground waters (Rech et al., 2003). The salts in groundwater are concentrated by evaporation within the capillary fringe and precipitate as evaporite minerals on playa surfaces. These evaporite minerals are highly susceptible to wind erosion (Reynolds et al., 2007), which is responsible for the dispersal of salts across the desert landscape. In a study of a Mojave Desert bolson landscape, salts were found to be concentrated in soils on slopes downwind from the playa (Hirmas and Graham, 2011).

30.3.2.3.3 Accumulation of Soluble Salts in Soils

Salts precipitate when their solubility is exceeded, usually as the soil solution is concentrated by ET. In well-drained soils where leaching is greater than ET, salts do not accumulate because the constituent ions are leached to groundwater. On the other hand, salts accumulate when leaching is minimal. Mineral leaching results from high ET rates and/or low rainfall, convex topography that disperses water flow, and soil conditions such as crusting that yield low infiltration rates. One special case of poor leaching conditions leading to salt accumulation involves the soils formed beneath desert pavement (a monolayer of closely packed clasts). The accumulation of salts beneath desert pavement corresponds strongly with the size, sorting, and percent cover of the surface clasts (Wood et al., 2005) and includes large nitrate pools in some soils in the Mojave Desert (Graham et al., 2008). In each of these cases, the areal and vertical distribution of salts reflects a balance between a downward or lateral flux of leaching water that removes salts, and the upward flux of ET that helps retain salts within the soil.

Although the dominant influence of plants on salt distribution is through their role in ET, some plants have a localized effect on soil salinity and sodicity. For example, surface soils under the canopy of grease-wood have EC values six times those of soils in interspaces between the shrubs, and ESP values are

FIGURE 30.2 Sources and redistribution of soluble salts in soils in relation to hydrology, lithology, and landscape position. When the downward leaching flux of water greatly exceeds upward flux due to ET, soluble salts, as indicated by EC of saturation extracts, are minimal throughout the soil profile. When leaching is slightly greater than ET, EC increases with depth, since salts are leached from the surface soil. When ET is greater than leaching, EC is greatest near the surface due to the predominating upward flux of water, carrying salts, to the evaporative surface.

more than 10 times greater due to biocycling of sodium ions (Fireman and Hayward, 1952).

Bare patches, known as slickspots, caused by salt accumulations, have been shown to grow due to interactions between the vegetation and soil properties (Reid et al., 1993). The grass-covered soils surrounding the slickspots have adequate infiltration rates and grass traps saline dust from a nearby soda lake

FIGURE 30.3 Diagrammatic representation of processes contributing to the salinization and expansion of slickspots on the Carrizo Plain, California. (Modified from Reid, D.A., R.C. Graham, R.J. Southard, and C. Amrhein. 1993. Slickspot soil genesis in the Carrizo Plain, California. Soil Sci. Soc. Am. J. 57:162–168. With permission of the Soil Science Society of America.)

playa (Figure 30.3). Under the slickspots, the much lower infiltration rates at the surface causes the subsoils to remain dry even during the wet season. This, in turn, establishes a matric potential gradient from the moist subsoils under grass to the slickspots, so that salts are carried laterally in solution as water is drawn by capillarity into the slickspot soils. The buildup of salts in the slickspot soils eventually has a toxic effect on the grasses at the slickspot margin, which exposes bare soil to raindrop impact and wind erosion, thereby promoting expansion of the slickspot.

30.3.2.3.4 Relation to Soil Morphology

Accumulation of soluble salts may be expressed in a morphological sequence similar to the stages of carbonate accumulation (discussed in the Section 30.3.2.4). Soils in stable landscape positions and hyperarid deserts, such as the Atacama Desert in South America and the Dry Valleys of Antarctica, can eventually develop horizons that are indurated with highly soluble minerals, including halite and soda niter (Bockheim, 1997; Rech et al., 2003). Soluble salt accumulation in moraines of the Transantarctic mountains progress with age from coatings on the bottoms of stones (stage I), to flecks (stages II–III), to a cemented pan (stages IV–V), and finally an indurated pan (stage VI) (Bockheim, 1990). In some cases, the morphology of salt accumulations may be a clue to their composition. Concentrations of gypsum may be identified in the field by a unique morphology, called gypsum snowballs, which are 0.5–1 mm, white, powdery spheres (Buck and Van Hoesen, 2002).

Exchangeable sodium can influence soil morphology by causing clay dispersion, which decreases aggregate stability and increases mobility of clays. Illuvial clay accumulates more rapidly in sodic soils, compared to nonsodic soils, thus complicating the relationships between soil morphology and surface age (Peterson, 1980). Strong development of vesicular horizons has been associated with sodic soils, due to dispersion and crust formation (Bouza et al., 1993).

The contrast between leached soils and salt-affected soils is well illustrated by the comparison of soils in slickspots with the surrounding vegetated landscape (Reid et al., 1993). The grass-covered soils have clayey Bt horizons with strong prismatic structure, which allows rapid infiltration of rainwater and leaching of salts. In these soils, less-soluble minerals occur near the surface and more-soluble minerals are concentrated deeper in the profile. With increasing depths, accumulations of calcite, gypsum, and soluble salts are observed, indicating a regime in which leaching predominates over ET. In contrast, <1 m away, the bare surface of the slickspot is crusted from raindrop impact, and the upper few centimeters are sufficiently leached of salts, so that the soil disperses. Crusting and dispersion result in very low infiltration rates, so that leaching of the profile is minimal. The predominance of evaporation over leaching is indicated by the accumulation of soluble salts above gypsum, and calcite accumulation deeper in the soil a trend opposite that in grassed areas (Figure 30.3).

30.3.2.4 Accumulation and Redistribution of Calcite

Where soil pH values are above about 7.2, $CaCO_3$ can precipitate in soils to form the mineral species calcite. Calcite contains <5 mol% Mg. At 25°C, pH 8, and P_{CO_2} of 10^{-4} MPa, calcite has a solubility of 0.0006 mol L^{-1} (0.06 g L^{-1}), which is much more soluble than silicate minerals but is considerably less soluble than the salts commonly found in soils. Under well-drained conditions, calcite is found in soils of arid, semiarid, and subhumid environments. With progressively greater leaching, it is found at greater depths below the soil surface. It can also accumulate in poorly drained soils where calcium and bicarbonate are concentrated in ponded water.

30.3.2.4.1 Natural Sources of Calcite in Soils

While calcite may be inherited in soils from parent materials such as limestone, pedogenic calcite precipitates from soil solution and requires sources of Ca^{2+} and CO_3^{2-} ions. Calcium ions may be derived from weathering of parent materials. Calcareous rocks provide the most abundant Ca, with limestone containing 30%–40% Ca and dolomite containing about 20% Ca. Calcareous tills, loess, and other sediments also provide a ready supply of Ca for pedogenic calcite formation. Calcium is present many in primary minerals, particularly Ca-rich plagioclase, but also in some amphiboles, pyroxenes, garnets, and epidote. Among igneous rocks, Ca is most abundant in the more mafic (but not ultramafic) rocks, such as gabbro and basalt, which contain about 7% Ca. Mafic rocks are also rich in Mg and soils formed on basalt may contain pedogenic high-Mg calcite and/or dolomite (Capo et al., 2000; Whipkey et al., 2002). Atmospheric deposition can be a substantial, even dominant, source of Ca, both in ionic form and as particulate $CaCO_3$. For example, in southern Nevada and California, dust contains 10%–30% $CaCO_3$, equating to a deposition rate of 1–6.6 g $CaCO_3$ m^{-2} $year^{-1}$ (Reheis and Kihl, 1995). In southern New Mexico, dust contributes about 0.4 g $CaCO_3$ m^{-2} $year^{-1}$, whereas rain delivers 1.2 g m^{-2} $year^{-1}$ (Birkeland, 1999). On the Edwards Plateau in Texas, no $CaCO_3$ is delivered

as dust, but rain supplies Ca equivalent to 2.3 g $CaCO_3$ m^{-2} $year^{-1}$ (Rabenhorst et al., 1984). In southern Arizona, isotopic studies have suggested that most $CaCO_3$ in dust actually comes from the erosion of soil carbonates; thus, dust can be viewed as a recycled source of Ca (Capo and Chadwick, 1999; Naiman et al., 2000). Pedogenic carbonates have a Sr isotope signature between that of dust and parent rock, indicating a mixed source of Ca. Calcium can also be derived from biocycled plant material and groundwater, which may be highly calcareous, depending on the lithologies exposed in the aquifer. When Ca-laden groundwater lies within or just below the soil zone, calcite can form in the soil as it is precipitated within the capillary fringe. At sites near the coast, Sr isotope ratios in pedogenic carbonates are similar to those of ocean water, suggesting that some Ca may come from sea spray (Naiman et al., 2000; Whipkey et al., 2000).

The source of the carbonate anion in calcite is the dissolution of CO_2 in water, which yields the following species:

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \qquad (30.4)$$

$$H_2CO_3 \rightleftharpoons HCO_3 + H^+ \quad pK = 6.4 \qquad (30.5)$$

$$HCO_3 \rightleftharpoons CO_3^{2-} + H^+ \quad pK = 10.3 \qquad (30.6)$$
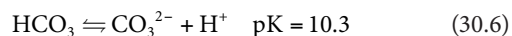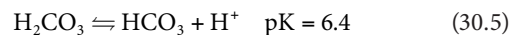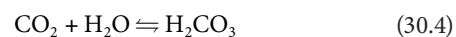
The amount of CO_2 in the soil solution depends on the relationships expressed in Henry's law:

$$K_H = \frac{(H_2CO_3)}{P_{CO_2}} = 10^{-1.5} \qquad (30.7)$$

An increase in the proportion of CO_2 in the gas phase increases the CO_2 concentration in the solution in contact with the gas. The P_{CO_2} in soils is on the order of 0.003–0.03, compared to 0.00033 in the atmosphere. The distribution of CO_2 in soils depends on a balance between production of CO_2 within the soil, by microbes and plant roots, and diffusion losses of CO_2 out of the soil to the atmosphere. As an example, the CO_2 concentration in a coarse-textured soil in southern Nevada was greatest in spring when biologic activity was at its peak (Figure 30.4; Terhune and Harden, 1991). Even at this time, CO_2 concentration in the surface soil was less than that in the subsoil due to diffusional losses to the atmosphere. Biologic activity and CO_2 production were low during the winter due to cold temperatures and during summer due to dryness.

30.3.2.4.2 Calcite Precipitation in Soils

The reaction for calcite precipitation can be expressed as follows:

$$Ca^{2+} + 2HCO_3 \rightleftharpoons CaCO_3 + H_2O + CO_2 \qquad (30.8)$$
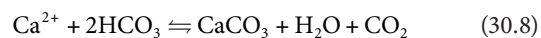
Precipitation of calcite is promoted by a number of factors that influence this relationship. An increase in pH drives the reaction to the right by supplying more HCO_3^-, as a result of $CO_2 + OH^-$. Decreasing the P_{CO_2}, through the loss of CO_2 to the atmosphere,
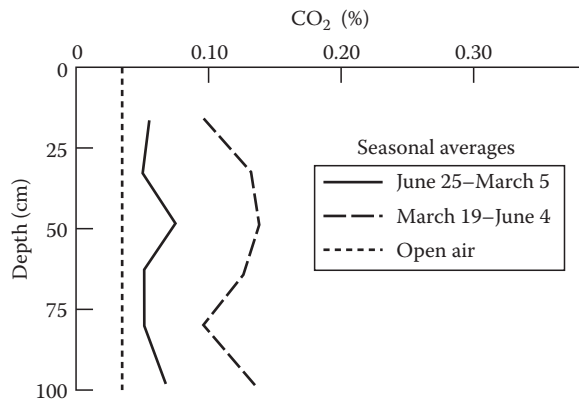
FIGURE 30.4 Seasonal changes of CO_2 with depth in an Entisol in southwestern Nevada. (Reprinted from Terhune, C.L., and J.W. Harden. 1991. Seasonal variations of carbon dioxide concentrations in stony, coarse-textured desert soils of southern Nevada, USA. Soil Sci. 151:417–429. Copyright Williams and Wilkins, Baltimore, MD. With permission.)

FIGURE 30.5 Idealized diagram of processes involved in calcite precipitation in (a) a well-drained soil (water flux is downward) and (b) a poorly drained soil (water flux is upward).

also drives the reaction to the right. Loss of water through ET increases the ionic concentration of the soil solution to the point where it exceeds the solubility product of calcite and results in precipitation. Precipitation of calcite in soils is not simply an inorganic chemical phenomenon. In fact, the role of microorganisms seems to be ubiquitous and perhaps essential in pedogenic calcite formation. Microbial influence is revealed by calcified biological structures, such as fungal hyphae, and experimental evidence that shows that both bacteria and, especially, fungi produce calcite as a byproduct of their metabolism. In Ca-rich soils, microbes excrete excess Ca, which concentrates on their external surfaces and reacts with HCO_3^- in the soil solution. In the laboratory, calcite precipitates rapidly in inoculated soil columns but not at all in sterile columns that were otherwise identical (Monger et al., 1991). Hyperarid regions of the Atacama Desert and Antarctica, in which no plant life is observed, have been found to contain only trace levels of $CaCO_3$ (Bockheim, 1990; Ewing et al., 2006; Quade et al., 2007). The isotopic signature of $CaCO_3$ in soils without plant life in the Atacama Desert is distinct from that in soils with plant life, suggesting possible low levels of abiotic precipitation.

Precipitation of calcite in soils occurs preferentially in certain types of morphologic sites. These favored sites include the vicinity of roots, where microbes and nucleation sites are abundant and plant water uptake concentrates the soil solution. Another favorable site is in large pores where drying is relatively rapid and the P_{CO_2} is lower than in the matrix due to effective gas exchange with the atmosphere. In gravelly soils, calcite tends to precipitate preferentially on the undersides of gravels, perhaps because water tends to collect and is last to dry in those sites. Once calcite precipitation is initiated at any of the sites described above, it is preferentially precipitated on preexisting calcite crystals.

30.3.2.4.3 Examples of Calcite Distribution and Precipitation in Soils

The processes governing precipitation of calcite in a well-drained soil, where water flux is downward, are illustrated in Figure 30.5a.
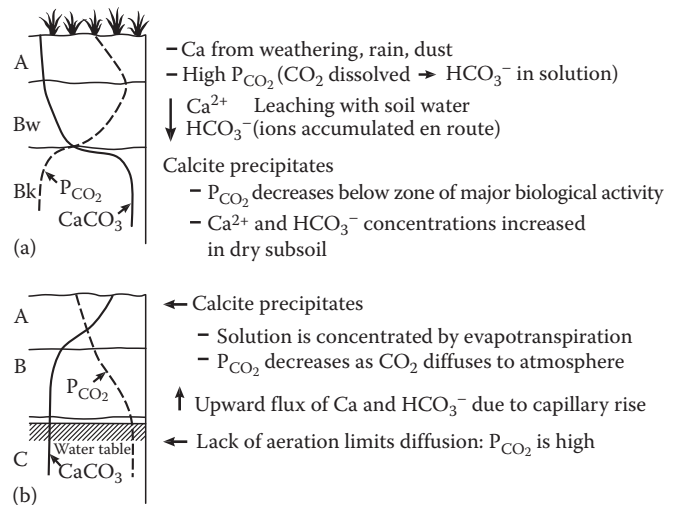
Calcium in the soil originates from weathering, rain, and/or dissolution of dust. Organic matter decomposition in the A horizon produces a high P_{CO_2}, although it decreases substantially near the soil surface due to gas exchange with the atmosphere. The high P_{CO_2} yields a relatively large amount of CO_2 dissolved in the soil solution, producing HCO_3^-. The Ca^{2+} and HCO_3^- ions are leached with soil water to a depth at which calcite precipitates. This depth is determined by decreased P_{CO_2} below the zone of major biological activity and, perhaps more importantly, by the soil solution being concentrated as it enters the dry subsoil and is depleted by ET.

In poorly drained soils of arid and semiarid regions, dominant water flux is upward in response to evaporation (Figure 30.5b). The relatively high water content in subsoil limits gas diffusion, so P_{CO_2} is high and decreases near the soil surface as CO_2 diffuses to the atmosphere. Calcium and HCO_3^- originating from groundwater sources or in situ from the soil move upward with the flux of water. Calcite precipitates in an upper soil zone where the P_{CO_2} is relatively low and the soil solution is concentrated by evaporation.

30.3.2.4.4 Relation to Soil Morphology

Soils with pedogenic calcite commonly progress through distinct evolutionary stages (Table 30.2; Gile et al., 1966). The stages and morphologic expressions are different for gravelly soils compared to fine-textured soils. Because gravelly soils have less total pore space than fine-textured soils, segregated calcite accumulates more rapidly, precipitating readily on pebbles bounded by relatively large pores. One of the distinct morphologic expressions of pedogenic $CaCO_3$ in gravelly soils is pendants precipitated on the bottom of clasts. The pendants grow by precipitation of $CaCO_3$ both at the clast–pendant contact and at the base of the pendant (Brock and Buck, 2005).

TABLE 30.2 Stages of Carbonate Morphology in Soils

Stage	Gravelly Parent Material	Nongravelly Parent Material
I	Thin discontinuous clast coatings; some filaments; matrix can be calcareous next to stones; about 4% $CaCO_3$	Few filaments or coatings on sand grains; <10% $CaCO_3$
I+	Many or all clast coatings are thin and continuous	Filaments are common
II	Continuous clast coatings; local cementation of few to several clasts; matrix is loose and calcareous enough to give somewhat whitened appearance	Few to common nodules; matrix between nodule is slightly whitened by carbonate (15%–50% by area), and the latter occurs in veinlets and as filaments; some matrix can be noncalcareous about 10%–15% $CaCO_3$ in whole sample, 15%–75% in nodules
II+	Same as stage II, except carbonate in matrix is more pervasive	Common nodules, 50%–90% of matrix is whitened; about 15% $CaCO_3$ in whole sample
Continuity of fabric high in carbonate		
III	Carbonate forms an essentially continuous medium in 50%–90% of horizon; color mostly white; carbonate-rich layers more common in upper part; about 20%–25% $CaCO_3$	Many nodules, and carbonate coats many grains such that over 90% of horizon is white; carbonate-rich layers are more common in upper part; about 20% $CaCO_3$ in whole sample
III+	Most clasts have thick carbonate coats; matrix particles continuously coated with carbonate or pores plugged by carbonate; cementation more or less continuous; >40% $CaCO_3$	Most grains coated with carbonate; most pores plugged; >40% $CaCO_3$ in whole sample
Partly or entirely cemented		
IV	Upper part of horizon is nearly pure cemented carbonate (75%–90% $CaCO_3$) and has a weak platy structure due to the weakly expressed laminar depositional layers of carbonate; the rest of the horizon is plugged with carbonate (50%–75% $CaCO_3$)	
V	Laminar layer and platy structure are strongly expressed incipient brecciation and pisolith (thin, multiple layers of carbonate surrounding particles) formation	
VI	Brecciation and recementation, as well as pisoliths, are common	

Source: After Birkeland, P.W. 1999. Soils and geomorphology. 3rd edn. Oxford University Press, New York. By permission of Oxford University Press.

In fine-textured soils, calcite precipitates first as filaments in root pores, then as soft masses. Eventually a calcite-cemented (petrocalcic) horizon forms, in which carbonate plugs nearly all the pores and laminar plates of carbonate build up at the interface between the overlying horizons and the plugged ones. Formation of a petrocalcic horizon (stage IV or greater) has major implications for the direction of further development. Plugged horizons form a barrier to most root penetration. They also resist erosion and the overlying horizons may be stripped off, leaving the petrocalcic horizon exposed at the surface. Burrowing animals and erosion commonly bring fragments of petrocalcic horizons to the surface, where the calcite is dissolved and recycled back into the soil (Eghbal and Southard, 1993).

30.3.2.5 Accumulation and Redistribution of Silica

After oxygen, Si is the most abundant element in Earth's crust. It exists in many primary mineral forms and is incorporated into soil-formed minerals as well. When Si is released by weathering, it is leached and can be lost or can be incorporated into aluminosilicate clay minerals, such as kaolinite or smectite. In humid environments, most Si is either leached or consumed during clay mineral synthesis. In subhumid and semiarid environments where there is enough water to support mineral weathering but not enough to completely leach Si, it can also precipitate as nanocrystalline silica that cements soil horizons.

30.3.2.5.1 Characteristics of Silica in Soils

Silica refers to compounds consisting of SiO_2 in crystalline, poorly crystalline, or amorphous forms, and is sometimes hydrated to some degree. Quartz is a common form of silica in many soils. It is highly crystalline SiO_2 that is inherited from parent materials and usually does not form pedogenically. Opal-A, the common pedogenic and biogenic silica, is a hydrated, x-ray amorphous, form of SiO_2. A somewhat more crystalline form of opal, opal-CT, may form in some very old soils, but it is not nearly as common as opal-A.

30.3.2.5.2 Sources of Soluble Silicon

The weathering of silicate minerals releases Si to solution that can be precipitated, under appropriate conditions, as opal-A. Soils rich in easily weathered silicates, such as olivine ([Mg, Fe]$_2$SiO$_4$) and anorthite (CaAl$_2$Si$_2$O$_8$), release abundant Si into solution. The accumulation of soluble silica in soils of southern California was attributed to feldspars weathering in the upper horizons (Kendrick and Graham, 2004). More resistant silicates, such as quartz, release Si very slowly. Amorphous silica is much more soluble than crystalline forms (Figure 30.6). Volcanic glass is a primary form of amorphous silica, which is abundant in many soils or parent materials, sometimes even in those far from current volcanic activity, since volcanic ash can travel great distances downwind from eruptions. Another source of amorphous silica is
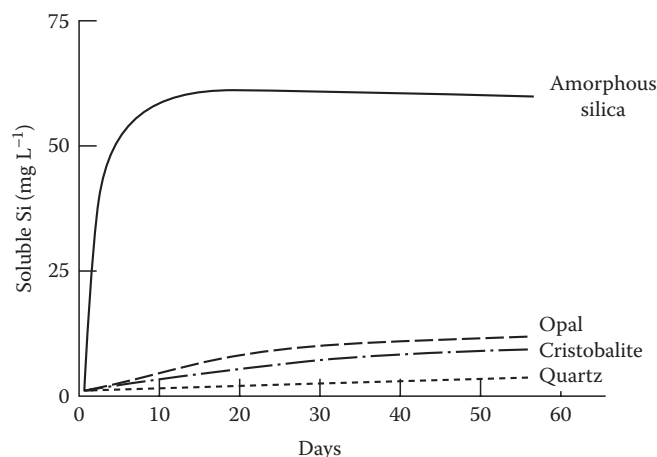
FIGURE 30.6 Dissolution of silicon in water from various silica phases as a function of time. (Reprinted from Drees, L.R., L.P. Wilding, N.E. Smeck, and A.L. Senkayi. 1989. Silica in soils: Quartz and disordered silica polymorphs, p. 913–974. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI. With permission of the Soil Science Society of America.)

biogenic opal, which is actually opal-A produced as part of plant structures, known as phytoliths, or as part of aquatic organisms, such as diatom tests or sponge spicules. Phytoliths are a biocycled form of silica in soils and can be very abundant in A horizons, particularly in grasslands. In soils that are depleted of primary mineral silica, weathering of phytoliths provides a significant portion of the $Si(OH)_4$ in solution (Alexandre et al., 1997; Derry et al., 2005). Diatoms and sponge spicules are most often found in soils derived from lacustrine, deltaic, or floodplain sediments.

30.3.2.5.3 Dissolution of Silica

Release of Si into solution from a solid is controlled by a number of factors. The inherent solubility of the mineral or amorphous compounds varies widely, with opal-A being the most soluble source of silica (Figure 30.6). Solubility also increases dramatically for very small particles ($<0.01\,\mu m$). External conditions play a large role, with Si solubilization increasing when the soil solution has a low ionic strength, high pH (>9), and relatively high temperature (Drees et al., 1989). Organic acids promote dissolution by complexing $Si(OH)_4$ released into solution, resulting in highest silica dissolution potentials in the root zone. Sorption of Si onto Fe/Al (hydr)oxides can make them a sink for soluble Si, thereby keeping solution concentrations of Si low and increasing dissolution. On the other hand, coatings of organic materials or Fe/Al (hydr)oxides can retard dissolution by isolating the reactive surface from the soil solution. Mobilization of silica in soils of the Pacific Northwest was decreased in response to management practices that promote accretion of organic carbon, presumably due to organic matter coatings on reactive surfaces (Gollany et al., 2005). Solubility of silica increases with increased pressure, a situation that may arise in petrocalcic horizons as calcite crystallization increases the subsoil volume and generates grain-to-grain pressure contacts (Monger and Daugherty, 1991).

30.3.2.5.4 Translocation and Precipitation of Silicon

In solution, Si exists mostly as silicic acid $[Si(OH)_4]$. It moves with percolating water and precipitates when conditions are favorable. The soil pH influences the way in which silica precipitates. At pH < 7, $Si(OH)_4$ precipitates as SiO_2 on adsorption sites as individual molecules or low-molecular-weight polymers, whereas at pH 7–10, Si in solution takes the form of high-molecular-weight polymers, and these are flocculated by cations in solution. Other factors may alter, or override, the effect of pH on Si precipitation in soils. Silica precipitation is promoted in soil solutions with high ionic strength and adsorption is promoted by a solid phase with a high surface area. Release, leaching, and precipitation of Si favors preservation of ^{28}Si in the solid phase of soils, creating strong negative shifts in $\partial^{30}Si$ signatures in soils during weathering (Ziegler et al., 2005a, 2005b).

In soils, $Si(OH)_4$ is adsorbed on exposed hydroxyl groups of clays, (hydr)oxides, and primary silicates (Figure 30.7). A high surface area (e.g., argillic horizon) or ionic strength (e.g., calcic horizon) promotes rapid adsorption of silica as individual or low-molecular-weight polymers. Adsorption is greatest at pH 7–9. Drying dehydrates the adsorbed $Si(OH)_4$ causing amorphous SiO_2 to precipitate on the surface (Figure 30.7a). The strong Si–O bond inhibits rehydration and desorption. The precipitated silica acts as a template for further precipitation (Figure 30.7b) and, eventually, opaline silica forms bridges between the grains on which it precipitates (Figure 30.7c; Chadwick et al., 1987).
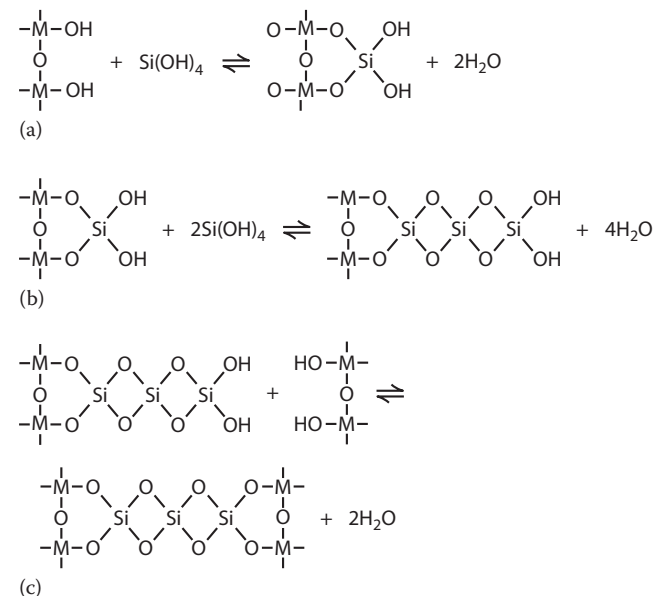
FIGURE 30.7 Model of silica precipitation in soils (a) reversible $Si(OH)_4$ adsorption on Fe and Al (hydr)oxides, silicates, and aluminosilicates (M may be Al, Fe, Mg, or Si); (b) further $Si(OH)_4$ adsorption followed by dehydration to SiO_2 during soil drying decreases the reversibility of the original adsorption process; (c) two soil components with $Si(OH)_4$ adsorption surfaces may be bonded together by opaline SiO_2 bridges. (Reprinted from Chadwick, O.A., D.M. Hendricks, and W.D. Nettleton. 1987. Silica in duric soils. 1. A depositional model. Soil Sci. Soc. Am. J. 51:975–982. With permission of the Soil Science Society of America.)

In fine-textured soil horizons, silica diffuses into smaller voids in the soil matrix, where there is more surface area for adsorption, and precipitates as polymerized layers. The zone of accumulation tends to correspond to the zone of clay accumulation because of the abundant sites for precipitation there (Kendrick and Graham, 2004). The silica polymers are found as flocs (≈1 μm diameter) on grains in the soil matrix and interlaminated with clay skins. The same mechanisms apply for coarse-textured soils, but silica not in direct contact with adsorption surfaces may precipitate by flocculation if the ionic strength is sufficiently high (Chadwick et al., 1987). Opaline silica may also precipitate preferentially on the undersides of gravels, even forming pendants (Munk and Southard, 1993).

In many soils of arid and semiarid regions, the depth distributions of silica and calcite overlap, but these pedogenic compounds are concentrated in different microsites by nature of their mode of precipitation. Calcite precipitates by self-nucleation and microbial processes, particularly in larger voids because these dry early, concentrating the soil solution. Larger voids also often have the most effective contact with the atmosphere, so the P_{CO_2} is relatively low. Silica specifically bonds with Al and Fe hydroxy compounds, whereas, under most soil conditions, calcium does not. Thus, silica precipitates primarily by adsorption on surfaces and so is found in the matrix, coating grains, and clay skins, rather than in voids. Furthermore, there is no direct chemical bonding between Ca and $Si(OH)_4$ at pH < 9, since $Si(OH)_4$ is neutral and nonpolar and cannot compete with polar molecules or anions for adsorption on calcite (Chadwick et al., 1987). Coatings of opal also occur as silans in macro voids when the soil matrix is plugged with calcite and silica cements.

30.3.2.5.5 Relation to Soil Morphology

A soil horizon that is thoroughly cemented by silica, such that it will not slake in water or HCl, is known as a duripan. Since the silica needs only to cement the matrix by bridging between grains, cementation can occur with very low concentrations of silica (e.g., ≈4%), particularly in coarse-textured soils. In a chronosequence study of silica accumulation in soils, an increase in rupture resistance was one of the first signs of silica precipitation, observed in 55 ka soils with no macroscopically visible silica precipitation (Kendrick and Graham, 2004). Duripans often have accessory cementing agents, such as Fe (hydr)oxides or calcite. Duripans exhibit two forms corresponding largely to the regional climate. In arid regions, duripans are commonly platy, with plates 1–15 cm thick. Pores and plates are coated with opal and are usually engulfed with pedogenic calcite. In Mediterranean climates, duripans take a different form in which opal coats the faces and pores of very coarse to extremely coarse (0.3–3 m diameter) prisms that make up the subsoil. The upper boundary is abrupt and the pan may have an opal coating on top in the strongly developed cases. Water often perches on top of the pan during the rainy season, so that Fe/Mn (hydr)oxide nodules and illuvial clay accumulate there. The matrix of this kind of duripan is extremely hard when dry but is brittle when moist and can be penetrated, with some difficulty, with a hand auger. Pedogenic calcite may, or may not, be present, but it is not a dominant component.

Reversible, weakly expressed silica cementation seems to play a role in the formation of fragipans. These horizons are dense and appear cemented when dry but will slake in water (Franzmeier et al., 1989; Marsan and Torrent, 1989). The strength of fragipans can be related to the ratio of $Si_d/(Si_d + Al_d)$, suggesting that an accumulation Si under low Al conditions is responsible for the hardening of fragipans, due to the precipitation of amorphous silica rather than aluminosilicates (Duncan and Franzmeier, 1999).

Silica cementation is often enhanced in soils near scarps or incisions, such as at terrace edges. Evaporation at the scarp surface concentrates the soil solution and promotes silica precipitation. The resulting induration impacts water movement and slope retreat, and renders the soil profile exposed along the scarp atypical of the soil under the broader geomorphic surface (Moody and Graham, 1997). Scarps in terrain with silica-cemented soils are often nearly vertical because cementation strengthens these edges (Kendrick and Graham, 2004).

30.3.2.6 Accumulation and Redistribution of Clay

Clay-sized particles have a dominant influence on soil properties due to their large, reactive surface area. Surfaces of clay minerals play a role in water retention, nutrient storage and exchange, and precipitation of secondary minerals. The redistribution of clay is a widely observed pedogenic process, which results in a horizon of clay accumulation (Bt horizon). The development of a Bt horizon, in turn, enhances other pedogenic processes, such as the reinforcement of soil structure and the precipitation of amorphous silica.

30.3.2.6.1 Sources of Clay

Clay (inorganic particles <2 μm diameter) is produced by weathering of primary mineral grains such as feldspars, micas, pyroxenes, and amphiboles. The ions released by weathering are carried in solution to other parts of the soil profile or the landscape and precipitated as colloids such as smectite, gibbsite, kaolin, or Fe (hydr)oxides. Commonly, as primary grains are weathered, secondary minerals replace them pseudomorphically. In this case, the secondary minerals are not initially dispersed as colloids in the soil (Nahon, 1991); they enter the clay size fraction when pseudomorph grains are crushed by turbation processes (Graham and Buol, 1990). Clay contained in some parent materials, especially sediments (e.g., alluvium, colluvium, lacustrine deposits, loess) and sedimentary rocks (e.g., shale, limestone, mudstone), is simply inherited in the soils derived from them. Another significant source of clay that is particularly important in arid regions is aerosolic input (Reheis and Kihl, 1995; Simonson, 1995). In the eastern Sierra, Nevada, California, spatial variability in eolian inputs helps to explain why pedogenic clay appears to increase with moraine age while other pedogenic processes are strongly confounded by erosion of the moraines (Rossi, 2009). In this chronosequence, older moraines are located closer to the valley floor, from which most

eolian material is derived, and, therefore, the oldest moraines receive the greatest inputs of eolian-derived clays.

30.3.2.6.2 Mobilization and Transport of Clay

As the smallest particle-size fraction, clay is most susceptible to suspension in and transport by water percolating through the soil. The mobility of clay particles depends on the characteristics of the clay particles, the pore water chemistry, and the physical nature of water movement through soil. In general, the finest clay particles are more readily transported, moving even at low pore water flow rates (Kaplan et al., 1993), but even silt particles move in suspension if conditions are favorable. Clay-sized minerals with relatively strong negative charge, such as smectite, are most mobile.

Dispersion, the chemical repulsion of colloids in suspension, greatly enhances clay movement. Chemical conditions that favor dispersion in soils are those that result in an expanded diffuse double layer (Chapter 15), most notably Na-saturated clay and absence of soluble salts or partially soluble minerals (e.g., calcite, gypsum) that would increase the electrolyte concentration in solution or put Ca on exchange sites at the expense of Na. Dispersion is also enhanced by high pH because it imparts a negative charge to variable charge materials (e.g., (hydr)oxides, kaolin), which are then repulsed by each other and by the permanently negatively charged minerals. An example of this in natural systems is the effect of wood ash from forest fires, which may raise the surface soil pH above 10 (Ulery et al., 1993), causing kaolinite to disperse (Durgin and Vogelsang, 1984). Low-molecular-weight organic acids can be effective dispersive agents (Jenny and Smith, 1935; Kaplan et al., 1993, 1997; Kretzschmar et al., 1993, 1995). They complex cations in solution, thereby keeping them from flocculating clays, and they can specifically adsorb to positively charged mineral edges, preventing edge-to-face bonding with negatively charged minerals (Durgin and Chaney, 1984; Heil and Sposito, 1993a, 1993b). On the other hand, dispersion is inhibited by processes that produce stable aggregates. Thus, interparticle bonding that is enhanced by organic matter and noncrystalline inorganic compounds minimizes clay redistribution.

Mobilization and transport of clay particles by water is affected by the velocity of pore water flow and by the wetting and drying of pores. Slaking, which is the physical detachment of clay, is enhanced by increased pore water flow rates because it increases shear stress needed to detach particles from the matrix (Kaplan et al., 1993). Water flow through macropores occurs at greater velocities than micropore flow. As a result, macropore flow can move more clay and larger particles to greater depths. In fact, most clay movement is driven by macroscopic flow, even though it occurs less frequently than micropore flow. Wetting–drying cycles also enhance slaking. As a soil rewets, moving water can dislodge loose particles and carry them in suspension (Hudson, 1977). Another mechanism by which wetting–drying cycles promote clay transport is enrichment of clays with pH-dependent charge at the air–water interface in partially drained pores (Wan and Tokunaga, 2002). This partitioning results in a concentration of colloids that are readily mobilized during the next wetting cycle (DeNovio et al., 2004).

30.3.2.6.3 Deposition of Clay

Physical processes that cause a deposition of clays include immobilization by physicochemical interaction with grain surfaces or immobile air–water interfaces and straining of large clays by fine pores or thin films of water on grain surfaces (DeNovio et al., 2004). Deposition of clays may be promoted by chemical factors such as flocculation by high electrolyte concentrations as in saline or calcareous subsoils and adsorption of phyllosilicate clays with permanent negative charge to Fe and Al (hydr)oxides in acid subsoils (Jenny and Smith, 1935). Deposition of clay frequently occurs as the suspension travels into a drier part of the soil where the moving water is imbibed and retained by capillary forces. For example, when a suspension of clay flows through a macropore (Figure 30.8), water is pulled by matric forces into the soil fabric and the clay in suspension is deposited in or on the pore wall. Deposition of this type occurs in tubular, inter-pedal, and intergranular pores, and results in the characteristic parallel orientation of platy phyllosilicate clays on pore walls. Depositions of clay may be promoted by textural contrasts. Interruption of gravitational water flow as by a fine layer overlying a coarse layer leads to the deposition of colloidal material from the leaching water above the contact (Bartelli and Odell, 1960). The accumulation of clay above the contact then serves to reinforce the textural contrast, creating a positive feedback (Schaetzl and Anderson, 2005).

ψ_m decreases
(soil is drier)

1. Water flows down pore carrying clay in suspension

2. Drier subsoil is encountered

3. Matric forces pull infiltrating water out of pore, into matrix

4. Clay from suspension is plastered on pore walls

5. Phyllosilicates have *c*-axis perpendicular to pore-walls
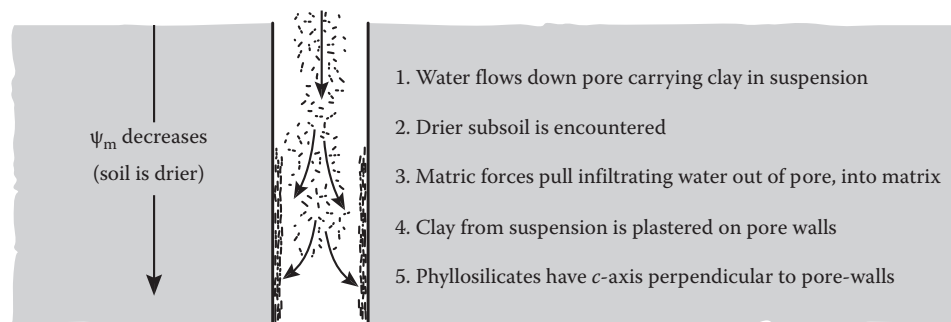
FIGURE 30.8 Illustration of clay deposition from suspension in a tubular pore to form an oriented clay lining (i.e., clay film, channel illuviation argillan).

30.3.2.6.4 Relation to Soil Morphology

Oriented clay coatings, known as clay skins, clay films, clay linings, or argillans, are recognized visually by a discernable thickness and a darker color compared to the inside of the soil peds. Clay skins form rapidly in the laboratory (Dalrymple and Theocharopoulos, 1984) and develop within several decades in the field if pedoturbation is minimal (Graham and Wood, 1991). They are often, but not always, enriched in finer clay sizes than the soil matrix as a whole. Clay coatings are typically most abundant in subsoils (Bt horizons) but are often best preserved and expressed within the fractures of underlying saprolite or bedrock (Graham et al., 1994). They may also occur in surface horizons in which eolian deposition provides clay that is then transported by water along ped and pore surfaces (Sullivan and Koppi, 1991). In sandy, quaternary-aged soils, illuvial clay may accumulate in bands called lamellae (Rawling, 2000). The deposition of clays in lamellae has been attributed to several factors including wetting-front-drying, chemical flocculation, and sieving caused by slight variations in texture of the parent material.

Clay deposition is a strong contributor to soil structure formation by creating distinct differences between the surface and interior of peds. Deposited clays on ped surfaces are strongly oriented and can be discerned easily using a petrographic microscope and cross-polarized light. Clay skins are destroyed in soils having extensive shrink/swell activity; they are replaced by shiny slickensides or pressure faces on ped surfaces, which can be distinguished microscopically from clay skins by a different pattern of clay orientation (Nettleton et al., 1969).

30.3.2.7 Complexation and Redistribution of Fe and Al

In many soils, particularly Spodosols, dissolved organic matter plays a critical role in pedogenic processes by complexing metals, predominantly Fe and Al, in surface horizons, translocating them, and depositing them in subsoils. This process enhances a distinctive style of mineral weathering, in which chelation removes weathering products, and produces a characteristic soil morphology, epitomized by an albic horizon overlying a spodic horizon (Section 33.9).

30.3.2.7.1 Reactive Agents and Sources

Dissolved organic acids act as the carriers of metal cations. Some are simple acids derived directly from leachates of relatively fresh plant material, either from the vegetative canopy or from the leaf litter at the soil surface. Typically, the more important dissolved organic acids in soils are the byproducts of microbial decomposition of organic matter produced in the O or A horizons. They are complex, heterogeneous, relatively low-molecular-weight organic acids referred to as fulvic acids (Chapter 11). Organic acids chelate and remove cations from the surface of mineral grains. This type of weathering is very effective, since it leaves a fresh grain surface with no coatings of secondary minerals to impede solution access to the surface, a condition that inhibits further weathering. Mycorrhizal fungi also promote weathering of primary minerals in the E horizon and may transport Fe and Al to the O horizon, where concentrations of dissolved organic carbon (DOC) are highest (Lundström et al., 2000). Since organic acids keep soil solution pH below the pK_{a1} of H_2CO_3 (6.4), bicarbonate weathering is not involved in these systems (Ugolini and Spaltenstein, 1992). Iron and Al are the cations preferentially removed by organic complexation since, being relatively small cations with high valence, they form the most stable chelates (Schnitzer, 1969). At a given pH, Fe and Al will remain in solution at a much higher concentration if organically complexed.

30.3.2.7.2 Translocation and Accumulation

Chelates, forming primarily in the O and A horizons, move in solution with percolating water. During intense leaching episodes, organometal colloids, as well as dissolved metal chelates, may be flushed downward in the profile (Stoner and Ugolini, 1988). The dissolved chelates precipitate when the metal:organic carbon ratio exceeds a critical value at which all polar-bonding sites are full. The precipitation of dissolved organic matter has been studied in the laboratory, by titrating extracts from organic horizons with solutions of Fe and Al (Nierop et al., 2002). Precipitation of dissolved organic matter by Fe^{3+} occurs at lower metal:organic carbon ratios and shows less pH dependence compared to precipitation by Al^{3+} (Figure 30.9). At higher
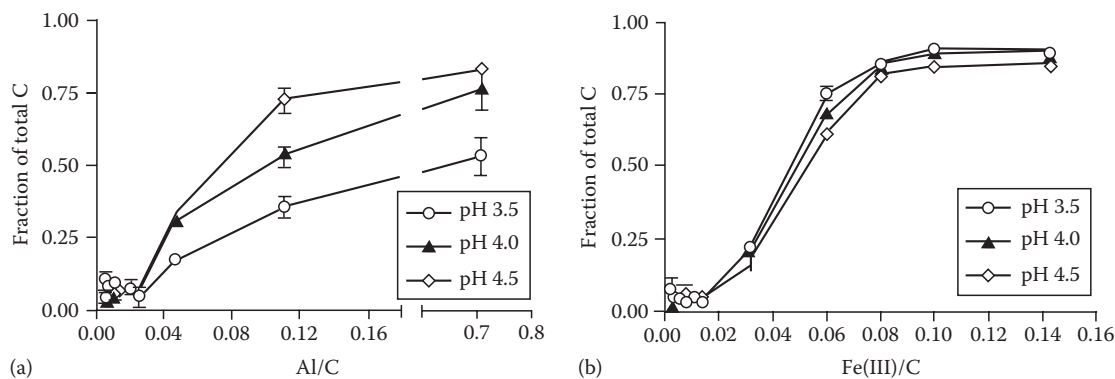
FIGURE 30.9 Fraction of total organic C precipitated form organic horizon extract solutions as a function of: (a) the Al/C molar ratio and (b) the Fe(III)/C molar ratio, at pH 3.5, 4.0, and 4.5, illustrating the precipitation of C in response to high metal-to-C ratios. (Reprinted from Nierop, K.G.J., B. Jansen, and J.A. Verstraten. 2002. Dissolved organic matter, aluminium and iron interactions: Precipitation induced by metal/carbon ratio, pH and competition. Sci. Total Environ. 300:201–211. Copyright (2002), with permission from Elsevier.)

pH, there is less proton competition, which leads to more metal interaction with dissolved organic matter and thus more precipitation of the organic–metal complexes (Nierop et al., 2002). The pH dependence of chelate solubility may explain why translocation of Fe and Al by dissolved organic matter does not occur in calcareous soils until leaching has lowered the soil pH to 5.0 or 5.5 (Schaetzl, 1996). In the field setting, chelates may be arrested even when the metal:carbon ratio is low, by adsorbing on positively charged material, such as Fe, Al (hydr) oxides, or high metal:organic C material already precipitated. Such reactions cause polymerization of soluble low-molecular-weight compounds into insoluble forms. Chelates may also be deposited by desiccation or by aggregation in a zone of relatively high ionic strength, and low H⁺ activity, exposed negative charges. Alternatively, precipitation may occur because microbial degradation of the organic ligands releases the Fe and Al from soluble organometal complexes (Lundström et al., 2000). Spodic horizons are often identified by microscopic examination of thin sections that reveal the presence of organic and Fe (hydr)oxide-rich silt-size aggregates, many with cracked coatings indicative of postdepositional dehydration (Deconinck, 1980). Fe and Al accumulate in the spodic horizons as poorly crystalline minerals, ferrihydrite and imogolite (Lundström et al., 2000).

30.3.2.7.3 Relation to Soil Morphology

The O horizon is the major source of dissolved organic acids, which strip Fe and Al from the mineral soil as they move downward (Table 30.3). This produces the bleached E (≈albic) horizon, which exhibits the colors of the fresh primary mineral grains. The zone in which the chelates are deposited is the Bhs (or Bh, or Bs; ≈spodic) horizon. Localized intensified leaching associated with tree-throw pits or macropores (e.g., old root channels) can cause irregularities in horizon boundaries, including overthickened tongues of E horizon.

As time passes, weatherable minerals in the E horizon may become deeply pitted by dissolution while the B horizon is enriched with humus and metals. Microbial activity within the B horizon oxidizes organic C, increasing the metal:organic C ratios and releasing metals from the organic complexes to precipitate as poorly crystalline (hydr)oxides. Thus, the B horizon takes on a dark reddish brown color reflecting the humus and Fe (hydr)oxide components. Metal–humus complexes can accumulate in such a way as to produce cemented horizons. One such feature is the placic horizon, which is a thin (2–10 mm), hard, brittle, and wavy zone cemented by Fe/Mn humus. Quite commonly these features accumulate at a hydrological boundary, such as a change in particle size. A more massive cemented horizon is ortstein, which is essentially a spodic horizon cemented by Fe/Al humus. Both of these features require soil that is relatively free of physical disruption in order to form and persist. A study of carbon storage in relation to Fe and Al in grassland soils has suggested that the processes of chelation and transport of Fe and Al can occur in soils that do not develop typical spodic features (Masiello et al., 2004). In these soils, organic carbon stability was found to be strongly correlated with chelated metal ions in the A horizon and noncrystalline Fe and Al in the B horizon, suggesting a carbon storage mechanism similar to that of Spodosols. However, the morphologic expression of these processes is not visible because of the simultaneous translocation and accumulation of clays in B horizon.

30.3.2.7.4 Relation to Environmental Conditions

The process of Fe and Al translocation and accumulation is associated with specific conditions of climate and vegetation. In general, the effective climate is wet, to provide a strong leaching environment, and cool, to produce a low decomposition rate of organics. In these cool, humid environments, coniferous forests and ericaceous shrubs often prevail and are particularly effective in promoting chelation since they have acidic foliage.

As a consequence of this close relationship between climate, vegetation, and the process of Fe/Al translocation and accumulation in soils, Spodosols and similar soils are found in vast areas north of 45° latitude and at lower latitudes where high precipitation and low temperature prevail, such as coastal regions or

TABLE 30.3 Chemical Variables and Their Roles in the Complexation and Redistribution of Fe and Al in Soils

Variable	Major Role and Interaction with Other Variables	Trend, Interaction with Compartment			
		O	E and/or A	B	C
DOC	Major driving variable mobile anion, acidity source, metal complexing agent	Major source	Minor source	Major sink	No trend
pH	Low pH controlled by DOC; major variable	Lowered greatly	Lowered slightly	Rises greatly	Rises slightly
HCO₃	Controlled by pH, P_{CO_2}	Lowered	Insignificant	Insignificant	Rises significantly
Fe	Complexed and mobilized by DOC, causes DOC immobilization in B	Source	Major source	Major sink	Insignificant
Al	Complexed and mobilized by DOC, causes DOC immobilization in B; may be mobilized inorganically at low pH	Minor source	Major source	Major sink	Insignificant or minor
Basic cations	Leached in association with DOC (upper horizons) and with HCO₃⁻ (C horizons)	Major source	Source	Sink	Deep leaching (sink)

Source: After Marrett, D.J. 1988. Acid soil processes in the Okpilak Valley, artic Alaska. Ph.D. Dissertation. University of Washington. Seattle, WA.

high elevations. On the other hand, Spodosols are common in warm humid regions where soils are poorly drained. Organic acids in the groundwater chelate Fe and Al, which are concentrated by the fluctuating water table to precipitate in a Bh horizon. The Bh horizons typically contain chelates but little or no free Fe as (hydr)oxides since it is reduced and leached away. Soils formed in this way have very pronounced albic and spodic horizons. They are common in the coastal plains of the southeastern United States and tropical forests such as the Orinoco basin in Brazil. The observation of close ties between organic matter stability and chelated forms of Fe and Al in moist grassland soils, without spodic properties, suggests that redistribution of Fe and Al in association of organic chelates may occur in a wider range of humid environments than previously recognized (Masiello et al., 2004).

Parent material composition is also very influential in the process of metal–humus translocation and accumulation. The process is promoted by relatively low levels of Ca, Fe, and Al. High levels of these cations prevent mobilization because a high metal:organic C complex forms quickly that is not soluble and readily translocated. Furthermore, high Ca contents promote microbial activity that decomposes soluble organics, so they are not available for chelation and leaching. As a rule, translocation of Fe– and Al–humus complexes is favored by silicic or felsic parent materials but not carbonate or mafic materials. It is further favored by coarse-textured materials (sand to coarse loamy) where surface area is low, water infiltration is rapid, and cation release is slower.

30.3.2.8 Desilication and Concentration of Resistant Oxides

Weathering of primary minerals occurs to different degrees in all soil environments. It releases highly mobile basic cations, moderately mobile $Si(OH)_4$, and relatively immobile Al^{3+} and Fe^{3+} into soil solution. The fate of these ions depends on soil leaching intensity, organic matter composition, the amount of reactive surface area, and pH. In this section, the cumulative effect of leaching on soil volume change, mineral composition, and soil fabric is discussed.

30.3.2.8.1 Leaching and Mineral Composition

In humid environments, well-drained soils lose many of the mobile constituents in a process that lowers pH and changes solution ionic composition from base cations and $Si(OH)_4$ to Al. At neutral pH, Si is more soluble than Al or Fe (Figure 30.10). In the pH range from 5 to 7, much silica can be lost by leaching. In weathered soils of the North Carolina Piedmont, 39%–75% of Si was lost relative to the parent material, but only 5%–13% of Fe and Al was lost (Oh and Richter, 2005).

Below pH 5, Al is also leached but commonly at slower rates than Si because Al can be strongly sorbed by organic matter. Usually there is enough Al in soil minerals that the pH is buffered near 5 and seldom drops to levels (<4) where Fe is dissolved due to acidity alone (Van Breemen et al., 1983). Reducing conditions are required to solubilize Fe as described
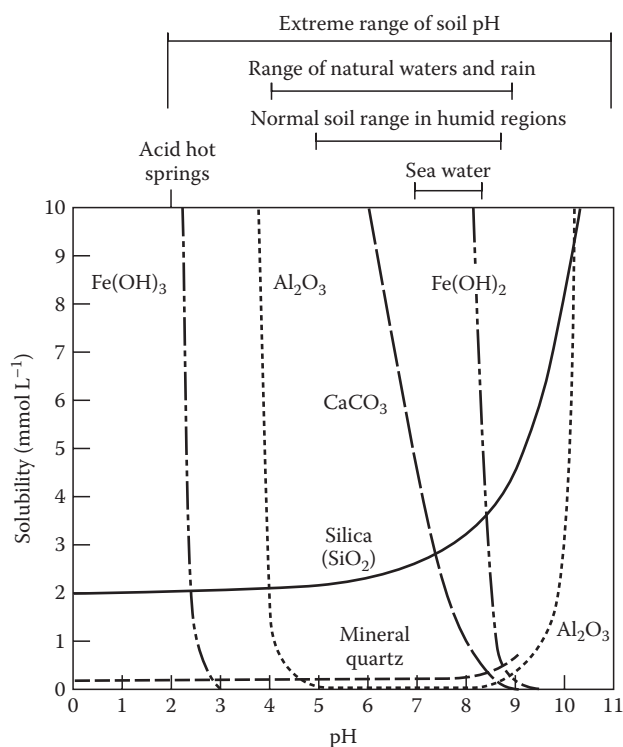
FIGURE 30.10   Mineral solubility as a function of pH.

in Section 30.3.2.9. Thus, leaching changes soil mineral stability fields in favor of minerals composed of Fe and Al and relatively small amounts of Si.

A typical soil mineral assemblage accumulated after intense weathering in a humid environment is shown in Figure 30.11.
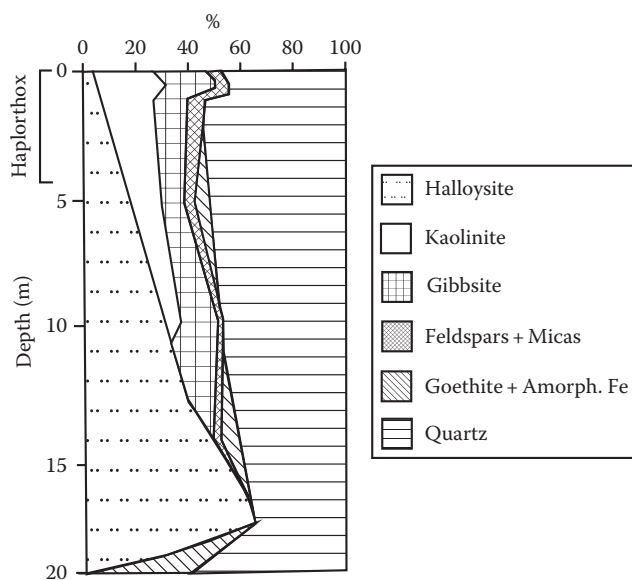
FIGURE 30.11   Mineral distribution in an Oxisol and associated saprolite on granite. (From Eswaran, H., and W.C. Bin. 1978. A study of a deep weathering profile on granite in peninsular Malaysia. I. Physicochemical and micromorphological properties. Soil Sci. Soc. Am. J. 42:144–149. With permission of the Soil Science Society of America.)

The secondary mineral assemblage of highly weathered soils includes kaolin and sometimes hydroxy-interlayered vermiculite (containing Si and Al), gibbsite (containing Al), and hematite and goethite (containing Fe) (Lynn et al., 2002). There is also quartz that is inherited from the parent material or added by dust, and concentrated by the dissolution of other primary minerals. These minerals have low nutrient retention and supply capabilities, which means that nearly all plant nutrients must be derived from breakdown of organic matter or from atmospheric deposition. Because Si has been dramatically depleted and that remaining resides in quartz and kaolinite, these minerals are subject to weathering to a greater extent than in soils that still contain weatherable primary minerals. Even though kaolinite is considered to be a stable end product of weathering, it can decompose (and form) quite rapidly in highly leached soils (Giral-Kacmarcik et al., 1998; Ziegler et al., 2005a). Silicon is conserved by biocycling between rainforest vegetation, where it forms opal phytoliths, and soils, where it is released by weathering of the phytoliths (Lucas et al., 1993; Alexandre et al., 1997; Derry et al., 2005). The preservation of kaolin minerals in the surface horizons of some highly weathered soils may be explained by biocycling of silicon (Kleber et al., 2007).

30.3.2.8.2 Leaching and Collapse

As primary rock minerals are dissolved, they commonly lose volume (Chadwick et al., 1990; Brimhall et al., 1992). This process is countered by any mass addition to soil, which will serve to expand or dilate it. Quantitative studies of leaching losses during soil formation account for these changes through the use of index minerals (or elements), which are extremely resistant to weathering (Brewer, 1964; Smeck and Wilding, 1980; Brimhall and Dietrich, 1987; Brimhall et al., 1992). Dilation is indicated when the quantity of an index element in a soil horizon is less than in the parent material and soil collapse is indicated when the opposite is true. As depicted in Figure 30.12, young soils are often dilated because accumulation of organic matter is more rapid than mineral weathering. Older soils show progressively greater collapse because mineral weathering becomes the dominant control as organic C accumulation is balanced by microbial respiration (Chadwick et al., 1994). Figure 30.12 was constructed by analyzing soils developing in beach sand on a suite of progressively older uplifted marine terraces (Brimhall et al., 1992; Merritts et al., 1992); similar pattern, of early C driven dilation followed by weathering driven collapse occur in soils on lava flows in Hawaii (Vitousek et al., 1997).
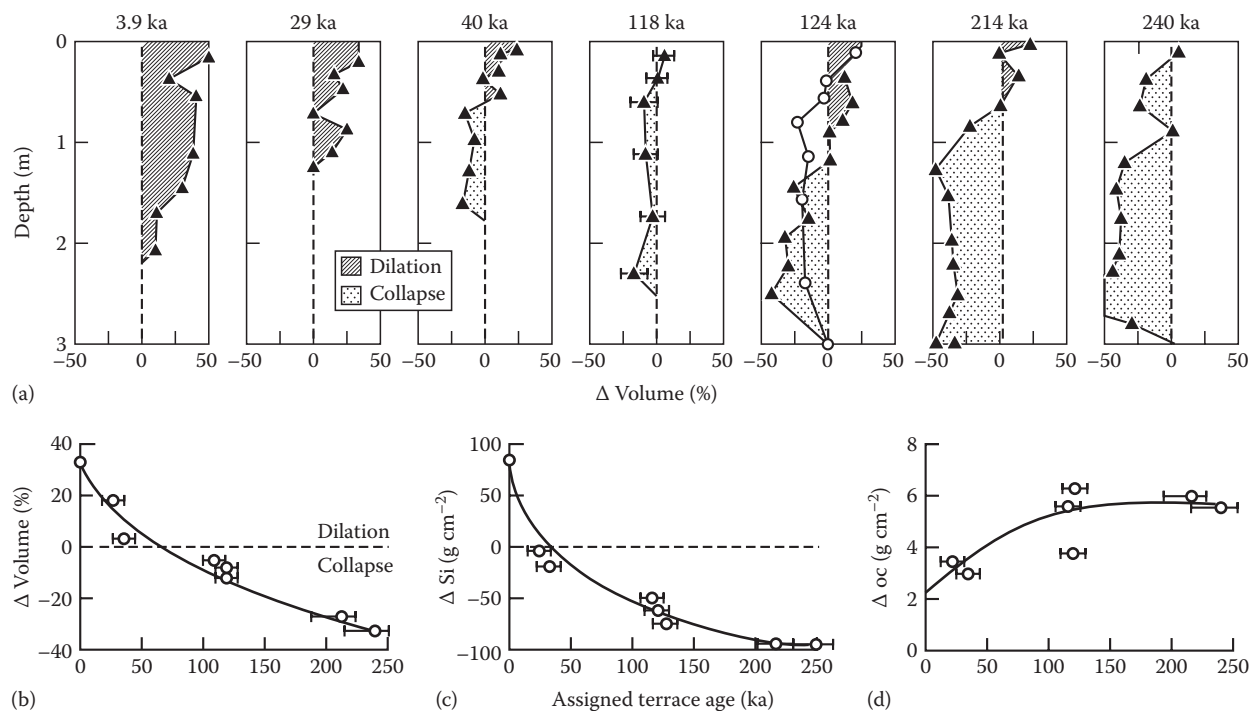
FIGURE 30.12 Parent material and soil values for bulk density and Zr are used to compute volume change that occurs as soils develop from beach sand to 240 ka on uplifted marine terraces. (a) Volume change plotted as a function of depth for each profile sampled; the 124 ka terrace was sampled in two different locations; (b) volume change integrated to the sampling depth (depth-weighted mean) to provide an average value for each profile plotted as a function of terrace age; (c) the quantity of Si leached from each profile as a summation over the soil sampling depth plotted as a function of terrace age. For these soils, base cations are leached more rapidly, and Al is not leached to a significant extent. Given the mass dominance of Si in the arkosic beach sand, Si leaching is the main control on soil collapse; (d) the quantity of organic carbon summed over the sampling depth plotted as a function of terrace age. Organic matter accumulation is the dominant factor driving early dilation and partly offsetting desilication in the older profiles. (Modified from Brimhall, G.H., O.A. Chadwick, C.J. Lewis, W. Compston, I.S. Williams, K.J. Danti, W.E. Dietrich, M.E. Power, D. Hendricks, and J. Bratt. 1992. Deformational mass-transport and invasive processes in soil evolution. Science 255:695–702. Copyright American Association for the Advancement of Science.)

30.3.2.8.3 Accumulation of Atmospherically Derived Constituents

Soils that reside on old stable geomorphic positions accumulate atmospherically transported minerals. For example, dust from Africa augments soils in the Amazon rainforest (Swap et al., 1992; Okin et al., 2004) and dust from Asia can be found in Hawaiian soils (Jackson et al., 1971; Kurtz et al., 2001). In parts of Africa and Australia, deeply weathered bauxite deposits are produced partly by the accumulation of chemically mature Fe and Al compounds that are blown in from other regions (Brimhall et al., 1988). These eolian additions are translocated into the top few meters of a deposit through macropores created by biological activity where they effectively dilate previously collapsed horizons (Brimhall et al., 1992). Below the lower limit of root growth, translocation is no longer accommodated by smaller pores in the saprolite, so leaching has led to collapse. Because these soils have accumulated eolian additions of highly weathered Fe and Al compounds, Si and basic cation losses could be overestimated when comparing soil horizons to underlying material. In contrast, when unweathered primary minerals are added to soils, it is possible that leaching losses will be underestimated. Rate of dust deposition may be similar to weathering rates, depending of mineralogy and climatic conditions (Porder et al., 2007). When weatherable dust is deposited on soils formed from resistant rock types, dust may be the main contributor to the weathering flux. On the other hand, in Hawaii, quartz-rich dust is added to soils formed on mafic bedrock, so that the dust is concentrated in the surface horizons as the bedrock parent material is broken down (Kurtz et al., 2001).

30.3.2.8.4 Relation to Soil Morphology

Horizons in the top 1–2 m of old, highly weathered soils display fine micropeds (<2 mm), which can agglomerate into larger structural units, but they rarely exceed 50 mm in diameter (van Wambeke et al., 1983). The peds are composed of strongly interbonded kaolinite and Al and Fe (hydr)oxides that have low CEC, which defines, in part, the existence of an oxic or kandic diagnostic horizon. In oxic horizons, clay films are rare both because of high aggregate stability and intense bioturbation caused by ants and termites. The soil has high microporosity within peds and relatively large interped pores, which produce excellent permeability.

In Fe-rich soils that are imperfectly drained, localized reduction and oxidation allows Fe to become mobile for short distances. Its redistribution produces localized accumulations of hematite or goethite, which appear as reddish or yellowish mottles. The location of initial deposition can be in small pores where it is thought that precipitation is favored because the chemical potential of water in close association with the solid phase is at a minimum (Tardy and Nahon, 1985), although precipitation can also occur in large pores where high P_{O_2} can lead to oxidation (Bouma, 1983). Initially, the accumulated Fe forms a weakly defined glaebule, or segregated body, that includes primary mineral grains and secondary clays as well as Fe (hydr)oxides. Slowly, the glaebule becomes more clearly defined because the engulfed minerals decompose and are replaced by Fe (hydr)oxide (Nahon, 1991). Mineral decomposition is driven by alternating oxidation and reduction of Fe, which creates acidity for enhanced hydrolysis (Brinkman, 1978). Continued growth of the glaebule produces an abrupt boundary between it and the surrounding soil matrix. In time, glaebules grow into each other producing a reticulate pattern of soil material enriched in Fe (hydr)oxides called plinthite, interspersed with less red, kaolin- and gibbsite-rich matrix. Plinthite in perennially moist soils remains soft, but under wetting and drying conditions, it can solidify into iron stone nodules or continuously cemented ferricrete. Wetting and drying can be the result of natural climate, climatic drying, erosional dissection of a plateau, or excavation of a road (Daniels et al., 1971; Nahon, 1991). There is no certain evidence on how long it takes for plinthite to harden when exposed to wetting and drying conditions.

30.3.2.9 Reduction and Oxidation Leading to Depletions and Concentrations

Environmental factors, particularly climate, topography, and the chemical and physical nature of the substratum create the drainage properties of soils, which, in turn, influence the intensity, duration, and spatial occurrence of anoxic conditions. Soils are susceptible to anoxia because O_2 is consumed by belowground microbial respiration but can only be supplied to soil pores by diffusion from the aboveground atmosphere. Soils with few macropores and many water-filled pores often consume O_2 more rapidly than it can be resupplied by diffusion. Many soils will be anoxic for short periods right after intense wetting events and the interiors of peds may be anoxic, even in otherwise well-aerated soils. When O_2 is depleted, microbially induced reduction reactions result in the dissolution of redox sensitive compounds, with the common result being an increase in their solubility, leading to selective elemental loss from the anoxic area and subsequent precipitation during exposure to higher O_2 levels. These processes leave long-lasting visible imprints on soil morphology.

30.3.2.9.1 Reactive Agents in Redox Processes

The reactive agents in redox processes include organic matter, oxygen, Fe and Mn (hydr)oxides, nitrates, sulfides and sulfates, and microbes (Chapter 14). Microbial activity is the key to reduction in soils. Microbes require a C source, supplied by solid or dissolved organic matter, and electron acceptors. In well-aerated soils, O_2 is the electron acceptor, but as it is used up, nitrates, Mn and Fe (hydr)oxides, and sulfates are used by different microbial populations. Such anaerobic conditions are usually associated with saturated or very wet soils in which there is little free pore space for the influx of O_2 from the atmosphere. Each electron acceptor compound is associated with a characteristic range of redox potentials (Chapter 14). Oxygen levels and redox potentials within a soil may show extreme variation at any given time, even on a scale of millimeters, such as from the exterior to the interior of an aggregate (Zausig et al., 1993). The behavior of nutrients that are not redox sensitive may be influenced through their interactions with redox-sensitive species. In oxidized soils

rich in Fe, considerable amounts of phosphorus can sorb to iron oxides; however, in reduced soils, the Fe sink for P is lessened and more of it is lost by leaching or remains in organic form in thickened O horizons (Miller et al., 2001).

In addition to Eh, pH also controls the form of redox sensitive species in soils. The theoretical relationship between Eh, pH, and the form of redox sensitive species can be described using stability diagrams (Vepraskas and Faulkner, 2001). However, caution is required in using these thermodynamic relationships to described real world soils, in which the soil solution composition and variation in reaction rates complicate the thermodynamic relationships expressed in a stability diagram. The effect of pH on redox reactions in soils can be observed in certain field settings. This is demonstrated by comparison of two soils with low-lying positions, seasonally high water tables, and reducing conditions on the North Carolina piedmont (McDaniel and Buol, 1991). An Ultisol formed in biotite gneiss with a very strong acid subsoil was observed to be gleyed with very little manganese, while an Alfisol formed in hornblende–epidote gneiss with a moderately acid subsoil contained concentrations of secondary Mn.

30.3.2.9.2 Oxidation of a Reduced Soil

As a wet soil drains, macropores are the first to lose water. Often shrinkage occurs during the drying of soils, so that cracks form between peds. Oxygen penetrates through the interpedal pores, root channels, macrofaunal burrows, and other macropores. In response to the higher Eh that develops within these voids, Fe and Mn (hydr)oxides precipitate from the soil solution within the matrix near the pore wall surface. Precipitation of these (hydr)oxides removes the metal from solution, thereby establishing a diffusion gradient, causing Fe^{2+} and Mn^{2+} to migrate from the still reduced soil matrix to the oxidizing zone adjacent to the macropores (Fanning and Fanning, 1989). Redox sensitive elements such as Fe and Mn often accumulate along macropores or the exterior of peds (Figure 30.13). When a fine-textured
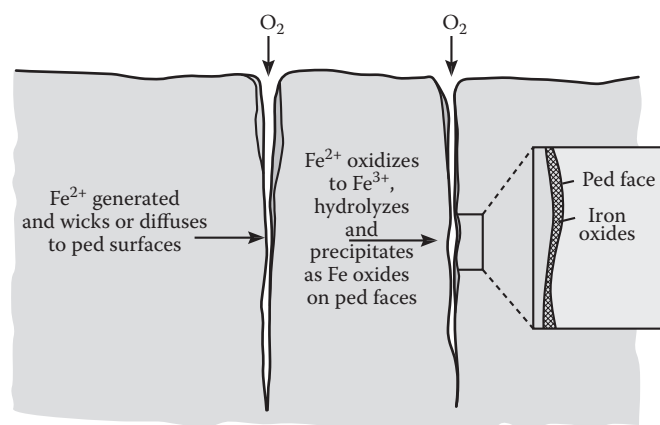
FIGURE 30.13 Macropores (e.g., cracks) facilitate O_2 precipitation of Fe (hydr)oxides on macropore wall surfaces (e.g., ped faces). (Modified from Fanning, D.S., and M.C.B. Fanning. 1989. Soil morphology, genesis, and classification. John Wiley & Sons, New York. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.)

horizon overlies a coarse layer, Fe and Mn (hydr)oxides may be concentrated in the upper part of the coarse deposit, reflecting the oxidation of reduced Fe and Mn as they move into the larger pores of the coarse-textured deposit (D'Amore et al., 2004). Oxidized microsites within a reduced matrix, which may be more common in coarse-textured soils (D'Amore et al., 2004), provide the conditions for formation of hard aggregates of Fe and Mn (hydr)oxides, described as nodules or concretions. These features form by an initial precipitation of Fe or Mn (hydr) oxides within the microsite, which acts as a template for further precipitation. When Fe (hydr)oxides precipitate, the color is red, orange, or yellow, depending on the species of Fe (hydr)oxide formed (Chapter 22). Manganese (hydr)oxides are a strong pigment and if they are present, even at levels of tenths of a percent, a pore wall will be blackened.

30.3.2.9.3 Reduction of an Oxidized Soil

Soils that are generally well aerated and oxidized periodically become so wet that macropores are filled with water and O_2 is excluded. Spatial and temporal variability in reducing conditions occurs within saturated soils, primarily due to factors that affect the rate of microbial activity, including soil temperature and organic matter content. The length of time required for a saturated soil to develop reducing conditions is variable. On floodplain soils of the Mid-Atlantic Piedmont, it took 20 days for reducing conditions to develop when soil temperatures were <4°C, but only 2 days when the soil temperature was >9°C (Vaughan et al., 2009). This relationship between temperature and development of reducing conditions is attributed to the lower activity of soil microorganisms at low temperatures.

The spatial distribution of reducing conditions is influenced by organic matter from roots, which typically persists in old root channels and along ped faces of strongly structured soils. This organic matter provides an abundant source of C for microbes, and Fe (hydr)oxides in the soil adjacent to the macropore serve as electron acceptors and are reduced to yield Fe^{2+} in solution. At the same time, particularly in highly structured, fine-textured soils, ped interiors and soil matrix at some distance from the macropores may contain sites that are not water saturated and retain relatively high Eh conditions. In this case, Fe^{2+} mobilized from soil adjacent to the pores diffuses into the higher Eh environments within the matrix, where it precipitates as Fe (hydr)oxides (Fanning and Fanning, 1989; Vepraskas, 1996). Soil materials near the pore is the site of redox depletion and is generally a gray or white color, whereas redox concentrations of Fe (red, orange, or yellow) occur within the soil matrix (Figure 30.14). Because Mn reduction and oxidation occur at higher Eh values than Fe, Mn (hydr)oxides should also be dissolved near the water-saturated pores and precipitated within the soil matrix but not as far from the pore as the Fe.

30.3.2.9.4 Redox Soil Systems

A classic expression of the effects of varying redox conditions is found in the catena concept, where upslope soils are usually well
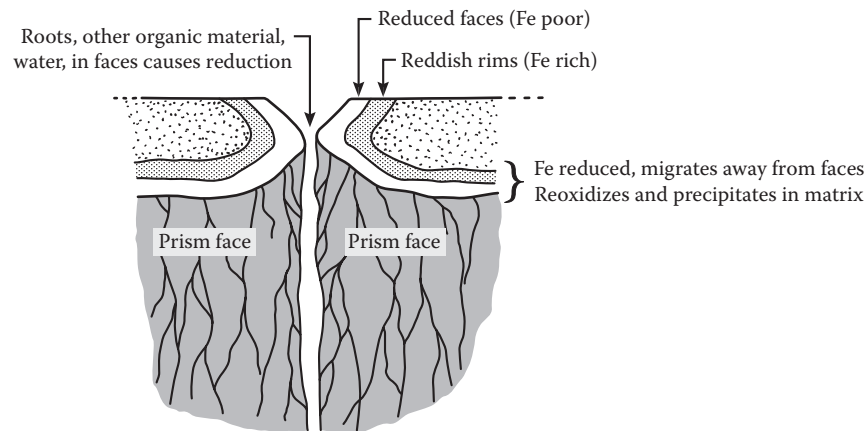
FIGURE 30.14 Occasional, localized reduction occurs in well-drained soils when macropores (e.g., cracks between peds) are temporarily filled with water. Fe (hydr)oxides in rims of peds are dissolved and removed or reprecipitated in more oxidized interior zones of the peds. (After Fanning, D.S., and M.C.B. Fanning. 1989. Soil morphology, genesis, and classification. John Wiley & Sons, New York. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.)

drained and oxidized, while soils at the base of the slope are, at least seasonally, poorly drained and reduced (Fanning and Fanning, 1989). As a result of these topographically induced pedochemical conditions, the upslope soils contain relatively abundant Fe (hydr)oxides and are reddish, whereas soils at the base of the slope contain few Fe (hydr)oxides, are enriched in Mn (hydr) oxides, and generally have low chroma colors (Weitkamp et al., 1996). Redox potentials in the upper slope soils are sufficiently high so that Fe (hydr)oxides are not reduced, although they are occasionally low enough to reduce Mn (hydr)oxides. Mn^{2+} is mobilized and transported to the base of the slope, where Eh is often low enough to cause reduction and loss of Fe (hydr)oxides; thus, Fe and Mn are lost from these soils, but only Mn is replaced by additions from upslope soils.

Soils that contain a slowly permeable layer, such as a dense argillic horizon, fragipan, or permafrost layer, may develop a seasonally perched water table. The resulting epiaquic conditions may not substantially deplete Fe or Mn from the soil as a whole, since there is little downward leaching. The redox sensitive elements are redistributed above and within the restrictive layer, largely by processes described above for the reduction of oxidized soils (Fanning and Fanning, 1989). An E horizon, depleted of Fe and Mn (hydr)oxides by reduction and lateral transport, may develop above the restrictive layer (McDaniel and Falen, 1994).

30.3.2.10 Mixing of Soil Materials

Processes that mix soil materials are sometimes viewed as regressive, with the idea that they oppose horizonation. In fact, while intensively mixed soils often display weak horizons, more moderate degrees of mixing can actually play a key role in forming soil horizons (Johnson et al., 1987). For example, bioturbation helps to form A horizons by mixing organic materials into the mineral soil (e.g., Graham and Wood, 1991; Figure 30.15) and various forms of mixing may obliterate rock structure, thus deepening the regolith and promoting soil formation (e.g., Heimsath et al., 1999).
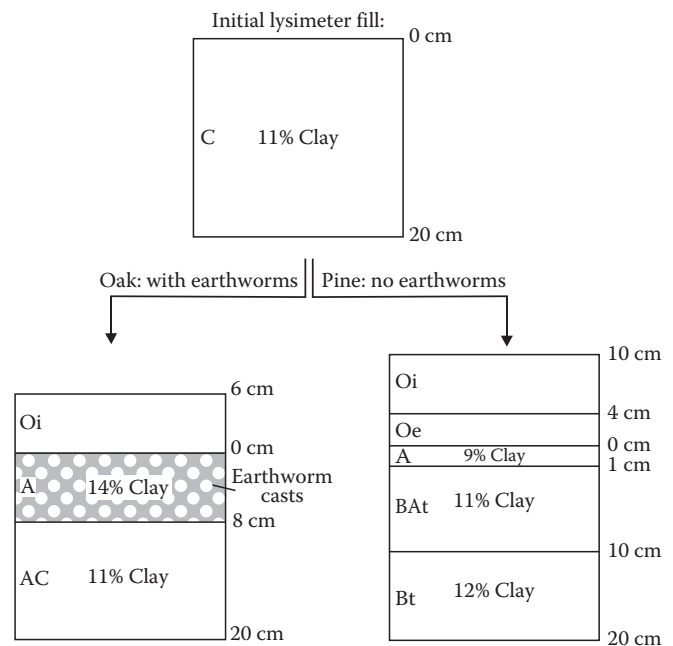
FIGURE 30.15 Soil development after 41 years in lysimeter fill that has been bioturbed by earthworms (oak lysimeter) and in earthworm-free lysimeter fill (pine lysimeter). The bioturbated soils have thinner organic horizons, deeper accumulation of organic C in the mineral soil, and clay accumulation at the surface rather than in the subsoil. (From Graham, R.C., and H.B. Wood. 1991. Morphological development and clay redistribution in lysimeter soils under chaparral and pine. Soil Sci. Soc. Am. J. 55:1638–1646.)

30.3.2.10.1 Agents and Conditions

Many agents of soil mixing can be identified (Hole, 1961); however, the most widely observed include soil biota, seasonally thawed ice, and expansive clays. Mixing by soil biota, referred to as bioturbation, can be caused by both plants and animals. When trees are uprooted, large volumes of soil materials are often moved with the root system. Bioturbation by tree-uprooting is promoted

by conditions that prevent deep-rooting (e.g., shallow bedrock, hardpan, shallow water table), making the trees less wind stable (Schaetzl et al., 1989). Animals that are strong agents of bioturbation include burrowing animals (e.g., gophers, earthworms) and mound-building animals (e.g., ants, termites). Mixing by ice, or cryoturbation, is the dominant soil-forming process in permafrost-affected soils (Bockheim and Tarnocai, 1998). Cryoturbation occurs primarily in the seasonally thawed zone above the permafrost, referred to as the active layer. Mixing caused by the shrinking and swelling of clays, referred to as argilliturbation, occurs when clay-rich soils are subject to seasonal wetting and drying cycles, causing vertical cracks to open and close. Argilliturbation is most pronounced in soils rich in expansive clays, such as smectites. Though these agents work to mix the soil by different mechanisms, they are related in that they apply forces that move bulk soil materials and open up voids through which soil material can move.

30.3.2.10.2 Bioturbation

Soil material can be moved upward, downward, or laterally by bioturbation. Burrowing animals move soil upward by depositing excavated soil material at the surface. The material deposited at the surface is size fractionated to different degrees, depending on the animal. Earthworms sort the fine-earth material by ingesting more silt and clay, relative to sand (Graham and Wood, 1991; Figure 30.15), while gophers sort the coarse fragments, by moving gravels, but not cobbles, stones, or boulders (Johnson et al., 1987). Downward movement of soil material by biotic agents can occur due to backfill of macropores left open by soil organisms or by intentional transport of materials by burrowing animals. For example, some earthworms will drag leaves into their burrows, thus mixing organic horizon material into the mineral soil. Lateral transport of soil material occurs when trees are uprooted; leaving an open pit where the tree was standing and creating a mound next to the pit, where soil material slumps off of the rootwad. Mound-building animals can also be responsible for lateral movement of soil materials. For example, when gophers inhabit shallow or poorly drained soils, they redistribute the soil into mounds that are deep enough to build their burrows (Cox and Scheffer, 1991).

30.3.2.10.3 Cryoturbation

Several mechanisms of cryoturbation have been described, which may act individually or together to cause soil mixing. These mechanisms may act during freezing, thawing, or in a multistep process that involves both. Mixing due to cryostatic pressure and the frost-pull mechanism of stone movement are processes associated with freezing. The active layer freezes in fronts moving both down from the surface and up from the permafrost. The volume expansion associated with freezing causes pressure that mixes the unfrozen material between the freezing fronts (Ping et al., 2008). Frost pull occurs when the soil freezes from the surface and the primary direction of expansion is upward. Coarse fragments carried with the upward-expanding soil undergo a net upward displacement if return to their original position upon thawing is prevented by fine material infilling beneath the coarse fragment (Washburn, 1980).

Convection cell mixing is a process that is associated with the thawing of frozen soil above the permafrost. Mixing by convection cells occurs over bowl-shaped features of the permafrost table, which cause inward slumping of the thawed soil and upward displacement at the center of the bowl (Mackay, 1980). Frost wedging is a mechanism of cryoturbation that occurs due to the freeze–thaw cycle. When the active layer is frozen, thermal contraction at temperatures below −10°C causes cracks to form, and then as it thaws, melt water fills in the cracks and refreezes. Expansion of the water as it refreezes (between 4°C and 0°C) causes pressure on the surrounding soil and distortion of the land surface (Washburn, 1980).

30.3.2.10.4 Argilliturbation

Mixing by argilliturbation occurs by drying-induced shrinkage of the soil matrix and formation of desiccation cracks, followed by partial infilling of cracks from overlying horizons. The infilling material may be knocked in by biota, washed in by rain, or otherwise carried by gravity. Soils in which shrinkage increases with depth are most susceptible to these transport mechanisms because desiccation cracks become wider with depth, causing the wall to overhang the void (Hallsworth et al., 1955). When the dry, cracked soil is exposed to heavy rains or flooding, the zone beneath the cracks may wet up and expand first, causing subsoil material to be thrust toward the surface. Coarse fragments may also be moved upward, due to the swelling of clays and expansion of the soil volume upward, in the direction of least resistance. If the coarse fragments are smaller than the desiccation cracks, they may be transported back down during subsequent drying cycles. However, if the fragments are larger than the desiccation cracks they are transported upward only and accumulate at the soil surface (Johnson and Hester, 1972).

30.3.2.10.5 Relation to Soil Morphology

Mixing counteracts redistribution processes that form subsurface horizons, transports organic matter to greater depths, and causes irregularities in horizon topography. For example, in a biosequence composed of soils formed under oak, which contained earthworms, and adjacent earthworm-free soils formed under pine, only the pine soils had an argillic horizon (Graham and Wood, 1991; Figure 30.15). Mixing by the earthworms prevented illuvial clay accumulation under the oak. Mixing can also create thicker organic-rich horizons due to the infilling of animal burrows or desiccation cracks with A horizon material. In cryoturbated soils, organic matter is often observed to accumulate on top of or within the permafrost, as a result of downward transport of organic-rich materials from the surface horizons (Bockheim and Tarnocai, 1998; Bockheim, 2007; Ping et al., 2008). Examples of horizon boundaries influenced by mixing processes included swirl-like horizon patterns caused by cryostatic pressure in Gelisols (Bockheim and Tarnocai, 1998) and irregular horizon boundaries caused by up-thrust C horizon material in Vertisols (Hallsworth et al., 1955).

Processes that mix the soil may cause coarse fragments to concentrate at a certain depth in the soil. Coarse fragments too

large for transport by burrowing or mound-building animals occur in a layer beneath a mantle of finer-textured material that the animals have brought to the surface (Johnson et al., 1987). Thus, bioturbation can result in a subsurface concentration of coarse fragments, which occurs at the maximum depth of faunal activity. Argilliturbation, on the other hand, concentrates coarse fragments at the surface, in the form of a stone pavement. Stone pavement formation by argilliturbation is exemplified by Vertisols on the Channel Islands, California, which are formed in sediments deposited above bedrock, yet are observed to have an enrichment of bedrock fragments at the soil surface (Johnson and Hester, 1972). Stone pavements may also form through freeze–thaw processes in glacial landscapes (Simón et al., 2000).

Mixing often occurs in localized patches, creating heterogeneity and microtopography in soil landscapes. Microrelief created by bioturbation includes mounds built by soil fauna and pit and mound topography caused by tree throw. Pit and mound topography in Spodosols has been related to variations in horizon thickness due to more intensive leaching of the pits (Schaetzl, 1990). Microhighs and lows, referred to as gilgai relief, are a common feature of argilliturbated soils. The microhighs are often associated with soil material from below the zone of mixing that has been up-thrust toward the surface (Hallsworth et al., 1955). In cryoturbated soils, microrelief and/or sorting of coarse and fine soil materials occurs in repeated shapes, referred to as patterned ground. Examples of patterned ground included polygons, ranging from a few centimeters to a few meters in diameter, formed by frost wedging (Ping et al., 2008) and hummocks, 1–2 m in diameter, formed by slumping of the active layer as it thaws over a curved permafrost table (Mackay, 1980). In between these units, upward injection of C horizon material may occur, resulting in a sorted form of patterned ground (Vandenberghe, 1992; Ugolini et al., 2006).

30.4 From Property to Process

For convenience, specific soil-forming processes have been identified and discussed separately. This leaves the impression that one can easily interpret soil-forming processes from soil properties. To a degree this inversion can be done quite well, but care is required. Many processes occur in the same pedon simultaneously, which imprints a less than clear suite of properties. Furthermore, secondary soil-forming processes may be initiated later in soil development in response to properties developed by the primary processes. Numerous examples of such intrinsic pedogenic thresholds illustrate the complex interactions that can occur between pedogenic processes (Chadwick and Chorover, 2001; Ewing et al., 2006). Development of cemented subsurface horizons can have a strong influence on the pedogenic processes that occur in the surface horizons. For example, by decreasing rooting depth, cemented horizons in forest soils become more susceptible to bioturbation by tree-uprooting (Schaetzl et al., 1989). Altered drainage due to a cemented horizon may also dramatically change the pedogenic processes, as demonstrated in a chronosequence study in southeast Alaska, in which initial

processes of Fe and Al translocation form Spodosols with a cemented placic horizon (Ugolini and Mann, 1979). The cemented horizon impedes drainage leading to anaerobic conditions and accumulation of organic material, resulting in the formation of Histosols. Development of surface horizons that impede infiltration, such as desert pavement and vesicular horizons, may also change the pedogenic processes. These horizons cause decreased leaching, leading to accumulation of salts in the subsoil (Young et al., 2004), and increased runoff, which may eventually lead to erosion of the restrictive surface (McDonald et al., 1995). The restriction of vegetation due to stone pavement formation may also reduce the accumulation of organic matter (Simón et al., 2000). The accumulation of clay in an argillic horizon may lead to the secondary process of argilliturbation and formation of a Vertisol, especially if erosion exposes the clay-rich subsoil to surface wetting/drying cycles (Graham and Southard, 1983).

Pedogenic processes may also change over the course of soil development because of changes in the soil-forming factors over time. Soils are often polygenetic because their properties developed under different climatic conditions. Past climate change can superimpose different processes on a pedon at different times during soil formation. Soils formed in glacial deposits are subject to extreme variation in climate between glacial and interglacial periods. Glacial deposits in the Rocky Mountains have undergone phases of carbonate accumulation during interglacial periods, interrupted by cryoturbation during glacial advances (Hall, 1999). Preglacial soils in the Sierra Nevada Mountains of southern Spain include unique features that reflect the transition of pedogenic processes related to changes in climate (Simón et al., 2000). In these soils, clay films in the argillic horizon (formed during the warmer, moisture preglacial climate) have been mixed by cryoturbation during glacial periods, resulting in abundant clay fragments in the matrix. It is important to be alert to the possibility that the present set of soil properties is a composite derived from a series of temporally varying processes. Interpretation of soil processes requires as full an Earth history context as possible. Smart sampling, keen observation, and an open mind are prerequisites to successful understanding.

References

Aber, J.D., and J.M. Melillo. 1991. Terrestrial ecosystems. Saunders College Publishing, Tokyo, Japan.

Alexandre, A., J.D. Meunier, F. Colin, and J.M. Koud. 1997. Plant impact on the biogeochemical cycle of silicon and related weathering processes. Geochim. Cosmochim. Acta 61:677–682.

Amundson, R., and H. Jenny. 1991. The place of humans in the state factor theory of ecosystems and their soils. Soil Sci. 151:99–109.

Amundson, R., and H. Jenny. 1997. Thinking of biology: On a state factor model of ecosystems. Bioscience 47:536–543.

Anderson, J.U., D. Silberman, and D. Rai. 1975. Humus accumulation in a forested haploboroll in south-central New Mexico. Soil Sci. Soc. Am. J. 39:905–908.

Atlas, R.M., and R. Bartha. 1981. Microbial ecology: Fundamentals and applications. Addison-Wesley Publishing Co., Redding, MA.

Bartelli, L.J., and R.T. Odell. 1960. Field studies of a clay-enriched horizon in the lowest part of the solum of some brunizem and gray-brown podzolic soils in Illinois. Soil Sci. Soc. Am. Proc. 24:388–390.

Berner, E.K., and R.A. Berner. 1996. Global environment: Water, air, and geochemical cycles. Prentice-Hall, Upper Saddle River, NJ.

Birkeland, P.W. 1999. Soils and geomorphology. 3rd edn. Oxford University Press, New York.

Bockheim, J.G. 1990. Soil development rates in the transantarctic mountains. Geoderma 47:59–77.

Bockheim, J.G. 1997. Properties and classification of cold desert soils from Antarctica. Soil Sci. Soc. Am. J. 61: 224–231.

Bockheim, J.G. 2007. Importance of cryoturbation in redistributing organic carbon in permafrost-affected soils. Soil Sci. Soc. Am. J. 71:1335–1342.

Bockheim, J.G., and C. Tarnocai. 1998. Recognition of cryoturbation for classifying permafrost-affected soils. Geoderma 81:281–293.

Bouma, J. 1983. Hydrology and soil genesis of sils with aquic moisture regimes, p. 253–281. *In* L.P. Wilding et al. (eds.) Pedogenesis and soil taxonomy. I. Concepts and interactions. John Wiley & Sons, Inc., New York.

Bouza, P., H.F. Delvalle, and P.A. Imbellone. 1993. Micromorphological, physical, and chemical characteristics of soil crust types of the Central Patagonia Region, Argentina. Arid Soil Res. Rehabil. 7:355–368.

Brewer, R. 1964. Fabric and mineral analysis of soil. John Wiley & Sons, Inc., New York.

Brimhall, G.H., O.A. Chadwick, C.J. Lewis, W. Compston, I.S. Williams, K.J. Danti, W.E. Dietrich, M.E. Power, D. Hendricks, and J. Bratt. 1992. Deformational mass-transport and invasive processes in soil evolution. Science 255:695–702.

Brimhall, G.H., and W.E. Dietrich. 1987. Constitutive mass balance relations between chemical-composition, volume, density, porosity, and strain in metasomatic hydrochemical systems—Results on weathering and pedogenesis. Geochim. Cosmochim. Acta 51:567–587.

Brimhall, G.H., C.J. Lewis, J.J. Ague, W.E. Dietrich, J. Hampel, T. Teague, and P. Rix. 1988. Metal enrichment in bauxites by deposition of chemically mature aeolian dust. Nature 333:819–824.

Brinkman, R.H. 1978. Ferrolysis. Elsevier, New York.

Brock, A.L., and B.J. Buck. 2005. A new formation process for calcic pendants from Pahranagat Valley, Nevada, USA, and implication for dating Quaternary landforms. Quat. Res. 63:359–367.

Buck, B.J., and J.G. Van Hoesen. 2002. Snowball morphology and SEM analysis of pedogenic gypsum, southern New Mexico, USA. J. Arid Environ. 51:469–487.

Capo, R.C., and O.A. Chadwick. 1999. Sources of strontium and calcium in desert soil and calcrete. Earth Planet. Sci. Lett. 170:61–72.

Capo, R.C., C.E. Whipkey, J. Blanchere, and O.A. Chadwick. 2000. Pedogenic origin of dolomite in a basaltic weathering profile, Kohala Peninsula, Hawaii. Geology 28:271–274.

Chadwick, O.A., G.H. Brimhall, and D.M. Hendricks. 1990. From a black to a gray box—A mass balance interpretation of pedogenesis. Geomorphology 3:369–390.

Chadwick, O.A., and J. Chorover. 2001. The chemistry of pedogenic thresholds. Geoderma 100:321–353.

Chadwick, O.A., L.A. Derry, P.M. Vitousek, B.M. Huebert, and L.O. Hedin. 1999. Changing sources of nutrients during four million years of ecosystem development. Nature 397:491–497.

Chadwick, O.A., D.M. Hendricks, and W.D. Nettleton. 1987. Silica in duric soils. 1. A depositional model. Soil Sci. Soc. Am. J. 51:975–982.

Chadwick, O.A., E.F. Kelly, D.M. Merritts, and R.G. Amundson. 1994. Carbon-dioxide consumption during soil development. Biogeochemistry 24:115–127.

Cox, G.W., and V.B. Scheffer. 1991. Pocket gophers and mima terrain in North America. Nat. Areas J. 11:193–198.

D'Amore, D.V., S.R. Stewart, and J.H. Huddleston. 2004. Saturation, reduction, and the formation of iron–manganese concretions in the Jackson-Frazier Wetland, Oregon. Soil Sci. Soc. Am. J. 68:1012–1022.

Dalrymple, J.B., and S.P. Theocharopoulos. 1984. Intrapedal cutans—Experimental production of depositional (illuviation) channel argillans. Geoderma 33:237–243.

Daniels, R.B., E.E. Gamble, and J.C. Cady. 1971. The relation between geomorphology and soil morphology and genesis. Adv. Agron. 23:51–88.

Deconinck, F. 1980. Major mechanisms in formation of spodic horizons. Geoderma 24:101–128.

DeNovio, N.M., J.E. Saiers, and J.N. Ryan. 2004. Colloid movement in unsaturated porous media: Recent advances and future directions. Vadose Zone J. 3:338–351.

Derry, L.A., and O.A. Chadwick. 2007. Contributions from earth's atmosphere to soil. Elements 3:333–338.

Derry, L.A., A.C. Kurtz, K. Ziegler, and O.A. Chadwick. 2005. Biological control of terrestrial silica cycling and export fluxes to watersheds. Nature 433:728–731.

Drees, L.R., L.P. Wilding, N.E. Smeck, and A.L. Senkayi. 1989. Silica in soils: Quartz and disordered silica polymorphs, p. 913–974. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Duncan, M.M., and D.P. Franzmeier. 1999. Role of free silicon, aluminum, and iron in fragipan formation. Soil Sci. Soc. Am. J. 63:923–929.

Durgin, P.B., and J.G. Chaney. 1984. Dispersion of kaolinite by dissolved organic-matter from douglas-fir roots. Can. J. Soil Sci. 64:445–455.

Durgin, P.B., and P.J. Vogelsang. 1984. Dispersion of kaolinite by water extracts of douglas-fir ash. Can. J. Soil Sci. 64:439–443.

Eghbal, M.K., and R.J. Southard. 1993. Micromorphological evidence of polygenesis of 3 aridisols, western Mojave Desert, California. Soil Sci. Soc. Am. J. 57:1041–1050.

Eswaran, H., and W.C. Bin. 1978. A study of a deep weathering profile on granite in peninsular Malaysia. I. Physicochemical and micromorphological properties. Soil Sci. Soc. Am. J. 42:144–149.

Ewing, S.A., B. Sutter, R. Amundson, J. Owen, K. Nishiizumi, W. Sharp, S.S. Cliff, K. Perry, W.E. Dietrich, and C.P. McKay. 2006. A threshold in soil formation at earth's arid-hyperarid transition. Geochim. Cosmochim. Acta 70:5293–5322.

Fanning, D.S., and M.C.B. Fanning. 1989. Soil morphology, genesis, and classification. John Wiley & Sons, New York.

Fireman, M., and H.E. Hayward. 1952. Indicator significance of some shrubs in the Escalante Desert. Utah. Bot. Gaz. (Chicago) 114:143–155.

Franzmeier, D.P., L.D. Norton, and G.C. Steinhardt. 1989. Fragipan formation in loess of the Midwestern United States, p. 69–98. *In* N.E. Smeck and E.J. Ciolkosz (eds.) Fragipans: Their occurrence, classification, and genesis. SSSA Special Publication No. 24. SSSA, Madison, WI.

Gile, L.H., F.F. Peterson, and R.B. Grossman. 1966. Morphological and genetic sequences of carbonate accumulation in desert soils. Soil Sci. 101:347–360.

Giral-Kacmarcik, S., S.M. Savin, D.B. Nahon, J.P. Girard, Y. Lucas, and L.J. Abel. 1998. Oxygen isotope geochemistry of kaolinite in laterite-forming processes, Manaus, Amazonas, Brazil. Geochim. Cosmochim. Acta 62:1865–1879.

Gollany, H.T., R.R. Allmaras, S.M. Copeland, S.L. Albrecht, and C.L. Douglas. 2005. Tillage and nitrogen fertilizer influence on carbon and soluble silica relations in a Pacific Northwest Mollisol. Soil Sci. Soc. Am. J. 69:1102–1109.

Graham, R.C., and S.W. Buol. 1990. Soil-geomorphic relations on the Blue Ridge Front. 2. Soil characteristics and pedogenesis. Soil Sci. Soc. Am. J. 54:1367–1377.

Graham, R.C., W.R. Guertal, and K.R. Tice. 1994. The pedologic nature of weathered rock, p. 21–40. *In* D.L. Cremeens et al. (eds.) Whole regolith pedology. SSSA Special Publication No. 34. SSSA, Madison, WI.

Graham, R.C., D.R. Hirmas, Y.A. Wood, and C. Amrhein. 2008. Large near-surface nitrate pools in soils capped by desert pavement in the Mojave Desert, California. Geology 36:259–262.

Graham, R.C., and A.R. Southard. 1983. Genesis of a Vertisol and an associated Mollisol in northern Utah. Soil Sci. Soc. Am. J. 47:552–559.

Graham, R.C., and H.B. Wood. 1991. Morphological development and clay redistribution in lysimeter soils under chaparral and pine. Soil Sci. Soc. Am. J. 55:1638–1646.

Hall, R.D. 1999. Effects of climate change on soils in glacial deposits, Wind River Basin, Wyoming. Quat. Res. 51:248–261.

Hallsworth, E.G., G.K. Robertson, and F.R. Gibbons. 1955. Studies in pedogenesis in New South Wales. VII. The 'gilgai' soils. J. Soil Sci. 6:1–31.

Heil, D., and G. Sposito. 1993a. Organic-matter role in illitic soil colloids flocculation. 1. Counter ions and pH. Soil Sci. Soc. Am. J. 57:1241–1246.

Heil, D., and G. Sposito. 1993b. Organic-matter role in illitic soil colloids flocculation. 2. Surface-charge. Soil Sci. Soc. Am. J. 57:1246–1253.

Heimsath, A.M., W.E. Dietrich, K. Nishiizumi, and R.C. Finkel. 1999. Cosmogenic nuclides, topography, and the spatial variation of soil depth. Geomorphology 27:151–172.

Hirmas, D.R. and R.C. Graham. 2011. Pedogenesis and soil-geomorphic relationships in an arid mountain range, Mojave Desert, California, Soil Sci. Soc. Am. J. 75:192–206.

Hole, F.D. 1961. A classification of perturbations and some other processes and factors of soil formation in relation to isotropism and anisotropism. Soil Sci. 91:375–377.

Hsu, P.H. 1989. Aluminum oxides and oxyhydroxides, p. 331–378. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Hudson, B.D. 1977. Cohesive water films as a factor in clay translocation. Soil Surv. Horiz. 18:9–15.

Jackson, M.L., T.W.M. Levelt, J.K. Syers, R.W. Rex, R.N. Clayton, G.D. Sherman, and G. Uehara. 1971. Geomorphological relationships of tropospherically derived quartz in soils of Hawaiian-Islands. Soil Sci. Soc. Am. Proc. 35:515–525.

Jenny, H. 1941. Factors of soil formation. McGraw-Hill Book Co., New York.

Jenny, H. 1980. The soil resource, origin and behavior. Springer-Verlag, New York.

Jenny, H., and G.D. Smith. 1935. Colloid chemical aspects of clay pan formation in soil profiles. Soil Sci. 39:377–389.

Johnson, D.L., and N.C. Hester. 1972. Origin of stone pavements on Pleistocene marine terraces in California. Proc. Assoc. Am. Geogr. 4:50–53.

Johnson, D.L., D. Watsonstegner, D.N. Johnson, and R.J. Schaetzl. 1987. Proisotropic and proanisotropic processes of pedoturbation. Soil Sci. 143:278–292.

Kaplan, D.I., P.M. Bertsch, and D.C. Adriano. 1997. Mineralogical and physicochemical differences between mobile and non-mobile colloidal phases in reconstructed pedons. Soil Sci. Soc. Am. J. 61:641–649.

Kaplan, D.I., P.M. Bertsch, D.C. Adriano, and A.J. Miller. 1993. Soil-borne mobile colloids as influenced by water flow and organic carbon. Environ. Sci. Technol. 27:1193–1200.

Kendrick, K.J., and R.C. Graham. 2004. Pedogenic silica accumulation in chronosequence soils, southern California. Soil Sci. Soc. Am. J. 68:1295–1303.

Kleber, M., L. Schwendenmann, E. Veldkamp, J. Rossner, and R. Jahn. 2007. Halloysite versus gibbsite: Silicon cycling as a pedogenetic process in two lowland neotropical rain forest soils of La Selva, Costa Rica. Geoderma 138:1–11.

Kretzschmar, R., W.P. Robarge, and A. Amoozegar. 1995. Influence of natural organic-matter on colloid transport through saprolite. Water Resour. Res. 31:435–445.

Kretzschmar, R., W.P. Robarge, and S.B. Weed. 1993. Flocculation of kaolinitic soil clays—Effects of humic substances and iron-oxides. Soil Sci. Soc. Am. J. 57:1277–1283.

Kurtz, A.C., L.A. Derry, and O.A. Chadwick. 2001. Accretion of Asian dust to Hawaiian soils. Geochim. Cosmochim. Acta 65:1971–1983.

Lee, K.E., and R.C. Foster. 1991. Soil fauna and soil structure. Aust. J. Soil Res. 29:745–775.

Lucas, Y., F.J. Luizao, A. Chauvel, J. Rouiller, and D. Nahon. 1993. The relation between biological-activity of the rain-forest and mineral-composition of soils. Science 260:521–523.

Lundström, U.S., N. van Breemen, D.C. Bain, P.A.W. van Hees, R. Giesler, J.P. Gustafsson, H. Ilvesniemi et al. 2000. Advances in understanding the podzolization process resulting from a multidisciplinary study of three coniferous forest soils in the Nordic Countries. Geoderma 94:335–353.

Lynn, W.C., R.J. Ahrens, and A.L. Smith. 2002. Soil minerals, their geographic distribution, and soil taxonomy, p. 691–709. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA, Madison, WI.

Mackay, J.R. 1980. The origin of hummocks, western arctic coast, Canada. Can. J. Earth Sci. 17:996–1006.

MacKenzie, M.D., E.J.B. McIntire, S.A. Quideau, and R.C. Graham. 2008. Charcoal distribution affects carbon and nitrogen contents in forest soils of California. Soil Sci. Soc. Am. J. 72:1774–1785.

Marrett, D.J. 1988. Acid soil processes in the Okpilak Valley, artic Alaska. Ph.D. Dissertation. University of Washington. Seattle, WA.

Marsan, F.A., and J. Torrent. 1989. Fragipan bonding by silica and iron-oxides in a soil from northwestern Italy. Soil Sci. Soc. Am. J. 53:1140–1145.

Masiello, C.A., O.A. Chadwick, J. Southon, M.S. Torn, and J.W. Harden. 2004. Weathering controls on mechanisms of carbon storage in grassland soils. Global Biogeochem. Cycles 18:GB4023. doi:10.1029/2004gb002219

McDaniel, P.A., and S.W. Buol. 1991. Manganese distributions in acid soils of the North Carolina Piedmont. Soil Sci. Soc. Am. J. 55:152–158.

McDaniel, P.A., and A.L. Falen. 1994. Temporal and spatial patterns of episaturation in a Fragixeralf landscape. Soil Sci. Soc. Am. J. 58:1451–1457.

McDonald, E.V., L.D. McFadden, and S.G. Wells. 1995. The relative influences of climate change, desert dust, and lithologic control on soil-geomorphic processes on alluvial fans, Mojave Desert, California: Summary of results, p. 35–42. *In* R.E. Reynolds and J. Reynolds (eds.) Ancient surfaces of the East Mojave Desert: A volume and field trip guide prepared in conjunction with the 1995 Desert Research Symposium. San Bernardino County Museum Association, Redlands, CA.

Melillo, J.M., J.D. Aber, A.E. Linkins, A. Ricca, B. Fry, and K.J. Nadelhoffer. 1989. Carbon and nitrogen dynamics along the decay continuum—Plant litter to soil organic-matter. Plant Soil 115:189–198.

Merritts, D.J., O.A. Chadwick, D.M. Hendricks, G.H. Brimhall, and C.J. Lewis. 1992. The mass balance of soil evolution on late Quaternary marine terraces, northern California. Geol. Soc. Am. Bull. 104:1456–1470.

Mikutta, R., M. Kleber, M.S. Torn, and R. Jahn. 2006. Stabilization of soil organic matter: Association with minerals or chemical recalcitrance? Biogeochemistry 77:25–56.

Miller, D.E. 1971. Formation of vesicular structure in soil. Soil Sci. Soc. Am. Proc. 35:635–637.

Miller, A.J., E.A.G. Schuur, and O.A. Chadwick. 2001. Redox control of phosphorus pools in Hawaiian montane forest soils. Geoderma 102:219–237.

Monger, H.C., and L.A. Daugherty. 1991. Pressure solution-possible mechanism for silicate grain dissolution in a petrocalcic horizon. Soil Sci. Soc. Am. J. 55:1625–1629.

Monger, H.C., L.A. Daugherty, W.C. Lindemann, and C.M. Liddell. 1991. Microbial precipitation of pedogenic calcite. Geology 19:997–1000.

Moody, L.E., and R.C. Graham. 1997. Silica-cemented terrace edges, central California coast. Soil Sci. Soc. Am. J. 61:1723–1729.

Munk, L.P., and R.J. Southard. 1993. Pedogenic implications of opaline pendants in some California late-Pleistocene Palexeralfs. Soil Sci. Soc. Am. J. 57:149–154.

Nahon, D.B. 1991. Introduction to the petrology of soils and chemical weathering. John Wiley & Sons, New York.

Naiman, Z., J. Quade, and P.J. Patchett. 2000. Isotopic evidence for eolian recycling of pedogenic carbonate and variations in carbonate dust sources throughout the southwest United States. Geochim. Cosmochim. Acta 64:3099–3109.

Nettleton, W.D., K.W. Flach, and B.R. Brasher. 1969. Argillic horizons without clay skins. Soil Sci. Soc. Am. Proc. 33:121–125.

Nierop, K.G.J., B. Jansen, and J.A. Verstraten. 2002. Dissolved organic matter, aluminium and iron interactions: Precipitation induced by metal/carbon ratio, pH and competition. Sci. Total Environ. 300:201–211.

Oades, J.M. 1989. An introduction to organic matter in mineral soils, p. 89–159. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Oades, J.M. 1993. The role of biology in the formation, stabilization and degradation of soil structure. Geoderma 56:377–400.

Oh, N.H., and D.D. Richter. 2005. Elemental translocation and loss from three highly weathered soil-bedrock profiles in the southeastern United States. Geoderma 126:5–25.

Okin, G.S., N. Mahowald, O.A. Chadwick, and P. Artaxo. 2004. The impact of desert dust on the biogeochemistry of phosphorus in terrestrial ecosystems. Global Biogeochem. Cycles 18:GB2005. doi:10.1029/2003GB002145

Paton, T.R., G.S. Humphreys, and P.B. Mitchell. 1995. Soils: A new global view. Yale University Press, New Haven, CT.

Peterson, F.F. 1980. Holocene desert soil formation under sodium-salt influence in a playa margin environment. Quat. Res. 13:172–186.

Pett-Ridge, J.C. 2009. Contributions of dust to phosphorus cycling in tropical forests of the Luquillo Mountains, Puerto Rico. Biogeochemistry 94:63–80.

Ping, C.L., G.J. Michaelson, J.M. Kimble, V.E. Romanovsky, Y.L. Shur, D.K. Swanson, and D.A. Walker. 2008. Cryogenesis and soil formation along a bioclimate gradient in Arctic North America. J. Geophys. Res. 113:G03S12. doi:10.1029/2008JG000744

Porder, S., G.E. Hilley, and O.A. Chadwick. 2007. Chemical weathering, mass loss, and dust inputs across a climate by time matrix in the Hawaiian Islands. Earth Planet. Sci. Lett. 258:414–427.

Quade, J., J.A. Rech, C. Latorre, J.L. Betancourt, E. Gleeson, and M.T.K. Kalin. 2007. Soils at the hyperarid margin: The isotopic composition of soil carbonate from the Atacama Desert, Northern Chile. Geochim. Cosmochim. Acta 71:3772–3795.

Quideau, S.A., R.C. Graham, O.A. Chadwick, and H.B. Wood. 1998. Organic carbon sequestration under chaparral and pine after four decades of soil development. Geoderma 83:227–242.

Rabenhorst, M.C., L.P. Wilding, and C.L. Girdner. 1984. Airborne dusts in the Edwards Plateau region of Texas. Soil Sci. Soc. Am. J. 48:621–627.

Rasmussen, C., R.J. Southard, and W.R. Horwath. 2006. Mineral control of organic carbon mineralization in a range of temperate conifer forest soils. Global Change Biol. 12:834–847.

Rawling, J.E., 3rd. 2000. A review of lamellae. Geomorphology 35:1–9.

Rech, J.A., J. Quade, and W.S. Hart. 2003. Isotopic evidence for the source of Ca and S in soil gypsum, anhydrite and calcite in the Atacama Desert, Chile. Geochim. Cosmochim. Acta 67:575–586.

Reheis, M.C., and R. Kihl. 1995. Dust deposition in southern Nevada and California, 1984–1989: Relations to climate, source area, and source lithology. J. Geophys. Res. 100:8893–8918.

Reid, D.A., R.C. Graham, R.J. Southard, and C. Amrhein. 1993. Slickspot soil genesis in the Carrizo Plain, California. Soil Sci. Soc. Am. J. 57:162–168.

Reynolds, R.L., J.C. Yount, M. Reheis, H. Goldstein, P. Chavez, R. Fulton, J. Whitney, C. Fuller, and R.M. Forester. 2007. Dust emission from wet and dry playas in the Mojave desert, USA. Earth Surf. Process. Landf. 32:1811–1827.

Rossi, A.M. 2009. Soil development and clast weathering on a moraine chronosequence, eastern Sierra Nevada, California. M.S. Thesis. University of California. Riverside, CA.

Schaetzl, R.J. 1990. Effects of treethrow microtopography on the characteristics and genesis of Spodosols, Michigan, USA. Catena 17:111–126.

Schaetzl, R.J. 1996. Spodosol-Alfisol intergrades: Bisequal soils in NE Michigan, USA. Geoderma 74:23–47.

Schaetzl, R.J., and S. Anderson. 2005. Soils: Genesis and geomorphology. Cambridge University Press, Cambridge, U.K.

Schaetzl, R.J., D.L. Johnson, S.F. Burns, and T.W. Small. 1989. Tree uprooting: Review of terminology, process, and environmental implications. Can. J. For. Res. 19:1–11.

Schlesinger, W.H. 1997. Biogeochemistry: An analysis of global change. Academic Press, San Diego, CA.

Schnitzer, M. 1969. Reactions between fulvic acid a soil humic compound and inorganic soil constituents. Soil Sci. Soc. Am. Proc. 33:75–81.

Schwertmann, U., and R.M. Taylor. 1989. Iron oxides, p. 379–438. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Simón, M., S. Sánchez, and I. García. 2000. Soil-landscape evolution on a Mediterranean high mountain. Catena 39:211–231.

Simonson, R.W. 1959. Outline of a generalized theory of soil genesis. Soil Sci. Soc. Am. Proc. 23:152–156.

Simonson, R.W. 1995. Airborne dust and its significance to soils. Geoderma 65:1–43.

Smeck, N.E., and L.P. Wilding. 1980. Quantitative evaluation of pedon formation in calcareous glacial deposits in Ohio. Geoderma 24:1–16.

Southard, R.J., and S.W. Buol. 1988. Subsoil blocky structure formation in some North Carolina Paleudults and Paleaquults. Soil Sci. Soc. Am. J. 52:1069–1076.

Sposito, G. 1989. The chemistry of soils. Oxford University Press, New York.

Stoner, M.G., and F.C. Ugolini. 1988. Arctic pedogenesis. II. Threshold-controlled subsurface leaching episodes. Soil Sci. 145:46–51.

Sullivan, L.A., and A.J. Koppi. 1991. Morphology and genesis of silt and clay coatings in the vesicular layer of a desert loam soil. Aust. J. Soil Res. 29:579–586.

Swap, R., M. Garstang, S. Greco, R. Talbot, and P. Kallberg. 1992. Saharan dust in the Amazon Basin. Tellus 44B:133–149.

Tardy, Y., and D. Nahon. 1985. Geochemistry of laterites, stability of Al-goethite, Al-hematite, and Fe3$^+$-kaolinite in bauxite and ferricretes. An approach to the mechanism of concretion formation. Am. J. Sci. 285:865–903.

Terhune, C.L., and J.W. Harden. 1991. Seasonal variations of carbon dioxide concentrations in stony, coarse-textured desert soils of southern Nevada, USA. Soil Sci. 151:417–429.

Tugel, A.J., J.E. Herrick, J.R. Brown, M.J. Mausbach, W. Puckett, and K. Hipple. 2005. Soil change, soil survey, and natural resources decision making: A blueprint for action. Soil Sci. Soc. Am. J. 69:738–747.

Turk, J.K., B.R. Goforth, R.C. Graham, and K.J. Kendrick. 2008. Soil morphology of a debris flow chronosequence in a coniferous forest, southern California, USA. Geoderma 146:157–165.

Ugolini, F.C., G. Corti, and G. Certini. 2006. Pedogenesis in the sorted patterned ground of Devon plateau, Devon Island, Nunavut, Canada. Geoderma 136:87–106.

Ugolini, F.C., and D.H. Mann. 1979. Biopedological origin of peatlands in south east Alaska. Nature 281:366–368.

Ugolini, F.C., and H. Spaltenstein. 1992. Pedosphere, p. 123–153. *In* S.S. Butcher and G.V. Wolfe (eds.) Global biogeochemical cycles. Academic Press, San Diego, CA.

Ulery, A.L., R.C. Graham, and C. Amrhein. 1993. Wood-ash composition and soil-pH following intense burning. Soil Sci. 156:358–364.

Van Breemen, N., J. Mulder, and C.T. Driscoll. 1983. Acidification and alkalization of soils. Plant Soil 75:283–308.

van Wambeke, A. 1992. Soils of the tropics. McGraw-Hill, New York.

van Wambeke, A., H. Eswaran, A.J. Herbillon, and J. Comerma. 1983. Oxisols, p. 325–354. *In* L.P. Wilding et al. (eds.) Pedogenesis and soil taxonomy. II. The soil orders. Elsevier Science Publishers, Amsterdam, the Netherlands.

Vandenberghe, J. 1992. Cryoturbations: A sediment structural analysis. Permafr. Periglac. Process. 3:343–352.

Vaughan, R.E., B.A. Needelman, P.J.A. Kleinman, and M.C. Rabenhorst. 2008. Morphology and characterization of ditch soils at an Atlantic Coastal Plain farm. Soil Sci. Soc. Am. J. 72:660–669.

Vaughan, K.L., M.C. Rabenhorst, and B.A. Needelman. 2009. Saturation and temperature effects on the development of reducing conditions in soils. Soil Sci. Soc. Am. J. 73:663–667.

Vepraskas, M.J. 1996. Redoximorphic features for identifying aquic conditions. NC Agric. Research Service., Tech. Bulletin. 301, Raleigh.

Vepraskas, M.J., and S.P. Faulkner. 2001. Redox chemistry of hydric soils, p. 85–106. *In* J.L. Richardson and M.J. Vepraskas (eds.) Wetland soils. Lewis Publishers, Boca Raton, FL.

Vitousek, P.M., O.A. Chadwick, T.E. Crews, J.H. Fownes, D.M. Hendricks, and D. Herbert. 1997. Soil and ecosystem development across the Hawaiian Islands. GSA Today 7:1–8.

Wan, J.M., and T.K. Tokunaga. 2002. Partitioning of clay colloids at air–water interfaces. J. Colloid Interface Sci. 247:54–61.

Washburn, A.L. 1980. Geocryology: A survey of periglacial processes and environments. John Wiley & Sons, New York.

Weitkamp, W.A., R.C. Graham, M.A. Anderson, and C. Amrhein. 1996. Pedogenesis of a vernal pool Entisol-Alfisol-Vertisol catena in southern California. Soil Sci. Soc. Am. J. 60:316–323.

Whipkey, C.E., R.C. Capo, O.A. Chadwick, and B.W. Stewart. 2000. The contribution of sea spray aerosol to the soil cation budget in a Hawaiian coastal environment. Chem. Geol. 168:37–48.

Whipkey, C.E., R.C. Capo, J.C.C. Hsieh, and O.A. Chadwick. 2002. Development of magnesian calcite and dolomite in quaternary soils on the island of Hawaii. J. Sediment. Res. 72:158–165.

Wilding, L.P., and D. Tessier. 1996. Genesis of Vertisols: Shrink/swell phenomena, p. 55–81. *In* L.P. Wilding and R. Puentes (eds.) Vertisols: Their distribution, properties, classification, and management. Texas A&M University Printing Center, College Station, TX.

Wood, Y.A., R.C. Graham, and S.G. Wells. 2005. Surface control of desert pavement pedologic process and landscape function, Cima volcanic field, Mojave Desert, California. Catena 59:205–230.

Young, M.H., E.V. McDonald, T.G. Caldwell, S.G. Benner, and D.G. Meadows. 2004. Hydraulic properties of a desert soil chronosequence in the Mojave desert, USA. Vadose Zone J. 3:956–963.

Zausig, J., W. Stepniewski, and R. Horn. 1993. Oxygen concentration and redox potential gradients in unsaturated model soil aggregates. Soil Sci. Soc. Am. J. 57:908–916.

Ziegler, K., O.A. Chadwick, M.A. Brzezinski, and E.F. Kelly. 2005a. Natural variation of $\delta^{30}Si$ ratios during progressive basalt weathering, Hawaiian Islands. Geochim. Cosmochim. Acta 69:4597–4610.

Ziegler, K., O.A. Chadwick, A.F. White, and M.A. Brzezinski. 2005b. $\delta^{30}Si$ systematics in a granitic saprolite, Puerto Rico. Geology 33:817–820.

31

Soil Taxonomy

Robert J. Ahrens
United States Department of Agriculture

Richard W. Arnold
United States Department of Agriculture

31.1 Conditions Favoring the Development of *Soil Taxonomy*

By 1960, the United States had 60 years of experience with a soil survey program, which mapped and interpreted soils in various parts of the country. Ever since the earliest mapping of areas with specialty croplands and problem saline soils, the primary purpose of the soil survey program has been to predict the consequences of alternative uses of soils.

During that time, significant events had taken place that affected the production and delivery of consistent products and services of the soil survey program. Among these were World War I, new and improved industries, enhanced energy distribution, gas motors and automobiles, mechanization of agriculture, and a shift from mainly family farms to more commercial enterprises. The Dust Bowl devastated the lives of many farmers as land degradation and unfavorable climatic conditions collided. New federal and state parks and forest reserves were established. World War II spanned the globe, followed by the Korean War. Economic reconstruction was promoted, and global markets expanded. Agriculture was also changing from a mode of unbridled production to one of increased emphasis on conservation of soil and water resources in more responsible ways. Many of these changes had impacts on the prediction of the consequences of alternative soil uses.

The model of soil changed during those times. When the soil survey started, it was perceived that soils were derived from the rocks, or from the transported materials, on which they rested, a classic geological conclusion. This concept gradually changed to that proposed by the Russian, Dokuchaev, where soils were considered to be the result of processes that were influenced by the interactions of soil-forming factors, namely, climate, biota, parent material, topography, and time. Soils were recognized as independent natural bodies worthy of study by, and for themselves, and, thus, the course of history of soil science in the United States began to change in the 1920s. Marbut promoted the independence of soils, presented his ideas of soil classification, and helped America become recognized in international affairs of soil science.

Similar climates existed over fairly large areas, as did vegetation groups such as forests and grasses, although both had microvariations. On more local scales, there were differences in parent materials and landform topography that comprised landscapes. The overlapping of the soil-forming factors in space was crucial to applying the model of soil as the result of processes that were influenced by the interactions of the soil-forming factors. Soil mappers discovered empirical relations linking sets of soil properties (generally called soil profiles) to specific landscape features, which represented soil factors. The correlation of soils with landscape segments was found to be consistent enough to be delineated on base maps and to be satisfactory for the purpose of the soil survey.

Soils, therefore, had certain predictability and so did their expected behavior. Where conditions had been the same, the soils would be the same; and where the soils were the same, those responses that depend on soil properties would be the same.

Where similar but not identical conditions or soils occurred, the soil responses would be similar, but not identical to those where the soils were identical.

Soil, which was considered to be a continuum covering the earth's terrestrial surface, could be subdivided into classes in a variety of ways, thereby creating a population or collection of individual soil bodies. Emphasis changed from thinking about the whole with loosely defined and indistinct parts to the concept in which the parts were sharply in focus and the whole was an organized collection of parts.

Locally, the individual soils were called soil types. The members of the same series, that is, soils having a similar sequence of the same kinds of horizons, were separated according to the general texture of the profile. This was a way to recognize different groups of parent material such as glacial till and loess. Initially, the soil types were grouped together into a soil series that had the same horizonation and commonality of properties. Later, soil type referred only to the texture of the surface soil and was considered as a phase of a soil series. The differences in parent materials became a basis for establishing separate series.

Mapping of soil types proceeded rapidly as there were hundreds of survey parties mapping in all parts of the country. Many new soil series were set up; however, an adequate system of correlation and classification lagged behind. It became more and more difficult to keep track of all the information being collected by the soil scientists, to compare soils from one region to another, and to communicate about soil properties and characteristics.

A number of differences in the responses to management and in land uses were found to vary geographically. Some regional variations were related to climate and age, and locally many variations were related to drainage condition and parent material. These observations supported the concept of important geographic differences and similarities and led to the need for some way of recognizing these similarities and differences. The significance of factors and factor interactions varied from region to region, and if reasonable ways to group or separate them could be found, they might serve as a basis for classification.

Over time the question about soil had changed from "How much yield can be expected from this soil with this amount of input?" to "How much input must be used on this soil to produce a given volume of produce?" This recognized that soil was dynamic and capable of modification and manipulation rather than being a static responder to management. Humans and their activities were accepted as major factors of soil formation; in fact, it was recognized that humans were the dominant influence on temporal soil quality. The changing concepts and attitudes about soil and the evolving world economic development were important contributors to the decision to develop a new soil classification system.

31.2 Recognition of Guiding Principles for a Soil Classification System

The leaders of the National Cooperative Soil Survey agreed that a classification system must serve the soil survey program of the United States. The program was undertaken for very practical reasons, namely, to identify and locate soils, and to predict the consequences of alternative uses of soils.

They indicated that a system should provide a basis for developing principles of soil genesis and behavior that would enable them to continue to provide predictions of soil behavior and responses to management and manipulation. Because soil genesis attempts to find cause and effect relationships rather than merely identifying empirical relationships, it was felt that genetic principles should serve well as a means of spatial extrapolation and to assist in interpreting soils for use (Smith, 1963).

It was clear to them that classes of a system must have real counterparts in mappable soil bodies. Some small geographic body was the entity to be classified; thus, the criteria of classification should not be applied independently of the values and variability within natural landscapes. This need to link with identifiable geographic bodies separated the practical from the purely hypothetical organization of soil property information (Cline, 1949).

It was implied that the component soil bodies in nature that were to be classified must be relevant to applied objectives. Geographic size and shape and soil attributes included within those areas were considered important for making relevant interpretations. Experience had demonstrated that this would be a problem in a classification with mutually exclusive classes; however, some of the concerns could be handled with mapping conventions.

Although the system would serve the practical uses of the soil survey, it was not implied that the classes themselves had to serve directly as the technical interpretive units. The leaders realized that valid interpretations of behavior and response likely would involve an additional step, level, or reasoning, which would rely on the known or assumed relationships between soil properties and behavior. This was also true for valid interpretations about soil genesis. It was recognized that no set of classes, or basis for groupings, or basis of subdivision, could provide units homogeneous enough for direct application to multiple objectives. Therefore, the classes had to be capable of being regrouped or subdivided as needed to satisfy different applied purposes. This was most important for the classes of the lowest category, which was the soil series (Smith, 1963).

There was agreement that a classification system must also serve to understand nature by revealing our understanding of soils as natural bodies. The accurate mapping of soils depended on this understanding. This departed from classification schemes of plants and animals where lineage relationships were assumed. This meant that science was to be basic to the process. In as much as such a system would be able to satisfy these ideas, it would be a reflection of the knowledge at a given point in time.

It was desired to have a system that could be applied uniformly by competent soil scientists. This was an important idea because it put soil scientists on an equal footing and allowed mappers to concentrate on other facets of the soil survey program. The new system would have to be objective in the sense that it should be based on properties of soils and not on beliefs of the classifier. The taxa should be defined, or recognized by observable or measurable properties selected to group soils of similar genesis, but

also capable of providing groups with similar properties, even if the genesis was uncertain or unknown. Soil genesis would be used to help select those properties that would provide useful groups of soils; however, the properties themselves would be the criteria for actual placement.

31.3 Science and Classification

Several lessons had been learned in trying to modify the 1938 classification. The concepts of some great soil groups were not clear and seemed to overlap with other groups. The methods of determining properties varied from place to place, or from time to time, and the correlations were not always certain. The mapping of soil types and soil series at the lower end of the spectrum, and the division of the pedosphere into regions and provinces and groups at the other end left an undecided middle ground with no good connections. There was great interest in not overly disturbing the thousands of soil series that had been established and used throughout the United States; yet there was a need to be able to link the series to classes in the higher categories of a classification system.

Mapping of soils and detailed laboratory studies of their properties gave rise to many consistent and repeatable correlations of soil features with landscape features. To maintain this consistency, the nature of the operations and methods used to obtain the data need to be specified and used. It is known that explanation is primarily a recognition of familiar correlations among phenomena in nature and that explanations are critical to understanding. It had been suggested that to go beyond empirical correlations to hypothesize the reasons for them will prejudice the future; therefore, there would be an effort to try to minimize using a priori principles, which determine or limit possibilities of new experiences. This meant that concepts of soil genesis would be useful to help select soil properties of interest, but that the operational character of soil facts would be used so as to not limit new experiences in the future.

There already were several thousand series considered to be classes of a lower category. They wanted some broader, more inclusive classes for making generalized maps, and an orderly organization of current understanding. It was also thought that they required flexibility to better serve other needs relying on soil information. Due to the magnitude and complexity of information, a hierarchical system of organization seemed most appropriate. For the most part, the logic of J.S. Mill as explained by Cline (1963) was followed. Some of the requirements follow.

Each category in a multicategoric system contains all members of the population of interest and is defined by abstracting the concepts that group the member classes together. Higher level categories are more abstract than those of the lower level.

The definition of a category suggests criteria that may be appropriate to separate the classes of that category. The criteria are then recognized by properties or features, which are called differentiating characteristics. These characteristics are observable or measurable properties of individuals that are grouped into a class.

Classes separated at a higher categorical level remain separated throughout the lower categories of the system. For ease of operation (placement of objects into the system), the classes are mutually exclusive, that is, without overlap. Operationally, the listing of the categories and the listing of the classes within each category need to be by priority because only a few of the characteristics of individuals are used to obtain placement within the scheme (Smith, 1983). This means that the scheme can be presented as a key or set of keys for consistent application of the definitions and the criteria (differentiating criteria).

The logic of the hierarchy represented by *Soil Taxonomy* is one of its major strengths, yet at the same time it retains the weaknesses of such organizational frameworks. The efficiency with which accessory characteristics are carried along with the differentiating ones is very powerful indeed. On the other hand, the uncertainty of the degree of accessory property relationships and the rigid class structure of the system limit the precision of statements that can be made for soils as they are observed in the field.

Recognition of a soil individual is crucial to the proper functioning of the system. Because the system was to support the soil survey, the individual had to be a recognizable geographic entity and one that could be characterized with limited sampling. The sample unit became the pedon and the soil individual became the polypedon. Their application has not been without some difficulty, which has continued even to the present.

The nomenclature of a hierarchical system is meaningful to the extent that there is consistent composition of the classes. A mnemonic scheme of phrases and combinations seemed a worthy goal. It was thought desirable to have a system of naming that would minimize previous biases and confusion and which might be translatable or usable in other languages. Consequently, roots of words from Latin and Greek and their combinations were proposed and tested. For some soil scientists the "new language" of soil classification was difficult to accept, but in retrospect was a brilliant idea.

And finally, it was agreed that the definitions must be continually tested by the nature of functioning of the soils grouped in a taxon. A taxonomy for the use of the soil survey needed to be tested by the nature of the interpretations that could be made (Smith, 1983). These interpretations included those for genesis and the mapping of soils, as well as for the behavior of soils under multiple uses.

31.4 Definitions of Categories of *Soil Taxonomy*

A category is the aggregate of classes that is formed by differentiation within a population on a single basis. A category includes the entire population; it includes all classes differentiated on one basis; it is distinguished from class that is only one part of a category; and it is definable only in terms of the basis of differentiation. In a hierarchy, the higher categories have fewer classes and they are more inclusive than the classes of the lower categories, which have accumulated attributes from all of the higher categories. Thus, high categorical levels are associated with high-level

generalizations or abstractions. These abstractions are used as the bases of differentiation; however, these ideas are expressed in terms of attributes that are assumed to be their consequences. Thus, soil attributes that are thought to be the results of such processes are the criteria used to segregate soils.

The definitions of categories, orders through series, although not stated very clearly, are intended to guide the selection and testing of properties and features used to characterize and classify soils. Soil properties that are thought to reflect processes or control processes are of great importance in this system. Although the processes of formation are influential in the selection of properties, it is the properties themselves that are used to determine the placement of soils into respective classes. The applications of quantified definitions and observations are the tests of adequacy of the theories behind this model of soil. Hence, *Soil Taxonom*y is considered a morphogenetic classification system.

31.4.1 Order

Classes at the order level are separated on the basis of properties resulting from the major processes and pathways of soil formation. Neither the genetic processes nor the courses of development are precisely known but the accepted concepts have influenced the selection of soil properties that are used to recognize and define the 12 classes currently considered. Many of the features are thought to have taken a reasonably long time to develop, are stable in a pedological sense, and are mostly static from a historical perspective.

31.4.2 Suborder

Classes at the suborder level are separated within each order on the basis of soil properties that are major controls, or reflect such controls on the current set of soil-forming processes. Most of the properties selected are dynamic such as soil moisture regime or cold soil temperatures. Other properties relate to materials or processes that retard horizon development, such as sand or alluvial sedimentation.

31.4.3 Great Group

Classes at the great group level are differentiated within each suborder on the basis of properties that constitute subordinate or additional controls, or reflect such controls on the current set of soil-forming processes. The properties selected are generally static, such as layers that retard percolation of water or root extension, but some are dynamic, such as the moisture regime where it was not a criterion at the suborder level.

31.4.4 Subgroup

Classes at the subgroup level are differentiated within each great group on the basis of properties resulting from either (1) a blending or overlapping of sets of processes in space or time that cause one kind of soil to develop from, or toward another kind of soil that has been recognized at the great group, suborder, or order level; or (2) sets of processes or conditions that have not been recognized as criteria for any class at a higher level. A third kind of subgroup fits neither (1) nor (2) but is considered to typify the central concept of the great group or simply represents what is not represented by the other subgroups.

31.4.5 Family

Classes at the family level are separated with the subgroup on the basis of properties that reflect important conditions affecting behavior or the potential for further change. Particle size, mineralogy, and soil depth are mainly capacity factors, whereas soil temperature and exchange activity are mainly intensity factors.

31.4.6 Series

Classes at the series level are separated within the family on the basis of properties that reflect relatively narrow ranges of soil-forming factors and of processes that transform parent materials into soils. Some properties are indicative of parent materials such as coarse fragments, sand or silt content, color, and horizon thickness or expression. Others reflect influences on processes such as differences in intensity or amount of precipitation, and depth to the presence or concentration of soluble compounds.

31.5 Differentiating Characteristics

The differentiating characteristics mentioned above are referred to in *Soil Taxonomy* as diagnostic horizons and characteristics or features. They are the building blocks of *Soil Taxonomy*. The diagnostic horizons and characteristics help define the criteria for the various taxa of *Soil Taxonomy*. They are also terms used by scientists to communicate the language of soil. Many soil scientists consider the concept of diagnostic horizons and features as one of the greatest contributions to soil classification.

31.5.1 Diagnostic Horizons

Diagnostic horizons are defined in the *Keys to Soil Taxonomy* (Soil Survey Staff, 2010). The intent in this chapter is not to reiterate the complete definition of the diagnostic horizons, but to provide a brief summary.

31.5.2 Diagnostic Epipedons

Diagnostic horizons that form at or near the surface are referred to as epipedons. There are eight epipedons recognized in *Soil Taxonomy*, which are briefly described below.

31.5.2.1 Anthropic Epipedon (Gr. *anthropos, Human Being*)

This is formed during long continuous use by humans, either as a place of residence (kitchen middens) or as a site for irrigated crops.

Anthropic epipedons often have the dark color of an umbric or a mollic epipedon, but were formed largely due to the actions of humans. Anthropic epipedons occur in a variety of soil orders, but are not widely recognized. Additional data and research are needed on this epipedon.

31.5.2.2 Folistic Epipedon (L. *folia*, Leaf)

This epipedon contains high amounts of organic C, but is not saturated with water for more than a few days after heavy rains. These epipedons occur mostly in cool, humid regions under forest vegetation. Folistic epipedons occur most commonly in Spodosols (see color insert) and Inceptisols (see color insert), but can occur in other orders.

31.5.2.3 Histic Epipedon (Gr. *histos*, Tissue)

This is 20–60 cm thick and contains high amounts of organic C. It differs from the folistic epipedon in that it is commonly saturated with water for long periods. Histic epipedons can occur in many soil orders.

31.5.2.4 Melanic Epipedon (Gr. *melas*, melan-, Black)

This is a thick, black horizon that contains high concentrations of organic C and short-range order minerals or Al–humus complexes. It commonly occurs in areas associated with a volcanic influence under forest vegetation. Melanic epipedons are associated with Andisols (see color insert for Andisol and Andisol (2)).

31.5.2.5 Mollic Epipedon (L. *mollis*, Soft)

This is a relatively thick, dark-colored, humus-rich surface horizon in which divalent cations are dominant on the exchange complex. It forms from the underground decomposition of organic residues, chiefly grasses, although particularly in ustic soil moisture regimes occur under a variety of vegetation, including shrubs and trees. Mollic epipedons have good structure and porosity and are associated with some of the richest agricultural soils in the world. Mollic epipedons are most commonly associated with Mollisols (see color insert for Mollisol and Mollisol (2)); they very rarely occur in other soil orders.

31.5.2.6 Ochric Epipedon (Gr. *ochros*, Pale)

This fails to meet the definition of any of the other diagnostic epipedons. It can be light colored, thin and dark colored, or even thin with high amounts of organic C. It has few or no accessory characteristics, but was defined to serve as a means of placing a name on the surface horizons that failed the criteria of any of the other seven epipedons. Ochric epipedons can occur in a variety of soil orders.

31.5.2.7 Plaggen Epipedon (Ger. *Plaggen*, Sod)

Plaggen epipedons are rare and occur mostly in Europe where in medieval times sod or other materials were used for bedding livestock, and the manure was spread on cultivated fields. The gradual additions of bedding materials raised the level of the fields. The plaggen epipedon is associated with Inceptisols.

31.5.2.8 Umbric Epipedon (L. *umbra*, Shade, Hence Dark)

Umbric epipedons resemble mollic epipedons but form largely under a forest vegetation and have a lower natural fertility (base saturation) than the mollic epipedon. The umbric epipedon is most commonly associated with Inceptisols, Andisols, Alfisols (see color insert), Ultisols (see color insert for Ultisol and Ultisol (2)), and Gelisols (see color insert).

31.5.3 Diagnostic Subsurface Horizons

The subsurface horizons form below the surface, but can be at or near the surface in eroded soils. Brief descriptions follow.

31.5.3.1 Agric Horizon (L. *ager*, Field)

This is an illuvial horizon, which has formed under cultivation and contains significant amounts of illuvial silt, clay, and humus. This horizon is relatively rare, or perhaps not readily recognized, and is often more associated with old world agriculture. Agric horizons are associated with Alfisols.

31.5.3.2 Albic Horizon (L. *albus*, White)

This is an eluvial horizon 1.0 cm or more thick with a color that is largely determined by the color of the primary sand and silt grains, rather than the color of their coatings. The clay and Fe oxides have been removed by pedogenesis. Albic horizons are most commonly associated with Spodosols, Alfisols, Ultisols, and Mollisols.

31.5.3.3 Argillic Horizon (L. *argilla*, White Clay)

An argillic horizon is commonly a subsurface horizon with a significantly higher content of phyllosilicate clay than the overlying horizon. There is also evidence of clay illuviation. Argillic horizons occur on stable landscapes and are most commonly associated with Alfisols, Ultisols, Aridisols (see color insert), and Mollisols.

31.5.3.4 Calcic Horizon (L. *calcis*, Lime)

This is an illuvial horizon in which secondary $CaCO_3$ has accumulated to a significant extent. In arid environments, precipitation is insufficient to move carbonates to great depths. In soils with water near the surface, capillary rise, evaporation, and transpiration concentrate $CaCO_3$ toward the surface. Calcic horizons occur in Aridisols, Alfisols, Andisols, Mollisols, Inceptisols, Vertisols (see color insert for Vertisol and Vertisol (2)), and Gelisols.

31.5.3.5 Cambic Horizon (L. *cambiare*, to Exchange)

The cambic horizon forms as a result of physical alterations, chemical transformations, accumulations, removals, or a combination of these processes. They are commonly identified by structure and a higher clay content, redder hue, or higher chroma than an overlying horizon. Cambic horizons are most commonly associated with Inceptisols, but they can occur in Aridisols, Mollisols, Vertisols, Andisols, and Gelisols.

31.5.3.6 Duripan (L. *durus*, Hard)

A duripan is cemented by illuvial silica. Duripans limit the downward growth of roots and movement of water. They are most commonly associated with the Aridisols, but are known to occur in Alfisols, Mollisols, Andisols, Inceptisols, Spodosols, and Vertisols.

31.5.3.7 Fragipan (L. *fragilis*, Brittle, and Pan)

This is a subsurface horizon that is noncemented, but restricts the entry of water and roots into the soil matrix. Fragipans most commonly are associated with Ultisols, Alfisols, Spodosols, and Inceptisols.

31.5.3.8 Glossic Horizon (Gr. *glossa*, Tongue)

The glossic horizon develops as a result of the degradation of an argillic, kandic, or natric horizon. Clay and Fe oxides are removed starting from the exterior of the peds. These horizons are transitional from argillic, natric, and kandic horizons to albic horizons. Glossic horizons occur most commonly in the Alfisols and Ultisols.

31.5.3.9 Gypsic Horizon (L. *gypsum*)

This is an illuvial horizon in which secondary gypsum has accumulated to a significant extent. Most gypsic horizons occur in arid environments where the parent materials are rich in gypsum. In soils with groundwater close to the surface, gypsum can accumulate by capillary rise, evaporation, and transpiration. Gypsic horizons most commonly occur in the Aridisols, but a few Inceptisols, Vertisols, and Gelisols also have gypsic horizons.

31.5.3.10 Kandic Horizon (Modified from Kandite)

This is a subsurface horizon with a significantly higher content of clay than an overlying horizon. Kandic horizons are dominated by low activity clays (1:1) and, therefore, have low cation exchange capacity (CEC). Kandic horizons occur in the Ultisols and Oxisols (see color insert for oxisol and oxisol (2)).

31.5.3.11 Natric Horizon (L. *natrium*, Sodium)

The natric horizon has all the characteristics of the argillic horizon (significant clay increase) and, in addition, has an accumulation of Na. Sodium adversely affects the physical properties of a soil. Of the soil orders, natric horizons are most commonly associated with the Aridisols, but they also occur in the Alfisols and Mollisols and to a very limited extent in the dry Gelisols.

31.5.3.12 Ortstein

Ortstein is a cemented horizon that consists of complexes of Al and organic matter with or without Fe (spodic materials). Ortstein limits root growth and the downward movement of water. Ortstein occurs in Spodosols.

31.5.3.13 Oxic Horizon (F. *oxide*)

This is a subsurface horizon of sandy loam or a finer texture with a low CEC and low amount of weatherable minerals. Oxic horizons occur in old, weathered soils or soils derived from highly weathered parent materials. Oxic horizons occur in Oxisols.

31.5.3.14 Petrocalcic Horizon (Gr. *petra*, Rock and Calcic)

This is cemented or indurated by $CaCO_3$ or Ca and Mg carbonates. It is a barrier to roots and the downward movement of water. Petrocalcic horizons are most commonly associated with the Aridisols, but also occur in the Alfisols, Mollisols, Vertisols, and Inceptisols.

31.5.3.15 Petrogypsic Horizon (Gr. *petra*, Rock and Gypsic)

This is cemented or indurated by gypsum. It is a barrier to roots and downward movement of water. Petrogypsic horizons are most commonly associated with Aridisols.

31.5.3.16 Placic Horizon

This is a thin (<25 mm), dark-colored horizon that is cemented by either Mn or Fe and Mn and organic matter. Commonly, they occur in moist, cool climates and are associated with the Spodosols and Inceptisols.

31.5.3.17 Salic Horizon (L. *sal*, Salt)

This is an accumulation of salts that are more soluble than gypsum in water. Many plants are intolerant of high concentrations of salt. Most salic horizons are associated with the Aridisols; a few occur in the drier Gelisols.

31.5.3.18 Sombric Horizon (F. *sombre*, Dark)

This is an illuvial horizon that contains humus that is not associated with Al nor dispersed by Na. It is largely confined to the cool, moist soils of high plateaus or mountains of the tropics and subtropics. Sombric horizons are associated with Inceptisols, Oxisols, and Ultisols.

31.5.3.19 Spodic Horizon (Gr. *spodos*, Wood Ash)

This is an illuvial horizon that contains active amorphous materials composed of organic matter and Al with or without Fe. Spodic horizons are associated with Spodosols.

31.5.3.20 Sulfuric Horizon (L. *sulfur*, Sulfur)

The sulfuric horizon forms when sulfide rich and organic materials are oxidized as a result of drainage, most commonly artificial. Sulfuric horizons are toxic to most plants. They occur in Entisols (see color insert), Histosols (see color insert), and Inceptisols.

31.5.4 Other Diagnostic Soil Characteristics

The diagnostic soil characteristics are features of the soil that are used repeatedly in *Soil Taxonomy*.

Abrupt textural change is a considerable increase in clay within a short distance, a very important feature used to help predict water movement in soils.

Albic materials are light-colored soil materials that reflect the color of the primary sand and silt particles rather than the color of the coatings.

Andic soil properties result from the presence of significant amounts of allophane, imogolite, ferrihydrite, or Al–humus complexes.

Anhydrous conditions refer to the moisture condition of soils in very cold deserts and other areas that have dry permafrost.

Aquic conditions arise in soils that currently experience continuous or periodic saturation and reduction. The presence of these features is indicated by redoximorphic features. There are several types of saturation: Episaturation describes a perched water table, endosaturation describes a groundwater table, and anthraquic saturation describes controlled flooding such as that used to grow rice and cranberries.

Coefficient of linear extensibility (COLE) is the ratio of the difference between the moist and dry lengths of a clod to its dry length. COLE is used to determine the shrink/swell potential of a horizon.

Cryoturbation (frost churning) is the mixing of the soil matrix that results in irregular or broken horizons.

Densic contact occurs between soil and densic materials and is a barrier to roots.

Densic materials are noncemented, relatively unweathered, mostly earthy materials such as till, volcanic mudflows, and noncemented rocks.

Durinodes are cemented to indurated nodules with SiO_2, presumably opal and microcrystalline forms of silica.

Fragic soil properties are similar to a fragipan, but do not have the required thickness or volume for a fragipan.

Gelic materials are soil materials that show evidence of cryoturbation or ice segregation in the active layer.

Glacic layer is massive ice in the form of ice lenses and wedges.

Identifiable secondary carbonates refer to translocated authigenic $CaCO_3$ that has been precipitated in place from the soil solution rather than precipitated in place from the parent material.

Interfingering of albic materials refers to albic materials that penetrate 5 cm or more into an underlying horizon.

Lamellae are thin (<7.5 cm) illuvial horizons that contain evidence of illuvial clay and occur only in coarse-textured soils.

Linear extensibility is the product of the thickness (cm) multiplied by the COLE of the soil layer in question and is used to predict shrink and swell.

Lithic contact occurs between soil and a coherent underlying material that is in a strongly or more cemented class. Lithic contacts occur at the boundary between soil and hard bedrock.

Lithologic discontinuities are significant differences in particle-size distribution or mineralogy that represent differences in lithology of a soil.

N value is used to predict if a soil can support loads and to the degree of subsidence after drainage.

Paralithic contact occurs between soil and paralithic materials (defined below) such as between soil and unconsolidated or weathered bedrock.

Paralithic materials are relatively unaltered (by pedogenesis) materials that are in a moderately or less cemented class.

Permafrost refers to soil material that remains below 0°C for 2 or more years in succession.

Petroferric contact is a boundary between soil and a continuous layer of indurated material in which Fe is an important cement and organic matter is either absent or present in trace amounts.

Plinthite is an Fe-rich, humus-poor material that irreversibly hardens when exposed to wetting and drying, especially if heated by the sun.

Resistant minerals refer to durable minerals in the 0.02–2.0 mm fraction, such as quartz, zircon, etc.

Slickensides are polished and grooved surfaces produced by one mass sliding past another. They are a feature of soils with a high capacity to shrink and swell.

Spodic materials are illuvial, amorphous materials composed of organic material and aluminum with or without iron.

Sulfidic materials contain oxidizable sulfur compounds.

31.5.5 Soil Moisture and Temperature Regimes

Soil Taxonomy is unique among many systems used to classify soils because it recognizes the temperature of a soil and the moisture status of a soil over time. Both temperature and moisture are important properties of the soil and convey essential information about soils. Because complete definitions for the soil moisture and temperature regimes can be found in the *Keys to Soil Taxonomy* (Soil Survey Staff, 2010), only brief descriptions are presented here.

31.5.5.1 Soil Moisture Regimes

For most soil orders, moisture regime is used to determine placement of a soil at the suborder level.

Aquic moisture regime signifies a reducing regime virtually free of dissolved O_2 from saturation with water. Most soils that have an aquic moisture regime are not saturated with water to the soil surface year round.

Aridic (torric) moisture regime applies to soils that commonly occur in arid climates and are dry in all parts more than half the time during the growing season and moist in some or all parts for less than 90 consecutive days during the growing season.

Udic moisture regime applies to soils that occur in climates with well-distributed rainfall or sufficient summer rain so that rainfall plus stored moisture equals or exceeds the amount of evaporation.

Ustic moisture regime is moister than aridic and drier than udic. The concept of ustic is one of limited moisture, but at least some moisture at a time when conditions are suitable for plant growth.

Xeric moisture regime applies to soils that occur in climates with cool, moist winters and warm, dry summers.

Perudic moisture regime applies to soils that occur in areas where precipitation exceeds evaporation in all months when the soil is not frozen.

Order	Suborder	Great Group	Subgroup	Family	Series
Aridisols	Argids	Calciargids	Ustic Calciargids	Fine-loamy, mixed, superactive, mesic	Barx "Clovis, Flaco".

31.5.5.2 Soil Temperature Regimes

The following soil temperature regimes are based on the mean annual temperature of a soil measured at 50 cm or at a lithic, paralithic, or densic contact, whichever is shallowest.

Cryic soil temperature regime applies to soils that have a mean annual soil temperature of <8°C, but no permafrost and a summer temperature cooler than soils in a frigid regime.

Frigid soil temperature regime applies to soils that have a mean annual soil temperature of <8°C.

Mesic temperature regime applies to soils that have a mean annual soil temperature of 8°C but <15°C.

Thermic temperature regime applies to soils that have a mean annual soil temperature of 15°C but <22°C.

Hyperthermic temperature regime applies to soils that have a mean annual soil temperature of 22°C.

When the difference between mean summer and winter soil temperatures is less than 6°C, *iso* is added to the name. *isofrigid*, *isomesic*, *isothermic*, and *isohyperthermic* are the only classes used. Soil temperature is most commonly used at the family level in *Soil Taxonomy*.

31.6 Categories of *Soil Taxonomy*

The information that follows is largely from the second edition of *Soil Taxonomy* (Soil Survey Staff, 1999). The diagnostic horizons and characteristics help define the categories of *Soil Taxonomy*.

A category of *Soil Taxonomy* is a set of classes that is defined approximately at the same level of generalization or abstraction and includes all soils. There are six categories or levels in *Soil Taxonomy*. In order of decreasing rank and increasing number of differentiae and classes, the categories are order, suborder, great group, subgroup, family, and series.

The nomenclature of *Soil Taxonomy* is based on the following premises: Each taxon requires a name if it is to be used in speech; a good name is short, easy to pronounce, and distinctive in meaning; a name is connotative, that is, capable of mnemonic attachment to the concept of the thing itself (Heller, 1963); it is useful if the name of a taxon indicates its position in the classification, if similarities in important properties are reflected by similarities in names, if the mnemonic attachments hold in many languages, and if the name fits into many languages without translation.

The name of each taxon above the category of series indicates its class in all categories of which it is a member. The name of a soil series indicates only the category of series. Thus, a series name may be recognized as a series, but one cannot tell from the name to what order, suborder, and so on, it belongs.

Table 31.1 gives a few examples of names of taxa in each category from order to series. Because of the system for assigning names and because most formative elements carry the same meaning in any combination, a name can convey a great deal of information about a soil.

31.7 Recognition of the Categories

31.7.1 Orders

The name of each order ends in sol (L. *solum*, soil) with the connecting vowel *o* for Greek roots and *i* for other roots. Each name of an order contains a formative element that begins with the vowel immediately preceding the connecting vowel and ends with the last consonant preceding the connecting vowel. In the order name, Entisol, the formative element is *ent*. In Aridisol, it is *id*. These formative elements are used as endings for the names of suborders, great groups, and subgroups. Thus, the names of all taxa higher than the series that are members of the Entisol order end in *ent* and can be recognized as belonging to the order of Entisol. Names ending in *id* are names of taxa belonging to the order of Aridisols. Table 31.2 lists all the soil orders and their formative elements.

31.7.2 Suborders

Names of suborders have exactly two syllables. The first syllable connotes something of the diagnostic properties of the soils or the soil moisture regime. The second is the formative element from the name of the order. Thirty formative elements

TABLE 31.2   Formative Elements in Names of Soil Orders in *Soil Taxonomy*

Formative Element in Name	Name of Order	Derivation of Formative Element	Pronunciation of Formative Element
Alf	Alfisol	Meaningless syllable	Pedalfer
And	Andisol	Modified from ando	Ando
Id	Aridisol	L. *aridus*, dry	Arid
Ent	Entisol	Meaningless syllable	Recent
El	Gelisol	L. *gelare*, to freeze	Jell
Ist	Histosol	Gr. *histos*, tissue	Histology
Ept	Inceptisol	L. *inceptum*, beginning	Inception
Oll	Mollisol	L. *mollis*, soft	Mollify
Ox	Oxisol	F. *oxide*, oxide	Oxide
Od	Spodosol	Gr. *spodos*, wood ash	Odd
Ult	Ultisol	L. *ultimus*, last	Ultimate
Ert	Vertisol	L. *verto*, turn	Invert

TABLE 31.3 Formative Elements in Names of Suborders in *Soil Taxonomy*

Formative Element	Derivation	Connotation
Alb	L. *albus*, white	Presence of albic horizon
Anth	Gr. *anthropos*, human	Modified by humans
Aqu	L. *aqua*, water	Aquic conditions
Ar	L. *arare*, to plow	Mixed horizon
Arg	L. *argilla*, white clay	Presence of argillic horizon
Cal	L. *calcis*, lime	Presence of calcic horizon
Camb	L. *cambiare*, to exchange	Presence of cambic horizon
Cry	Gr. *kryos*, icy cold	Cold
Dur	L. *durus*, hard	Presence of duripan
Fibr	L. *fibra*, fiber	Least decomposed stage
Fluv	L. *fluvius*, river	Floodplain
Fol	L. *folia*, leaf	Mass of leaves
Gel	L. *gelare*, to freeze	Jell
Gyps	L. *gypsum*, gypsum	Presence of gypsic horizon
Hapl	Gr. *haplos*, simple	Minimum horizon development
Hem	Gr. *hemi*, half	Intermediate stage of decomposition
Hist	Gr. *histos*, tissue	Presence of organic materials
Hum	L. *humus*, earth	Presence of organic matter
Orth	Gr. *orthos*, true	The common ones
Per	L. *per*, throughout time	Perudic moisture regime
Psamm	Gr. *psammos*, sand	Sandy texture
Rend	Modified from rendzina	High carbonate content
Sal	L. *sal*, salt	Presence of salic horizon
Sapr	Gr. *saprose*, rotten	Most decomposed stage
Torr	L. *torridus*, hot and dry	Torric moisture regime
Turb	L. *turbidis*, disturbed	Presence of cryoturbation
Ud	L. *udus*, humid	Udic moisture regime
Ust	L. *ustus*, burnt	Ustic moisture regime
Vitr	L. *vitrum*, glass	Presence of glass
Xer	Gr. *xeros*, dry	Xeric moisture regime

(Table 31.3) are used with the 12 formative elements from names of the orders to make names of over 65 suborders. The suborder of Entisols that has aquic conditions throughout is called Aquents (L. *aqua*, water, plus *ent* from Entisol). The formative element *aqu* is used with this meaning in 9 of the 12 orders. The suborder of Entisols that consists of very young sediments is called Fluvents (L. *fluvius*, river, plus *ent* from Entisol).

31.7.3 Great Groups

The name of a great group consists of the name of a suborder and a prefix that consists of one or two formative elements suggesting something of the diagnostic properties. Names of great groups, therefore, have three or four syllables and end with the name of a suborder. Fluvents that have a cryic temperature regime are called Cryofluvents (Gr. *kryos*, icy cold, plus fluvent). Fluvents that have a torric moisture regime are called Torrifluvents (L. *torridus*, hot and dry). The formative elements for the great groups are listed in Table 31.4.

31.7.4 Subgroups

The name of a subgroup consists of the name of a great group modified by one or more adjectives. The adjective, typic, represents in some instances what is thought to typify the great group and in other instances, typic subgroups simply do not have any of the characteristics used to define the other subgroups in a great group. Each typic subgroup has, in clearly expressed form, all the diagnostic properties of the order, suborder, and great group to which it belongs. Typic subgroups also have no additional properties indicating a transition to another great group. Typic subgroups are not necessarily the most extensive subgroup of a great group.

Intergrade subgroups are those that belong to one great group but that have some properties of another order, suborder, or great group. They are named by using the adjectival form of the name of the appropriate taxon as a modifier of the great group name. Thus, the Torrifluvents that have some of the properties or properties closely associated with Vertisols are called Vertic Torrifluvents. Vertic Torrifluvents have some of the properties of Vertisols superimposed on the complete set of diagnostic properties of Torrifluvents.

Extragrade subgroups are those that have important properties that are not representative of the great group but that do not indicate transitions to any other known kind of soil. They are named by modifying the great group name with an adjective that connotes something of the nature of the aberrant properties. Thus, a Cryorthent that has bedrock that is at least strongly cemented within 50 cm of the mineral soil surface is called a Lithic Cryorthent (lithic, Gr. *lithos*, stone). This subgroup is listed as an example in Table 31.5.

31.7.5 Families

Names of families are polynomial. Each consists of the name of a subgroup and descriptive terms, generally three or more, to indicate the particle-size class (or combinations thereof, if strongly contrasting), the mineralogy (26 classes), the cation exchange activity (4 classes), the calcareous and reaction (4 classes), the temperature (8 classes), and, in a few families, depth of soil (3 classes), rupture resistance (2 classes), and classes of coatings and classes of cracks (3 classes). Names of most families have three to five descriptive terms that modify the subgroup name, but a few have only one or two and a few as many as six. An example given in Table 31.1 is a family of fine-loamy (particle size), mixed (mineralogy), superactive (cation exchange activity), and mesic (soil temperature) Ustic Calciargids.

31.7.6 Series

Names of series as a rule are abstract place names. The name usually is taken from a place near where the series was first recognized. It may be the name of a town, a county, or some local feature. Some series have coined names while many have been carried over from earlier classifications with some having been

TABLE 31.4 Formative Elements in Names of Great Groups in *Soil Taxonomy*

Formative Element	Derivation	Connotation
Acr	Gr. *akros*, at the end	Extreme weathering
Al	Modified from aluminum	High aluminum, low iron
Alb	L. *albus*, white	Presence of an albic horizon
Anhy	Gr. *anhydros*, waterless	Very dry
Anthr	Gr. *anthropos*, human	An anthropic epipedon
Aqu	L. *aqua*, water	Aquic conditions
Arg	L. *argilla*, white clay	Presence of an argillic horizon
Calc	L. *calcis*, lime	A calcic horizon
Cry	Gr. *kryos*, icy cold	Cold
Dur	L. *durus*, hard	A duripan
Dystr (Dys)	Gr. *dys*, ill; dystrophic, infertile	Low base saturation
Endo	Gr. *endo*, within	Implying a groundwater table
Epi	Gr. *epi*, on, above	Implying a perched water table
Eutr (Eu)	Gr. *eu*, good; eutrophic, fertile	High base saturation
Ferr	L. *ferrum*, iron	Presence of iron
Fibr	L. *fibra*, fiber	Least decomposed stage
Fluv	L. *fluvius*, river	Floodplain
Fol	L. *folia*, leaf	Mass of leaves
Frag	L. *fragilis*, brittle	Presence of fragipan
Fragloss	Compound of fra(g) and gloss	See Frag and Gloss
Fulv	H. *fulvus*, dull	Dark brown color, presence of organic C
Gel	L. *gelare*, to freeze	Jell
Glac	L. *glacialis*, icy	Ice lenses or wedges
Gyps	L. *gypsum*, gypsum	Presence of gypsic horizon
Gloss	Gr. *glossa*, tongue	Presence of glossic horizon
Hal	Gr. *hals*, salt	Salty
Hapl	Gr. *haplos*, simple	Minimum horizon
Hem	Gr. *hem*, half	Intermediate stage of decomposition
Hist	Gr. *histos*, tissue	Presence of organic materials
Hum	L. *humus*, earth	Presence of organic matter
Hydr	Gr. *hydor*, water	Presence of water
Kand	Modified from kandite	1:1 layer silicates
Luv	Gr. *louo*, to bath	Illuvial
Melan	Gr. *melasanos*, black	Black, presence of organic C
Molli	L. *mollis*, soft	Presence of a mollic epipedon
Natr	L. *natrium*, sodium	Presence of a natric horizon
Pale	Gr. *paleos*, old	Excessive development
Petro	Gr. *petra*, rock	A cemented horizon
Plac	Gr. *plax*, flat stone	Presence of a thin pan
Plagg	Ger. *plaggen*, sod	Presence of a plaggen epipedon
Plinth	Gr. *plinthos*, brick	Presence of plinthite
Psamm	Gr. *psammos*, sand	Sand texture
Quartz	Ger. *quart*, quartz	High quartz content
Rhod	Gr. *rhodon*, rose	Dark red color
Sal	L. *sal*, salt	Presence of a salic horizon
Sapr	Gr. *saprose*, rotten	Most decomposed stage
Somb	F. *sombre*, dark	Presence of sombric horizon
Sphagn	Gr. *sphagnos*, bog	Presence of sphagnum
Sulf	L. *sulfur*, sulfur	Presence of sulfides or their oxidation products
Torr	L. *torridus*, hot	Tonic moisture regime and dry
Ud	L. *udus*, humid	Udic moisture regime
Umbr	L. *umbra*, shade	Presence of an umbric horizon
Ust	L. *ustus*, burnt	Ustic moisture regime
Verm	L. *vermes*, worm	Wormy or mixed by animals
Vitr	L. *vitrum*, glass	Presence of volcanic glass
Xer	Gr. *xeros*, dry	Xeric moisture regime

TABLE 31.5 Adjectives in Names of Extragrades in *Soil Taxonomy* and Their Meaning

Adjective	Derivation	Connotation
Abruptic	L. *abruptum*, torn off	Abrupt textural change
Aeric[a]	Gr. *aerios*, air	Aeration
Albic	L. *albus*, white	Presence of an albic horizon
Alic	Modified from aluminum	High Al^{3+} status
Anionic	Gr. *anion*	Positively charged colloid
Anthraquic	Gr. *anthropos*, human and L. *aqua*, water	Controlled flooding
Anthropic	Gr. *anthropos*, human	An anthropic epipedon
Arenic	L. *arena*, sand	Sandy between 50 and 100 cm thick
Calcic	L. *calcis*, lime	Presence of a calcic horizon
Chromic	Gr. *chroma*, color	High chroma
Cumulic	L. *cumulus*, heap	Thickened epipedon
Durinodic	L. *durus*, hard	Presence of durinodes
Dystic	Gr. *dys*, ill	Low base status
Eutric	Gr. *eu*, good; eutrophic, fertile	High base status
Fragic	L. *fragilis*, brittle	Presence of fragic properties
Glacic	Gr. *glacialis*, icy	Presence of ice lenses or wedges
Glossic	Gr. *glossa*, tongue	Tongued horizon boundaries
Grossarenic	L. *grossus*, thick; *arena*, sand	Thick, sandy layer
Gypsic	L. *gypsum*, gypsum	Presence of a gypsic horizon
Halic	Gr. *hals*, salt	Salty
Humic	L. *humus*, earthy	Presence of organic matter
Hydric	Gr. *hydor*, water	Presence of water
Kandic	Modified from kandite	Presence of 1:1 layer silicates
Lamellic	L. *lamella*, dim	Presence of lamellae
Leptic	Gr. *leptos*, thin	A thin soil
Limnic	Gr. *limn*, lake	Presence of a limnic layer
Lithic	Gr. *lithos*, stone	Presence of a shallow lithic contact
Natric	L. *natrium*, sodium	Presence of sodium
Nitric	Modified from *nitron*	Presence of nitrate salts
Ombroaquic	Gr. *ombros*, rain; L. *aqua*, water	Surface wetness
Oxyaquic	Oxy representing oxygen and aquic	Aerated
Pachic	Gr. *pachys*, thick	A thick epipedon
Petrocalcic	Gr. *petra*, rock; L. *calcis*, lime	Presence of a petrocalcic horizon
Petroferric	Gr. *petra*, rock; L. *ferrum*,, iron	Presence of a petroferric contact (ironstone)
Petrogypsic	Gr. *petra*, rock; L. *gypsum*, gypsum	Presence of a petrogypsic horizon
Petronodic	Gr. *petra*, rock; L. *nodulus*, a little knot	Presence of concretions and/or nodules
Placic	Gr. *plax*, flat stone	Presence of a thin pan (placic horizon)
Plinthic	Gr. *plinthos*, brick	Presence of plinthite
Rhodic	Gr. *rhodon*, rose	Dark red color
Ruptic[a]	L. *ruptum*, broken	Intermittent or broken horizons
Sodic	Modified from sodium	Presence of sodium
Sombric	F. *sombric*, dark	Presence of sombric horizon
Sulfic	L. *sulfur*, sulfur	Presence of sulfides or their oxidation products
Terric	L. *terra*, earth	A mineral substratum
Thapto(ic)[a]	Gr. *thapto*, buried	A buried soil
Ultic	L. *ultimus*, last	Low base status
Umbric	L. *umbra*, shade	Presence of umbric epipedon
Xanthic	Gr. *xanthos*, yellow	Yellow color

[a] Not strictly an extragrade. Name used to indicate a special departure from the typic subgroup.

in use since 1900. The name of a series carries no meaning to people who have no other source of information about the soils in it. There are over 21,000 soil series currently recognized in the United States.

The Barx, Clovis, and Flaco series (Table 31.1) are three members of the fine-loamy, mixed, superactive, mesic family of Ustic Calciargids. The meaning of each of these terms is defined later, but in a general way the name tells us the following.

Fine-loamy means that from a depth of 25–100 cm there is no marked contrast in particle-size class, the content of clay is between 18% and 35%, 15% or more of the material is coarser than 0.1 mm in diameter (fine sand to very coarse sand plus gravel), but less than 35% by volume of the material is rock fragments 2.0 mm or more in diameter (less than about 50% by weight). The average texture, then, is more likely to be loam, clay loam, or sandy clay loam. *Mixed* means a mixed mineralogy that is less than 40% of any one mineral other than quartz in the fraction between 0.02 and 2.0 mm in diameter; less than 20% (by weight) glauconitic pellets in the fine earth fraction; a total iron plus gibbsite (by weight) in the fine earth fraction of 5% or less; a fine earth fraction that has at least one of the following: free carbonates, the pH of a suspension of 1 g soil in 50 mL 1 M NaF of 8.4 or less after 2 min, or a ratio of −1500 kPa water to measured clay of 0.6 or less. *Superactive* means the CEC divided by the clay content (%) is 0.60 or more, which indicates the soil is relatively high in bases. *Mesic* indicates a mesic temperature regime, that is, the mean annual soil temperature is between 8°C and 15°C (47°F and 59°F) and the soil temperature fluctuates more than 6°C between summer and winter. In other words, the soil is somewhere in the midlatitudes, summer is warm or hot, and winter is cool or cold. *Depth of soil* when no class is used, in Ustic Calciargids, means the soil is 50 cm or more deep.

It is not necessary to know the exact criteria in the *Keys to Soil Taxonomy* in order to communicate useful information about a soil. It is necessary to know the language of *Soil Taxonomy*. For example, from Ustic Calciargids one knows the following.

Ustic at the subgroup level of an Aridisol means that the soil is moister than what is typical for an Aridisol or aridic soil moisture regime and close to having an ustic moisture regime. *Calci* means the soil has a calcic horizon within 150 cm of the soil surface. *Argi* denotes presence of an argillic horizon. The *id* means it is an Aridisol. Therefore, one can say that the soil occurs in an arid, temperate climate. Presence of an argillic horizon implies that the soil occurs on stable, older landscape. It has a calcic horizon and under irrigation, Fe chlorosis may be a problem in sensitive plants. If the soil is not irrigated, it can be used only for limited grazing.

31.8 Forming Names

31.8.1 Names of Orders, Suborders, and Great Groups

The names of the orders, suborders, and great groups that are currently recognized are presented in Chapter 33 under each order.

31.8.1.1 Names of Subgroups

The name of a subgroup consists of the name of a great group modified by one or more adjectives. As explained earlier, the adjective typic is used for the subgroup that is thought to typify the central concept of the great group, or for soils that fail to meet the criteria of the other subgroups defined for a great group.

Intergrade subgroups that have, in addition to the properties of their great group, some properties of another taxon carry the name of the other taxon in the form of an adjective. The names of orders, suborders, or great groups or any of the prior (first) formative elements of those names may be used in the form of an adjective in subgroup names. A few soils may have aberrant properties of two great groups that belong in different orders or suborders. For these, it is necessary to use two names of taxa as adjectives in the subgroup name.

Extragrade subgroups carry the name of one or more special descriptive adjectives (Table 31.5) to modify the name of the great group and connote the nature of the aberrant properties.

31.8.1.2 Names of Intergrades Toward Other Great Groups in the Same Suborder

If the aberrant property of a soil is one that is characteristic of another great group in the same suborder, only the distinctive formative element of the great group name is used to indicate the aberrant properties. Thus, Typic Argidurid is defined, in part, as having an indurated or very strongly cemented duripan. If the only aberrant feature of an Argidic Argidurid is that the duripan is strongly cemented or less cemented throughout, it is considered to intergrade toward the Argids. The name, however, is Argidic Argidurids, not Haplargidic Argidurids. Only the prior (first) formative element is used in adjectival form if the two great groups are in the same suborder.

31.8.1.3 Names of Intergrades Toward a Great Group in the Same Order but Different Suborder

Two kinds of names have been chosen here. If the only aberrant features are color and moisture regime and the hue is too yellow or the chroma is too high or too low for the typic subgroup, the adjectives aeric or aquic are used.

If an Epiaquult has chroma too high for the typic subgroup, but has no other aberrant feature, it is placed in an aeric subgroup. Using an adjective taken from the suborder, udic, would not suggest that the difference is one of aeration alone.

If the only aberrant feature of a Hapludult is redoximorphic features that are too shallow for a Typic Hapludult, the adjective aquic is used in the subgroup name. If redox depletions (accompanied by aquic conditions, unless artificially drained) appear within the upper 60 cm of the argillic horizon, the soil is called an Aquic Hapludult, not an Aquultic Hapludult.

In other instances, the adjective in the subgroup name is made from the first two formative elements of the appropriate great group name in that suborder. For example, if a Paleudult has both shallow redoximorphic features and some plinthite, it is called a Plinthaquic Paleudult, not a Plinthaquultic Paleudult.

Note that the formative element for the order is not repeated in the adjective if the two great groups are in the same order.

31.8.1.4 Names of Subgroups not Intergrading Toward Any Known Kind of Soil (Extragrade)

Some soils have aberrant properties that are not characteristic of a class in a higher category of any order, suborder, or great group. One example might be taken from the concave pedons that are at the base of slopes, in depressions, or in other places where new soil material accumulates slowly on the surface. In these soils, material is added to the A horizon. The presence of an overthickened A horizon is not used to define any great group, but the soils lie outside the range of the typic subgroup and there is no class toward which they intergrade. Hence, a descriptive adjective is required. For this particular situation, the adjective cumulic (L. *cumulus*, heap, plus ic, Gr. *ikos*) is used to form the subgroup names. Pachic is used to indicate an overthickened epipedon if there is no evidence of new material at the surface.

Other soils lie outside the range of typic subgroups in an opposite direction. Such soils are, in effect, truncated by hard rock and are shallow or are intermittent between rock outcrops. They are, in effect, intergrades to nonsoil and are called lithic subgroups. The names of formative elements in groups of this sort, which are called extragrades, are listed together with their derivation in Table 31.5.

31.8.2 Names of Families

Each family requires one or more names. The technical family name consists of a series of descriptive terms modifying the subgroup name. For these terms the class names that are given later for particle-size class, mineralogy, and so on, in family differentiae are used. To have consistent nomenclature, the order of descriptive terms in names of families is particle-size class, mineralogy class, cation exchange activity class, calcareous and reaction class, soil temperature class, soil depth class, rupture resistance class, classes of coatings, and classes of cracks.

Redundancy in names of families is avoided. Particle-size class and temperature classes should not be used in the family name if they are specified above the family. Psamments, by definition, all have a sand or loamy sand texture and are in a sandy particle-size class, unless they are ashy. It is, therefore, redundant to use a particle-size class for Psamments, unless they are ashy.

References

Cline, M.G. 1949. Basic principles of soil classification. Soil Sci. 67:81–91.

Cline, M.G. 1963. Logic of the new system of soil classification. Soil Sci. 96:17–22.

Heller, J.L. 1963. The nomenclature of soils or what's in a name? Soil Sci. Soc. Am. Proc. 27:216–220.

Smith, G.D. 1963. Objectives and basic assumptions of the new classification system. Soil Sci. 96:6–16.

Smith, G.D. 1983. Historical development of soil taxonomy: Background, p. 23–49. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy. I. Concepts and interactions. Elsevier, New York.

Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd edn. USDA handbook 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. U.S. Government Printing Office, Washington, DC.

32

Other Systems of Soil Classification

32

Other Systems of Soil Classification

Erika Michéli
Szent István University

Otto C. Spaargaren
*World Reference and
Information Centre*

32.1 Introduction

Soil classification is probably as old as farming. The fact that, around 8000 BP, the first farming communities in Europe settled on the better loess soils indicates that, during these times, farmers were already capable of distinguishing between the more and less productive soils. The oldest historical record of soil classification is most likely the Chinese book, *Yugong*, in which the soils of China were classified into three categories and nine classes, based on soil color, texture, and hydrological features (Gong, 1994). Even today, such criteria are still in use by farmers to differentiate soils. Studies on indigenous soil knowledge in northern Ghana, for example, have shown that farmers use texture, color, stoniness, and soil depth to stratify the soils (Asiamah et al., 1997).

This chapter will describe (1) the *Legend of the Soil Map of the World* (FAO–UNESCO, 1974), its revised version (FAO–UNESCO, 1988), and the *World Reference Base for Soil Resources* (WRB) (ISSS–ISRIC–FAO, 1994, 1998; IUSS Working Group WRB, 2006) as international systems of soil classification and correlation; (2) the French system of soil classification (CPCS, 1967), which is still the system in use in large parts of West and Central Africa and the new *Référentiel Pédologique Français* (AFES–INRA, 1990, 1992); (3) the systems of soil classification used in Russia (VASKhNIL, 1986; Shishov et al., 2004); (4) the recently developed soil classification system of China (Gong and Zhang, 2007); and (5) the newly developed system of soil classification for Australia (Isbell, 1996) as classification systems, which are covering a range of ecological regions (from the arctic to the [sub]tropics). In addition, some attention will be paid to the soil classification systems in use in Brazil, Canada, England and Wales, Germany, New Zealand, and South Africa, which are more oriented toward one ecological region.

32.2 The FAO–UNESCO Legend of the Soil Map of the World 1:5,000,000

32.2.1 Introduction

The *Soil Map of the World* is a response to recommendations made by the international soil community in the 1950s to give special attention to developing the classification and correlation of the soils of great regions of the world (FAO–UNESCO, 1974). It was the first attempt to prepare, on the basis of international cooperation, a soil map covering all continents of the world using a uniform legend, thus enabling the correlation of soil units and the comparison of soils on a global scale. It should be borne in mind that the *Soil Map of the World* project produced a legend accompanying the map and not a global soil classification system.

32.2.2 History

The Food and Agriculture Organization (FAO) and United Nations Educational, Scientific, and Cultural Organization (UNESCO), in association with the International Society of Soil Science (ISSS), jointly took up the recommendations made during the Sixth and Seventh ISSS Congresses in 1956 and 1960 to prepare a *Soil Map of the World* at scale 1:5,000,000. The project started in 1961 and was based on the compilation of available soil survey material and field correlation. A scientific advisory panel was convened to study the scientific and methodological problems relative to the preparation of such a soil map of the world.

During a number of meetings, the advisory panel worked out the organization of the field correlation, selected the scale of the map and its topographic base, and prepared the first draft definitions of soil units. These were presented to the Eighth ISSS Congress in 1964. In 1966, a general agreement was reached on the principles for constructing the international legend, on the preparation of the definitions of soil units, and on the adoption of a unified nomenclature. The first draft was presented in 1968 to the Ninth ISSS Congress, which approved the outline of the legend, the definitions, and the nomenclature. Moreover, it recommended that the *Soil Map of the World* be published as soon as possible.

32.2.3 Objectives

The objectives of the *Soil Map of the World* project (FAO–UNESCO, 1974) were to (1) make a first appraisal of the world's soil resources, (2) supply a scientific basis for the transfer of experience between areas with similar environments, (3) promote the establishment of a generally accepted soil classification and nomenclature, (4) establish a common framework for more detailed investigations in developing areas, (5) serve as a basic document for educational, research, and development activities, and (6) strengthen international contacts in the field of soil science.

32.2.4 The Soil Units

The soil units that form the basis of the *FAO–UNESCO Legend* have been defined in terms of measurable and observable properties of the soil itself. They form a monocategorical and not a taxonomic system with different levels of generalization. However, for the sake of logical presentation, they can be grouped on generally accepted principles of soil formation (FAO–UNESCO, 1974). Based on soil development status, material, and major geographical zone, 24 major soils and 106 soil units are distinguished (Table 32.1).

The soil units are characterized by the presence or absence of diagnostic horizons and properties. Key properties have been selected on the basis of generally accepted principles of soil formation so as to correlate with as many other characteristics as possible. Clusters of properties have been combined into so-called diagnostic horizons, which have been adopted to formulate the definitions of the soil units. The definitions and nomenclature of the diagnostic horizons and properties used are drawn from those adopted in *Soil Taxonomy* (Soil Survey Staff, 1975), but the definitions have been summarized and sometimes simplified to serve the purpose of the legend. For full background and details, the user is referred to *Soil Taxonomy* (Soil Survey Staff, 1975, 1996). A brief overview of the diagnostic horizons and properties used in the *FAO–UNESCO Legend* and their meaning are given in Table 32.2.

Volume I of the *Legend of the Soil Map of the World* provides a key to the soil units, which can be used to identify the soil (FAO–UNESCO, 1974). An abbreviated version listing the major soil units is reproduced in Table 32.3.

32.3 The Revised Legend of the FAO–UNESCO Soil Map of the World

32.3.1 Introduction

In 1988, a revised version of the *Legend of the Soil Map of the World* was issued (FAO–UNESCO, 1988), which assessed to what extent the objectives of the original *Legend of the Soil Map of the World* (FAO–UNESCO, 1974; Section 32.2.3) were met and analyzed its present day function. It was realized that, in order to keep the maps and accompanying legend up to date, revisions were necessary. New knowledge on soils, particularly from the developing world, had emerged and more recent soil surveys had yielded better insight into the distribution of soils in the world. The *Revised Legend* did not replace the *1974 Legend*, which is still the reference for the *Soil Map of the World*. The *Revised Legend* served also as a framework for the establishment by the ISSS of the WRB (Section 32.4).

32.3.2 Amendments to the 1974 Legend of the Soil Map of the World

The monocategorical character of the *1974 Legend*, using only soil units, was transformed into a multicategorical system with major soil groupings (MSGs; e.g., Fluvisols), soil units

TABLE 32.1 List of 106 Soil Units of the *FAO–UNESCO Legend of the Soil Map of the World 1:5,000,000*

Non or weakly developed soils

J	**Fluvisols**	G	**Gleysols**	R	**Regosols**	I	**Lithosols**		
Je	Eutric Fluvisols	Ge	Eutric Gleysols	Re	Eutric Regosols				
Jc	Calcaric Fluvisols	Gc	Calcaric Gleysols	RC	Calcaric Regosols				
Jd	Dystric Fluvisols	Gd	Dystric Gleysols	Rd	Dystric Regosols				
Jt	Thionic Fluvisols	Gm	Mollic Gleysols	Rx	Gelic Regosols				
		Gh	Humic Gleysols						
		Gp	Plinthic Gleysols						
		Gx	Gelic Gleysols						

Soils conditioned by parent material

Q	**Arenosols**	E	**Rendzinas**	U	**Rankers**	T	**Andosols**	V	**Vertisols**
Qc	Cambic Arenosols					To	Ochric Andosols	Vp	Pellic Vertisols
Ql	Luvic Arenosols					Tm	Mollic Andosols	Vc	Chromic Vertisols
Qf	Ferralic Arenosols					Th	Humic Andosols		
Qa	Albic Arenosols					Tv	Vitric Andosols		

Soils from (semi-)arid region

Z	**Solonchaks**	S	**Solonetz**	Y	**Yermosols**	X	**Xerosols**		
Zo	Orthic Solonchaks	So	Orthic Solonetz	Yh	Haplic Yermosols	Xh	Haplic Xerosols		
Zm	Mollic Solonchaks	Sm	Mollic Solonetz	Yk	Calcic Yermosols	Xk	Calcic Xerosols		
Zt	Takyric Solonchaks	Sg	Gleyic Solonetz	Yy	Gypsic Yermosols	Xy	Gypsic Xerosols		
Zg	Gleyic Solonchaks			Yl	Luvic Yermosols	Xl	Luvic Xerosols		
				Yt	Takyric Yermosols				

Soils from the steppe regions

K	**Kastanozems**	C	**Chernozems**	H	**Phaeozems**	M	**Greyzems**		
Kh	Haplic Kastanozems	Ch	Haplic Chernozems	Hh	Haplic Phaeozems	Mo	Orthic Greyzems		
Kk	Calcic Kastanozems	Ck	Calcic Chernozems	Hc	Calcaric Phaeozems	Mg	Gleyic Greyzems		
Kl	Luvic Kastanozems	Cl	Luvic Chernozems	Hl	Luvic Phaeozems				
		Cg	Glossic Chernozems	Hg	Gleyic Phaeozems				

Moderately developed soils mainly from temperate regions

B	**Cambisols**	L	**Luvisols**	D	**Podzoluvisols**	P	**Podzols**	W	**Planosols**
Be	Eutric Cambisols	Lo	Orthic Luvisols	De	Eutric Podzoluvisols	Po	Orthic Podzols	We	Eutric Planosols
Bd	Dystric Cambisols	Lc	Chromic Luvisols	Dd	Dystric Podzoluvisols	Pl	Leptic Podzols	Wd	Dystric Planosols
Bh	Humic Cambisols	Lk	Calcic Luvisols	Dg	Gleyic Podzoluvisols	Pf	Ferric Podzols	Wm	Mollic Planosols
Bg	Gleyic Cambisols	Lv	Vertic Luvisols			Ph	Humic Podzols	Wh	Humic Planosols
Bx	Gelic Cambisols	Lf	Ferric Luvisols			Pp	Placic Podzols	Ws	Solodic Planosols
Bk	Calcic Cambisols	La	Albic Luvisols			Pg	Gleyic Podzols	Wx	Gelic Planosols
Bc	Chromic Cambisols	Lp	Plinthic Luvisols						
Bv	Vertic Cambisols	Lg	Gleyic Luvisols						
Bf	Ferralic Cambisols								

Strongly weathered soils mainly from the tropical regions

A	**Acrisols**	N	**Nitosols**	F	**Ferralsols**		
Ao	Orthic Acrisols	Ne	Eutric Nitosols	Fo	Orthic Ferralsols		
Af	Ferric Acrisols	Nd	Dystric Nitosols	Fx	Xanthic Ferralsols		
Ah	Humic Acrisols	Nh	Humic Nitosols	Fr	Rhodic Ferralsols		
Ap	Plinthic Acrisols			Fh	Humic Ferralsols		
Ag	Gleyic Acrisols			Fa	Acric Ferralsols		
				Fp	Plinthic Ferralsols		

Organic soils

O	**Histosols**
Oe	Eutric Histosols
Od	Dystric Histosols
Ox	Gelic Histosols

Source: FAO–UNESCO. 1974. Soil map of the world 1:5,000,000. Volume I, Legend. United Nations Educational, Scientific and Cultural Organization, Paris, France.

TABLE 32.2 Diagnostic Horizons and Properties in the *FAO–UNESCO Legend of the Soil Map of the World 1:5,000,000*

Horizons

Albic E	Light colored eluvial horizon generally associated with argic and spodic horizons
Argillic B	Subsurface horizon with distinct clay accumulation
Calcic	Horizon with accumulation of calcium carbonate
Cambic B	Subsurface horizon showing evidence of alteration relative to the underlying horizon(s)
Gypsic	Horizon with accumulation of gypsum
Histic H	Poorly aerated, waterlogged, highly organic surface horizon
Mollic A	Thick, dark-colored surface horizon with high base saturation and moderate to high organic matter content
Natric B	Subsurface horizon with distinct clay accumulation and a high exchangeable sodium percentage
Ochric A	Weakly developed surface horizon, either light colored, or thin, or having a low organic matter content
Oxic B	Strongly weathered subsurface horizon with low cation exchange capacity (CEC)
Spodic B	Dark-colored subsurface horizon with illuvial aluminoorganic complexes, with or without iron
Sulfuric	Extremely acid subsurface horizon with sulfuric acid resulting from oxidation of sulfides
Umbric A	Thick, dark-colored surface horizon with low base saturation and moderate to high organic matter content

Properties and materials

Abrupt textural change	Sharp increase in clay content within a limited depth range
Albic material	Light colored mineral soil material
Aridic moisture regime	No available water in any part of the moisture control section for as long as 90 consecutive days or more than half the time when soil temperature (at 50 cm depth) is above 5°C
Exchange complex dominated by amorphous material	CEC (pH 8.2) more than 150 $cmol_c \cdot kg^{-1}$ clay; if 15 bar water content is 20% or more, pH of NaF is more than 9.4; ratio of 15 bar water content to clay is more than 1.0; more than 0.6% organic carbon; DTA shows low-temperature endotherm; bulk density is 0.85 $g\ cm^{-3}$ at 1/3 bar tension, low CEC (<24 $cmol_c\ kg^{-1}$ clay by NH_4Cl)
Ferralic ferric	Iron concentrated in large mottles or concretions, or low CEC (<24 $cmol_c\ kg^{-1}$ clay by NH_4Cl)
Gilgai microrelief	Succession of enclosed microbasins and microknolls in level areas or microvalleys and microridges on slopes
High organic matter content	Organic matter content of 1.35% or more averaged to a depth of 100 cm or 1.5% organic matter in the upper part of the B horizon (acrisols only)
High salinity	EC of saturation extract more than 15 dS m^{-1} at 25°C at specified depths or more than 4 dS m^{-1} within 25 cm of the surface if pH (H_2O, 1:1) exceeds 8.5
Hydromorphic properties	Saturation with groundwater; occurrence of a histic H horizon; dominant neutral (N) hues or hues bluer than 10Y; saturation with water at some period of the year (unless artificially drained) with evidence of reduction or of reduction and segregation
Interfingering	Penetrations of an albic E horizon into an underlying argillic or natric B horizon, not wide enough to constitute tonguing
Permafrost	Perennial temperature at or below 0°C
Plinthite	Iron-rich, humus-poor soil material, which hardens irreversibly upon repeated wetting and drying
Slickensides	Polished and grooved ped surfaces that are produced by sliding past another
Smeary consistence	Presence of thixotropic soil material
Soft powdery lime	Accumulation of translocated calcium carbonate in soft powdery form
Sulfidic materials	Waterlogged deposit containing 0.75% or more sulfur and less than three times as much carbonates ($CaCO_3$ equivalent) as sulfur
Takyric features	Combination of heavy texture, polygonal cracks, and platy or massive surface crust
Thin iron pan	Black to dark reddish layer cemented by iron, by iron and manganese, or by iron–organic matter complexes
Tonguing	Penetrations of an albic E horizon into an argillic B horizon with specified dimensions
Vertic properties	Cracks 1 cm or more wide within 50 cm of the upper boundary of the B horizon, extending to the surface or at least to the upper part of the B horizon
Weatherable minerals	Presence of minerals considered to be unstable relative to other minerals such as quartz and 1:1 lattice clays and which produce plant nutrients upon weathering

Source: FAO–UNESCO. 1974. Soil map of the world 1:5,000,000. Volume I, Legend. United Nations Educational, Scientific and Cultural Organization, Paris, France.

(e.g., Dystric Fluvisols), and soil subunits (e.g., Gleyi-dystric Fluvisols). The introduction of a multicategorical system was necessitated by the increasing use of the *1974 Legend* in more detailed surveys, particularly in Africa, where during the 1980s 1:1,000,000 soil maps were produced of, for example, Kenya, Botswana, and Zambia.

The *Revised Legend of the Soil Map of the World* distinguishes 28 MSGs, 4 more than the *1974 Legend*, and 153 soil units, 47 more than in 1974. Major changes are (1) amalgamation of the Lithosols, Rendzinas, and Rankers into one MSG (Leptosols), since the three had been difficult to show on the map; (2) deletion of the Xerosols and Yermosols, which were characterized by an

TABLE 32.3 Key to Major Soil Units of the 1974 *FAO–UNESCO Legend*

Soils having an H horizon of 40 cm or more (60 cm or more if the organic material consists mainly of sphagnum or moss or has a bulk density of less than 0.1 g cm⁻³) either extending down from the surface or taken cumulatively within the upper 80 cm of the soil; the thickness of the H horizon maybe less when it rests on rock or on fragmental material of which the interstices are filled with organic matter

<div align="right">

Histosols (O)

</div>

Other soils that are limited in depth by continuous coherent and hard rock within 10 cm of the surface

<div align="right">

Lithosols (I)

</div>

Other soils that, after the upper 20 cm have been mixed, have 30% or more clay in all horizons to at least 50 cm from the surface; at some period in most years have cracks at least 1 cm wide at a depth of 50 cm, unless irrigated, and have one or more of the following characteristics: gilgai microrelief, intersecting slickensides, or wedge-shaped or parallelepiped structural aggregates at some depth between 25 and 100 cm from the surface

<div align="right">

Vertisols (V)

</div>

Other soils developed from recent alluvial deposits, having no diagnostic horizons other than (unless buried by 50 cm or more new material) an ochric or an umbric A horizon, a histic H horizon, or a sulfuric horizon

<div align="right">

Fluvisols (J)

</div>

Other soils having high salinity and having no diagnostic horizons (unless buried by 50 cm or more new material) an A horizon, an H horizon, a cambic B horizon, a calcic or a gypsic horizon

<div align="right">

Solonchaks (Z)

</div>

Other soils showing hydromorphic properties within 50 cm of the surface; having no diagnostic horizons other than (unless buried by 50 cm or more new material) an A horizon, an H horizon, a cambic B horizon, a calcic or gypsic horizon

<div align="right">

Gleysols (G)

</div>

Other soils having either a mollic or an umbric A horizon possibly overlying a cambic B horizon or an ochric A horizon and a cambic B horizon; having no other diagnostic horizons (unless buried by 50 cm or more new material); having to a depth of 35 cm or more one or both of (a) a bulk density (at 1/3 bar water retention) of the fine earth (less than 2 mm) fraction of the soil of less than 0.85 g cm⁻³ and an exchange complex dominated by amorphous material; (b) 60% or more vitric volcanic ash, cinders, or other vitric pyroclastic material in the silt, sand, and gravel fractions

<div align="right">

Andosols (T)

</div>

Other soils of coarse texture consisting of albic material occurring over a depth of at least 50 cm from the surface or showing characteristics of argillic, cambic, or oxic B horizons which, however, do not qualify as diagnostic horizons because of textural requirements; having no diagnostic horizons other than (unless buried by 50 cm or more new material) an ochric A horizon

<div align="right">

Arenosols (Q)

</div>

Other soils having no diagnostic horizons or none other than (unless buried by 50 cm or more new material) an ochric A horizon

<div align="right">

Regosols (R)

</div>

Other soils having an umbric A horizon, which is not more than 25 cm thick; having no other diagnostic horizons (unless buried by 50 cm or more new material)

<div align="right">

Rankers (U)

</div>

TABLE 32.3 (continued) Key to Major Soil Units of the 1974 *FAO–UNESCO Legend*

Other soils having a mollic A horizon that contains or immediately overlies calcareous material with a calcium carbonate equivalent of more than 40% (when the A horizon contains a high amount of finely divided calcium carbonate the color requirements of the mollic A horizon may be waived)

<div align="right">

Rendzinas (E)

</div>

Other soils having a spodic B horizon

<div align="right">

Podzols (P)

</div>

Other soils having an oxic B horizon

<div align="right">

Ferralsols (F)

</div>

Other soils having an albic E horizon overlying a slowly permeable horizon (e.g., an argillic or natric B horizon showing an abrupt textural change, a heavy clay, a fragipan) within 125 cm of the surface; showing hydromorphic properties at least in a part of the E horizon

<div align="right">

Planosols (W)

</div>

Other soils having a natric B horizon

<div align="right">

Solonetz (S)

</div>

Other soils having a mollic A horizon with a moist chroma of 2 or less to a depth of at least 15 cm, showing bleached coatings on structural ped surfaces

<div align="right">

Greyzems (M)

</div>

Other soils having a mollic A horizon with a moist chroma of 2 or less to a depth of at least 15 cm, having one or more of the following: a calcic or a gypsic horizon or concentrations of soft powdery lime within 125 cm of the surface, when the weighted average textural class is coarse, within 90 cm for medium textures, within 75 cm for fine textures

<div align="right">

Chernozems (C)

</div>

Other soils having a mollic A horizon with a moist chroma of more than 2 to a depth of at least 15 cm, having one or more of the following: a calcic or a gypsic horizon or concentrations of soft powdery lime within 125 cm of the surface, when the weighted average textural class is coarse, within 90 cm for medium textures, within 75 cm for fine textures

<div align="right">

Kastanozems (K)

</div>

Other soils having a mollic A horizon

<div align="right">

Phaeozems (H)

</div>

Other soils having an argillic B horizon showing an irregular or broken upper boundary resulting from deep tonguing of the E into the B horizon or from the formation of discrete nodules (ranging from 2 to 5 cm up to 30 cm in diameter) the exteriors of which are enriched and weakly cemented or indurated with iron and having redder hues and stronger chroma than the interiors

<div align="right">

Podzoluvisols (D)

</div>

Other soils having a weak ochric A horizon and an aridic moisture regime; lacking permafrost within 200 cm of the surface

<div align="right">

Xerosols (X)

</div>

Other soils having a very weak ochric A horizon and an aridic moisture regime; lacking permafrost within 200 cm of the surface

<div align="right">

Yermosols (Y)

(continued)

</div>

TABLE 32.3 (continued)   Key to Major Soil Units of the 1974
FAO–UNESCO Legend

Other soils having an argillic B horizon with a clay distribution, where the
  percentage of clay does not decrease from its maximum amount by as much
  as 20% within 150 cm of the surface; lacking plinthite within 125 cm of the
  surface; lacking vertic and ferric properties

Nitosols (N)

Other soils having an argillic B horizon; having a base saturation, which is
  less than 50% (by NH_4OAc) in at least some part of the B horizon within
  125 cm of the surface

Acrisols (A)

Other soils having an argillic B horizon

Luvisols (L)

Other soils having a cambic B horizon or an umbric A horizon, which is
  more than 25 cm thick

Cambisols (B)

Source: FAO–UNESCO. 1974. Soil map of the world 1:5,000,000. Volume
I, Legend. United Nations Educational, Scientific and Cultural Organization,
Paris, France.

aridic moisture regime (as a general principle, climatic criteria
were not to be used in separating the soil units); (3) division of
the Acrisols and Luvisols, in 1974 distinguished by base satura-
tion, into four MSGs, introducing clay activity as an additional
separating criterion; (4) introduction of new MSGs (Alisols,
Calcisols, Gypsisols, Lixisols, Plinthosols, and Anthrosols); and
(5) renaming of Nitosols (Nitisols). A listing of the MSGs and the
soil units is given in Table 32.4.

The diagnostic horizons and properties have also been
largely adapted. In 1974, they were fairly similar to those in
Soil Taxonomy (Soil Survey Staff, 1975), but in 1988, many were
redefined and renamed and new additions made. The argillic
and oxic B horizons were renamed argic and ferralic B horizons,
respectively. The argic B horizon now includes both the argillic
and the kandic horizon as defined in *Soil Taxonomy* (Soil Survey
Staff, 1996), while the ferralic B horizon includes additional cri-
teria such as silt/clay ratio and water-dispersible clay content. An
addition is the thick, manmade, fimic A horizon, which includes
both the anthropic and plaggen epipedons of *Soil Taxonomy*.

Newly defined diagnostic properties include continuous hard
rock, fluvic, geric, nitic, and sodic properties. The 1974 hydro-
morphic properties were split into gleyic (wetness conditioned
by groundwater) and stagnic (wetness conditioned by surface
water) properties. Definitions of albic material, aridic moisture
regime, high organic matter content, and thin iron pan were
deleted as they are no longer used in defining soil units. Andic
properties have replaced the 1974 exchange complex dominated
by amorphous material.

A brief overview of the diagnostic horizons and proper-
ties used in the *Revised Legend* and their meaning is given in
Table 32.5. A key to the MSGs in the *Revised Legend* is given
in Table 32.6.

32.4 The World Reference Base for Soil Resources

32.4.1 Introduction

The first official version of the *World Reference Base for Soil
Resources* (WRB) was released at the 16th World Congress of
Soil Science at Montpellier in 1998. At the same event, it was
also endorsed and adopted as the system for soil correlation and
international communication of the International Union of Soil
Sciences (IUSS). It is also serving as a correlation scheme for soil
maps and databases for the European Union. Although it was not
designed for, in several countries it is also applied for different
scale of soil mapping. The current official version was released
at the 18th World Congress of Soil Science in Philadelphia in
2006, and an updated digital version was published in 2007 on
the official, FAO-hosted Web site of the WRB. The main text was
translated into 14 languages (Arabic, Chinese, French, German,
Hungarian, Italian, Japanese, Latvian, Lithuanian, Polish,
Rumanian, Russian, Spanish, and Vietnamese) and most trans-
lations are also available online on the Web site http://www.fao.
org/nr/land/soils/soil/wrb-documents/jp/

32.4.2 History

The WRB was initiated by the ISSS, FAO, and ISRIC to "provide
scientific depth and background to the 1988 FAO–UNESCO–
ISRIC *Revised Legend of the Soil Map of the World*, so that it incor-
porates the latest knowledge relating to global soil resources and
interrelationships" (ISSS–ISRIC–FAO, 1994). The initiative was
closely related to the *Soil Map of the World 1:5,000,000* project
of FAO and UNESCO. After its completion in the early 1980s,
it was realized that 20 years had elapsed since the project had
started. During this period, numerous new soil surveys, often
using local or national soil classification systems, had been car-
ried out both in developing and developed countries, generating
new knowledge and insight on the distribution and potential of
our soil resources. If the *Soil Map of the World* were to retain its
value as a global soil resource inventory, it needed regular updat-
ing with the most recent information.

Preliminary discussions were started on the establishment
of an International Reference Base for Soil Classification (IRB),
an initiative undertaken by FAO and UNESCO, supported by
the United Nations Environmental Program (UNEP) and the
ISSS. The intention of the IRB project was to work toward the
establishment of a framework through which existing soil clas-
sification systems could be correlated and ongoing soil classi-
fication work could be harmonized (Dudal, 1990). Meetings
were organized in Sofia, Bulgaria, to commence such an interna-
tional program. The outcomes were draft definitions of 16 MSGs
occurring globally.

During the ISSS congresses in 1982 and 1986 and expert con-
sultations in 1987 and 1988, the IRB took form, and as a result,
20 MSGs were identified and agreed upon as being representa-
tive of the principal components of the world's soil cover.

TABLE 32.4   MSGs and Soil Units, 1988 *Revised Legend of the Soil Map of the World*

AC, Acrisols

ACh, haplic Acrisols
ACf, ferric Acrisols
ACu, humic Acrisols
ACp, plinthic Acrisols
ACg, gleyic Acrisols

AL, Alisols

ALh, haplic Alisols
ALf, ferric Alisols
ALu, humic Alisols
ALp, plinthic Alisols
ALj, stagnic Alisols
ALg, gleyic Alisols

AR, Arenosols

ARh, haplic Arenosols
ARb, cambic Arenosols
ARl, luvic Arenosols
ARo, ferralic Arenosols
ARa, albic Arenosols
ARc, calcaric Arenosols
ARg, gleyic Arenosols

AN, Andosols

ANh, haplic Andosols
ANm, mollic Andosols
ANu, umbric Andosols
ANz, vitric Andosols
ANg, gleyic Andosols
ANl, gelic Andosols

AT, Anthrosols

ATa, aric Anthrosols
ATc, cumulic Anthrosols
ATf, fimic Anthrosols
ATu, urbic Anthrosols

CL, Calcisols

CLh, haplic Calcisols
CLc, luvic Calcisols
CLp, petric Calcisols

CM, Cambisols

CMe, eutric Cambisols
CMd, dystric Cambisols
CMu, humic Cambisols
CMc, calcaric Cambisols
CMv, vertic Cambisols
CMo, ferralic Cambisols
CMg, gleyic Cambisols
CMi, gelic Cambisols

CH, Chernozems

CHh, haplic Chernozems
CHk, calcic Chernozems
CHl, luvic Chernozems
CHw, glossic Chernozems
CHg, gleyic Chernozems

FL, Fluvisols

FLe, eutric Fluvisols
FLc, calcaric Fluvisols
FLd, dystric Fluvisols
FLm, mollic Fluvisols
FLu, umbric Fluvisols
FLt, thionic Fluvisols
FLs, salic Fluvisols

FR, Ferralsols

FRh, haplic Ferralsols
FRx, xanthic Ferralsols
FRr, rhodic Ferralsols
FRu, humic Ferralsols
FRg, geric Ferralsols
FRp, plinthic Ferralsols

GL, Gleysols

GLe, eutric Gleysols
GLk, calcic Gleysols
GLd, dystric Gleysols
GLa, andic Gleysols
GLm, mollic Gleysols
GLu, umbric Gleysols
GLt, thionic Gleysols
GLi, gelic Gleysols

GR, Grgreyzems

GRh, haplic Greyzems
GRg, gleyic Greyzems

GY, Gypsisols

GYh, haplic Gypsisols
GYk, calcic Gypsisols
GYl, luvic Gypsisols
GYp, petric Gypsisols

HS, Histosols

HSl, folic Histosols
HSs, terric Histosols
HSf, fibric Histosols
HSt, thionic Histosols
HSi, gelic Histosols

KS, Kastanozems

KSh, haplic Kastanozems
KSl, luvic Kastanozems
KSk, calcic Kastanozems
KSy, gypsic Kastanozems

LP, Leptosols

LPe, eutric Leptosols
LPd, dystric Leptosols
LPk, rendzic Leptosols
LPm, mollic Leptosols
LPu, umbric Leptosols
LPq, lithic Leptosols
LPi, gelic Leptosols

LX, Lixisols

LXh, haplic Lixisols
LXr, ferric Lixisols
LXp, plinthic Lixisols
LXa, albic Lixisols
LXj, stagnic Lixisols
LXg, gleyic Lixisols

LV, Luvisols

LVh, haplic Luvisols
LVf, ferric Luvisols
LVx, chromic Luvisols
LVk, calcic Luvisols
LVv, vertic Luvisols
LVa, albic Luvisols
LVj, stagnic Luvisols
LVg, gleyic Luvisols

NT, Nitisols

NTh, haplic Nitisols
ATr, rhodic Nitisols
NTu, humic Nitisols

PH, Phaeozems

PHh, haplic Phaeozems
PHc, calcaric Phaeozems
PHl, luvic Phaeozems
PHj, stagnic Phaeozems
PHg, gleyic Phaeozems

PL, Planosols

PLe, eutric Planosols
PLd, dystric Planosols
PLm, mollic Planosols
PLu, umbric Planosols
PLi, gelic Planosols

(continued)

TABLE 32.4 (continued)   MSGs and Soil Units, 1988 *Revised Legend of the Soil Map of the World*

PT, **Plinthosols**	**PZc**, carbic Podzols	**SCn**, sodic Solonchaks
PTe, eutric Plinthosols	**PZg**, gleyic Podzols	**SCg**, gleyic Solonchaks
PTd, dystric Plinthosols	**PZi**, gelic Podzols	**SCi**, gelic Solonchaks
PTu, humic Plinthosols		
PTa, albic Plinthosols	**RG**, **Regosols**	**SN**, **Solonetz**
	RGe, eutric Regosols	**SNh**, haplic Solonetz
PD, **Podzoluvisols**	**RGc**, calcaric Regosols	**SNm**, mollic Solonetz
PDe, eutric Podzoluvisols	**RGy**, gypsiric Regosols	**SNk**, calcic Solonetz
PLd, dystric Podzoluvisols	**RGd**, dystric Regosols	**SNy**, gypsic Solonetz
PLj, stagnic Podzoluvisols	**RGu**, umbric Regosols	**SNj**, stagnic Solonetz
PDg, gleyic Podzoluvisols	**RGi**, gelic Regosols	**SNg**, gleyic Solonetz
PDi, gelic Podzoluvisols		
	SC, **Solonchaks**	**VR**, **Vertisols**
PZ, **Podzols**	**SCh**, haplic Solonchaks	**VRe**, eutric Vertisols
PZh, haplic Podzols	**SCm**, mollic Solonchak	**VRd**, dystric Vertisols
PZb, cambic Podzols	**SCk**, calcic Solonchaks	**Vrk**, calcic Vertisols
PZf, ferric Podzols	**SCy**, gypsic Solonchaks	**VRy**, gypsic Vertisols

Source: FAO–UNESCO. 1988. FAO–UNESCO soil map of the world, revised legend. World Soil Resources Report No. 60. Food and Agriculture Organization, Rome, Italy.

Subsequently, it became clear that some of the proposed 20 MSGs were too broad to be defined consistently and, consequently, had to be subdivided. By doing so, the list of MSGs became very close to those of the *Revised Legend of the Soil Map of the World* (FAO–UNESCO, 1988). As a result, it was decided in 1992 to adopt the revised legend as the frame for further IRB work. This was also prompted by the fact that both the *Revised Legend* and the IRB were supported by the ISSS and that it would be inappropriate to pursue two programs, which essentially had the same goal, namely, to arrive at a rational inventory of global soil resources (ISSS–ISRIC–FAO, 1994, FAO–ISRIC–ISSS, 1998). The two programs were, therefore, merged under the name WRB. After 8 years of data collection and intensive worldwide testing of the first edition (1998), the second edition (2006 and 2007) was presented after broad discussions. The text of the chapters to follow will be based mostly on the current version (IUSS Working Group WRB, 2006) indicating significant changes from the earlier edition.

32.4.3 Objectives

The specific objectives of the WRB are to (1) develop an internationally acceptable framework for delineating soil resources to which national classifications can be attached and related, using the *FAO Revised Legend* as a guideline; (2) provide this framework with a sound scientific base, so that it can also serve different applications in related fields, such as agriculture, geology, hydrology, and ecology; (3) acknowledge in the framework important lateral aspects of soils and soil horizon distribution as characterized by topo- and chronosequences; and (4) emphasize the morphological characterization of soils rather than to follow a purely analytical approach.

32.4.4 Concepts and Principles

The general principles on which the WRB is based were laid down during the early Sofia meetings in 1980 and 1981 and further elaborated upon by the working groups entrusted with its development. These general principles can be summarized as follows: (1) The classification of soils is based on soil properties defined in terms of diagnostic horizons, properties, and materials, which to the greatest extent possible should be measurable and observable in the field; (2) selection of diagnostic horizons, properties, and materials takes into account their relationship with soil-forming processes; however, at a high level of generalization, it also attempts to select, to the extent possible, diagnostic features, which are of significance for management purposes; (3) climatic parameters are not applied in the classification of soils. It is fully realized that they should be used for interpretation purposes, in dynamic combination with soil properties, but they should not form part of soil definitions; (4) the WRB is a comprehensive classification system that enables people to accommodate their national classification system. It comprises two tiers of categorical detail: *the reference base*, limited to the first level only and having 32 reference soil groups (RSGs) and the *WRB classification system*, consisting of combinations of a set of prefix and suffix qualifiers that are uniquely defined and added to the name of the RSG, allowing very precise characterization and classification of individual soil profiles. The first edition of the *WRB*, published in 1998, comprised 30 RSGs; the second edition, published in 2006, has 32 RSGs. Definitions and descriptions of lower level soil units reflect variations in soil characteristics both vertically and laterally so as to account for spatial linkages within the landscape. The term "Reference Base" is connotative

TABLE 32.5 Diagnostic Horizons and Properties, 1988 *Revised Legend of the Soil Map of the World*

Horizons

Albic E	Light colored eluvial horizon generally associated with argic and spodic horizons
Argic B	Subsurface horizon with distinct clay accumulation
Calcic	Horizon with accumulation of calcium carbonate
Cambic B	Subsurface horizon showing evidence of alteration relative to the underlying horizon(s)
Ferralic B	Strongly weathered subsurface horizon with low CEC
Fimic A	Surface and subsurface horizons resulting from long-continued cultivation
Gypsic	Horizon with accumulation of gypsum
Histic H	Poorly aerated, waterlogged, highly organic surface horizon
Mollic A	Thick, dark-colored surface horizon with high base saturation and moderate to high organic matter content
Natric B	Subsurface horizon with distinct clay accumulation and a high exchangeable sodium percentage
Ochric A	Weakly developed surface horizon, either light colored, or thin, or having a low organic matter content
Petrocalcic	Continuous cemented or indurated calcic horizon
Petrogypsic	Continuous cemented or indurated gypsic horizon
Spodic B	Dark-colored subsurface horizon with illuvial aluminoorganic complexes, with or without iron
Sulfuric	Extremely acid subsurface horizon with sulfuric acid resulting from oxidation of sulfides
Umbric A	Thick, dark-colored surface horizon with low base saturation and moderate to high organic matter content

Properties

Abrupt textural change	Sharp increase in clay content within a limited depth range
Andic	High acid oxalate extractable Al and Fe content, low bulk density, high phosphate retention, high amount of coarse volcanoclastic material
Calcareous	Strong effervescence with 10% HCl (more than 2% calcium carbonate)
Calcaric	Presence of calcareous soil material between 20 and 50 cm depth
Continuous hard rock	Presence of coherent rock, practically impermeable for roots
Ferralic	Low CEC ($<24\,cmol_c\,kg^{-1}$ clay or $4\,cmol_c\,kg^{-1}$ fine earth)
Ferric	Presence of many coarse mottles or large iron concretions
Fluvic	Presence of fresh fluviatile, lacustrine, or marine sediments at the surface
Geric	Extremely low to negative effective CEC
Gleyic	Wetness producing reduced conditions caused by groundwater
Gypsiferous	Presence ≥5% gypsum
Interfingering	Narrow penetrations of an albic E horizon into an argic or natric B horizon
Nitic	Presence of strongly developed, nut-shaped structure and shiny pedfaces
Organic material	Material containing a very high amount of organic debris
Permafrost	Perennial temperature at or below 0°C
Plinthite	Presence of iron-rich, humus-poor material, which hardens irreversibly upon repeated wetting and drying
Salic	High soluble salt content (electrical conductivity $>15\,dS\,m^{-1}$ or >4 if pH exceeds 8.5)
Slickensides	Presence of polished and grooved surfaces produced by one mass sliding past another
Smeary consistence	Presence of thixotropic soil material
Sodic	ESP ≥ 15% or exchangeable Na + Mg percentage ≥50%
Soft powdery lime	Accumulation of translocated calcium carbonate in soft powdery form
Stagnic	Wetness producing reduced conditions caused by stagnating surface water
Strongly humic	High organic matter content
Sulfidic material	Waterlogged deposit containing sulfides and only moderate amounts of calcium carbonate
Tonguing	Wide penetrations of an albic E horizon into an argic B horizon or penetrations of a mollic A horizon into an underlying cambic B horizon or into a C horizon (Chernozems only)
Vertic	Presence of cracks, slickensides, wedge-shaped, or parallelepiped structural aggregates
Weatherable minerals	Presence of minerals unstable in a humid climate relative to other minerals

Source: FAO–UNESCO. 1988. FAO–UNESCO soil map of the world, revised legend. World Soil Resources Report No. 60. Food and Agriculture Organization, Rome, Italy.

TABLE 32.6 Key to MSGs of the 1988 *Revised Legend of the Soil Map of the World*

Soils having an H horizon, or an O horizon, of 40 cm or more (60 cm or more if the organic material consists mainly of sphagnum or moss or has a bulk density of less than 0.1 Mg m^{-3}) either extending down from the surface or taken cumulatively within the upper 80 cm of the soil; the thickness of the H or O horizon may be less when it rests on rocks or on fragmental material of which the interstices are filled with organic matter

Histosols (HS)

Other soils that are coarser than sandy loam to a depth of at least 100 cm from the surface, having <35% of rock fragments or other coarse fragments in all subhorizons within 100 cm of the surface, having no diagnostic horizons other than an ochric A horizon or an albic E horizon

Arenosols (AR)

Other soils in which human activities have resulted in a profound modification or burial of the original soil horizons, through removal or disturbance of surface horizons, cuts and fills, secular additions of organic materials, long-continued irrigation, etc.

Anthrosols (AT)

Other soils having no diagnostic horizons other than an ochric or umbric A horizon; lacking soft powdery lime

Regosols (RG)

Other soils that are limited in depth by continuous hard rock or highly calcareous materials (calcium carbonate equivalent >40%) or a continuous cemented layer within 30 cm of the surface or having <20% of fine earth over a depth of 75 cm from the surface. Diagnostic horizons may be present

Leptosols (LP)

Other soils having a spodic B horizon

Podzols (PZ)

Other soils having, after the upper 18 cm have been mixed. Thirty percent or more clay in all horizons to a depth of 50 cm; developing cracks from the soil surface downward which at some period in most years (unless the soil is irrigated) are at least 1 cm wide to a depth of 50 cm; having one or more of the following: Intersecting slickensides or wedge-shaped or parallelepiped structural aggregates at some depth between 25 and 100 cm from the surface

Vertisols (VR)

Other soils having ≥25% plinthite by volume in a horizon, which is at least 15 cm thick within 50 cm of the surface or within a depth of 125 cm when underlying an albic E horizon or a horizon that shows stagnic properties within 50 cm of the surface or gleyic properties within 100 cm of the surface

Plinthosols (PT)

Other soils showing fluvic properties and having no diagnostic horizons other than an ochric, mollic, an umbric A horizon, or a histic H horizon, or a sulfuric horizon, or sulfidic material within 125 cm of the surface

Fluvisols (FL)

Other soils having a ferralic B horizon

Ferralsols (FR)

Other soils showing salic properties and having no diagnostic horizons other than an ochric, umbric, or mollic A horizon, a histic H horizon, a cambic B horizon, a calcic, or a gypsic horizon

Solonchaks (SC)

TABLE 32.6 (continued) Key to MSGs of the 1988 *Revised Legend of the Soil Map of the World*

Other soils having an E horizon showing stagnic properties at least in part of the horizon and abruptly overlying a slowly permeable horizon within 125 cm of the surface and lacking a natric or a spodic B horizon

Planosols (PL)

Other soils, exclusive of coarse textured materials (except when a histic H horizon is present), showing gleyic properties within 50 cm of the surface; having no diagnostic horizons other than an A horizon, a histic H horizon, a cambic B horizon, a sulfuric horizon, a calcic, or a gypsic horizon; lacking plinthite within 125 cm of the surface

Gleysols (GL)

Other soils having a natric B horizon

Solonetz (SN)

Other soils showing andic properties to a depth of 35 cm or more from the surface and having a mollic or an umbric A horizon possibly overlying a cambic B horizon or an ochric A horizon and a cambic B horizon; having no other diagnostic horizons

Andosols (AN)

Other soils having a mollic A horizon with a moist chroma of 2 or less to a depth of at least 15 cm, showing uncoated silt and sand grains on structural pedfaces; having an argic B horizon

Greyzems (GR)

Other soils having an argic B horizon showing an irregular or broken upper boundary resulting from deep tonguing of the A into the B horizon or from the formation of discrete nodules larger than 2 cm, the exteriors of which are enriched and weakly cemented or indurated and have redder hues and stronger chromas than the interiors

Podzoluvisols (PD)

Other soils having a mollic A horizon with a moist chroma of 2 or less to a depth of at least 15 cm; having a calcic or petrocalcic horizon, or concentrations of soft powdery lime within 125 cm of the surface, or both

Chernozems (CH)

Other soils having a gypsic or petrogypsic horizon within 125 cm of the surface; having no diagnostic horizons other than an ochric A horizon, a cambic B horizon or an argic B horizon permeated with gypsum or calcium carbonate, a calcic or petrocalcic horizon

Gypsisols (GY)

Other soils having a mollic A horizon with a moist chroma of more than 2 to a depth of at least 15 cm; having one or more of the following: a calcic, petrocalcic, or gypsic horizon or concentrations of soft powdery lime within 125 cm of the surface

Kastanozems (KS)

Other soils having a calcic or a petrocalcic horizon, or a concentration of soft powdery lime, within 125 cm of the surface; having no diagnostic horizons other than an ochric A horizon, a cambic B horizon, or an argic B horizon, which is calcareous

Calcisols (CL)

Other soils having a mollic A horizon; having a base saturation (by NH$_4$OAc) of ≥50% throughout the upper 125 cm of the soil

Phaeozems (PH)

TABLE 32.6 (continued) Key to MSGs of the 1988 *Revised Legend of the Soil Map of the World*

Other soils having an argic B horizon with a clay distribution, which does not show a relative decrease from its maximum of more than 20% within 150 cm of the surface; showing gradual to diffuse horizon boundaries between the A and B horizons; having nitic properties in some subhorizon within 125 cm of the surface

Nitisols (NT)

Other soils having an argic B horizon, which has a CEC equal to or more than 24 cmol$_c$ kg^{-1} clay and a base saturation (by NH$_4$OAc) of less than 50% in at least some part of the B horizon within 125 cm of the surface

Alisols (AL)

Other soils having an argic B horizon, which has a CEC of less than 24 cmol$_c$ kg^{-1} clay and a base saturation (by NH$_4$OAc) of less than 50% in at least some part of the B horizon within 125 cm of the surface

Acrisols (AC)

Other soils having an argic B horizon, which has a CEC equal to or more than 24 cmol$_c$ kg^{-1} clay and a base saturation (by NH$_4$OAc) of 50% or more throughout the B horizon within 125 cm of the surface

Luvisols (LV)

Other soils having an argic B horizon, which has a CEC of less than 24 cmol$_c$ kg^{-1} clay and a base saturation (by NH$_4$OAc) of 50% or more throughout the B horizon within 125 cm of the surface

Lixisols (LX)

Other soils having a cambic B horizon

Cambisols (CM)

Source: FAO–UNESCO. 1988. FAO–UNESCO soil map of the world, revised legend. World Soil Resources Report No. 60. Food and Agriculture Organization, Rome, Italy.

of the common denominator function that the WRB assumes. Its units have sufficient width to stimulate harmonization and correlation of existing national systems; (5) the reference base is not meant to substitute for national soil classification systems but rather to serve as a common denominator for communication at an international level; (6) in addition to serving as a link between existing classification systems, the WRB also serves as a consistent communication tool for compiling global soil databases and for the inventory and monitoring of the world's soil resources. The nomenclature used to distinguish soil groups retains terms that have been used traditionally or that can be introduced easily in current language.

32.4.5 The Architecture of the WRB

Currently, the WRB comprises two tiers of categorical detail: (1) the RSGs and (2) the combination of RSGs with qualifiers, detailing the properties of the RSGs by adding a set of uniquely defined qualifiers. From the *FAO Revised Legend*, through the 1998 WRB edition, several changes occurred. In 1998, one MSG has been omitted (Greyzems), and three new ones are introduced (Cryosols, Durisols, and Umbrisols). Greyzems

were deleted as they constitute the smallest MSG and were amalgamated with the Phaeozems. Cryosols were newly introduced to identify a group of soils, which occur under the unique environmental conditions of thawing and freezing. Durisols have been added to group soils together, which have accumulation of secondary silica, analogous to the Calcisols and Gypsisols. Umbrisols constitute the group of soils that have a thick accumulation of desaturated organic matter at the surface and are the natural counterpart of Chernozems, Kastanozems, and Phaeozems.

The second edition of the *WRB* has undergone a major revision. Technosols and Stagnosols have been introduced, leading to 32 RSGs instead of 30. The Technosols are soils with a certain amount of artifacts, a constructed geomembrane or technic hard rock. The Stagnosols unify the former Epistagnic subunits of many other RSGs. Some rearrangement has taken place in the order of the key, with Anthrosols, Solonetz, Nitisols, and Arenosols moving upward. The definitions of many diagnostic soil horizons, soil properties, and materials have been adjusted. An overview of the diagnostic categories is presented in Table 32.7. The number of qualifiers almost doubled (currently 179), and a significant change was the subdivision of qualifiers into prefix and suffix ones. Prefix qualifiers comprise those that are typically associated with the RSG (in order of their importance) and the intergrades to other RSGs (in order of the key). All other qualifiers are listed as suffix qualifiers.

The RSGs are defined by the key. For each RSG number of possible prefix and suffix qualifiers are listed in priority order. After determining the RSG, the applying prefix qualifier names are put before the RSG name and applying suffix qualifier names are placed between brackets following the RSG name.

The current scheme of the WRB proved to be capable of indicating most of the soil's properties and performed properly for correlation purposes. However, recent applications for mapping purposes indicated that when generalization is required, important information may not show with the current set of the qualifiers. Although WRB was not primarily designed to serve mapping purposes, it is increasingly used for that. Therefore, an addendum has been developed to serve the need for small-scale mapping. The "Guidelines for constructing small-scale map legends using the World Reference Base for Soil Resources" (IUSS Working Group WRB, 2010) is available online at the WRB Web site.

32.4.6 The WRB Reference Soil Groups

The key to the RSGs in the WRB stems from the *Legend of the Soil Map of the World*. The history behind the key to the major soil units of the *Legend of the Soil Map of the World* reveals that it is mainly based on functionality; the key was conceived to derive the correct classification as efficiently as possible. The sequence of the major soil units was such that the central concept of the major soils would come out almost automatically by specifying

TABLE 32.7   An Overview of the Diagnostic Horizons, Properties, and Materials of the WRB

Surface horizons and subsurface diagnostic horizons at shallow depth

Anthraquic horizon	An anthraquic horizon (from Greek anthropos, human, and Latin aqua, water) is a human-induced surface horizon that comprises a puddled layer and a plow pan
Anthric horizon	A moderately thick, dark-colored surface horizon that is the result of long-term cultivation (plowing, liming, fertilization, etc.)
Folic horizon	Surface horizon or subsurface horizon at shallow depth, consisting of well-aerated organic soil material
Fulvic horizon	Thick, black surface horizon having a low bulk density and high organic carbon content conditioned by short-range-order minerals (usually allophane) and/or organoaluminum complexes
Histic horizon	A surface horizon or a subsurface horizon occurring at shallow depth that consists of poorly aerated organic material
Hortic horizon	A human-induced mineral surface horizon that results from deep cultivation, intensive fertilization, and/or long-continued application of human and animal wastes and other organic residues
Hydragric horizon	A human-induced subsurface horizon associated with wet cultivation
Irragric horizon	Human-induced mineral surface horizon that builds up gradually through continuous application of irrigation water with substantial amounts of sediments
Melanic horizon	Thick, black surface horizon conditioned by short-range-order minerals (usually allophane) and/or organoaluminum complexes
Mollic horizon	Well-structured, dark surface horizons with high base saturation and moderate to high organic carbon content
Plaggic horizon	An organic matter—rich, thick, black or brown human-induced low base mineral surface horizon that has been produced by long-continued manuring
Takyric horizon	Finely textured surface horizon consisting of a dense surface crust and a platy lower part; formed under arid conditions in periodically flooded soils
Terric horizon	A human-induced mineral surface horizon that develops through addition of earthy manures, compost, beach sands, or mud over a long period of time
Umbric horizon	Well-structured, dark surface horizon with low base saturation and moderate to high organic matter content
Yermic horizon	Surface horizon of rock fragments ("desert pavement") usually, but not always, embedded in a vesicular crust and covered by a thin eolian sand or loess layer
Voronic horizon	Deep, well-structured, blackish surface horizon with a high base saturation, high organic matter content, strong biological activity, and well-developed, usually granular, structure. Its carbon content is intermediate between a mollic horizon and a histic horizon

Subsurface diagnostic horizons

Albic horizon	Bleached eluviation horizon with the color of uncoated soil material, usually overlying an illuviation horizon
Argic horizon	Subsurface horizon having distinctly more clay than the overlying horizon as a result of illuvial accumulation of clay and/or pedogenetic formation of clay in the subsoil and/or destruction or selective erosion of clay in the surface soil
Cambic horizon	Genetically young subsurface horizon showing evidence of alteration relative to underlying horizons: modified color, removal of carbonates, or presence of soil structure
Cryic horizon	Perennially frozen horizon in mineral or organic soil materials
Calcic horizon	Horizon with distinct calcium carbonate enrichment
Duric horizon	Subsurface horizon with weakly cemented to indurated nodules cemented by silica (SiO_2) known as "durinodes"
Ferralic horizon	Strongly weathered horizon in which the clay fraction is dominated by low activity clays and the sand fraction by resistant materials such as iron-, aluminum-, manganese-, and titanium oxides
Ferric horizon	Subsurface horizon in which segregation of iron has taken place to the extent that large mottles or concretions have formed in a matrix that is largely depleted of iron
Fragic horizon	Dense, noncemented subsurface horizon that can only be penetrated by roots and water along natural cracks and streaks
Gypsic horizon	Horizon with distinct calcium sulfate enrichment
Natric horizon	Subsurface horizon with more clay than any overlying horizon(s) and high exchangeable sodium percentage; usually dense, with columnar or prismatic structure
Nitic horizon	Clay-rich subsurface horizon with a moderate to strong polyhedric or nutty structure with shiny pedfaces
Petrocalcic horizon	Continuous, cemented, or indurated calcic horizon
Petroduric horizon	Continuous subsurface horizon cemented mainly by secondary silica (SiO_2), also known as a "duripan"
Petrogypsic horizon	Cemented horizon containing secondary accumulations of gypsum ($CaSO_4 \cdot 2H_2O$)
Petroplinthic horizon	Continuous layer indurated by iron compounds and without more than traces of organic matter
Pisoplinthic horizon	A pisoplinthic horizon contains nodules that are strongly cemented or indurated with Fe (and in some cases also with Mn)

TABLE 32.7 (continued)   An Overview of the Diagnostic Horizons, Properties, and Materials of the WRB

Plinthic horizon	Subsurface horizon consisting of an iron-rich, humus-poor mixture of kaolinitic clay with quartz and other constituents and which changes irreversibly to a hardpan or to irregular aggregates on exposure to repeated wetting and drying with free access of oxygen
Salic horizon	Surface or shallow subsurface horizon containing 1% of readily soluble salts or more
Sombric horizon	A dark-colored subsurface horizon containing illuvial humus that is neither associated with Al nor dispersed by Na
Spodic horizon	Dark-colored subsurface horizon with illuvial amorphous substances composed of organic matter and aluminum, with or without iron
Thionic horizon	An extremely acid subsurface horizon in which sulfuric acid is formed through oxidation of sulfides
Vertic horizon	Subsurface horizon rich in expanding clays and having polished and grooved ped surfaces ("slickensides") or wedge-shaped structural aggregates formed upon repeated swelling and shrinking
Diagnostic properties	
Abrupt textural change	Very sharp increase in clay content within a limited vertical distance
Albeluvic tonguing	Iron-depleted material penetrating into an argic horizon along ped surfaces
Andic properties	Result from moderate weathering of mainly pyroclastic deposits. The presence of short-range-order minerals and/or organometallic complexes is characteristic for andic properties
Aridic properties	Refer to soil material low in organic matter, with evidence of eolian activity, light in color and (virtually) base-saturated
Continuous rock	Continuous rock is consolidated material underlying the soil, exclusive of cemented pedogenetic horizons such as petrocalcic, petroduric, petrogypsic, and petroplinthic horizons
Ferralic properties	Indicate that the (mineral) soil material has a "low" CEC or would have qualified for a ferralic horizon if it had been less coarsely textured
Geric properties	Mark soil material of very low-effective CEC or even acting as anion exchanger
Gleyic color pattern	Visible evidence of prolonged waterlogging and reducing conditions by shallow groundwater
Lithological discontinuity	Significant changes in particle-size distribution or mineralogy that represent differences in lithology within a soil
Reducing conditions	Lack of oxygen due the saturation of moisture in some parts of the soil
Secondary carbonates	Significant quantities of translocated lime, soft enough to be readily cut with a finger nail, precipitated from the soil solution rather than being inherited from the soil parent material
Stagnic color pattern	Visible evidence of prolonged waterlogging and reducing conditions by a perched water table
Vertic properties	Vertic properties are characterized with slickensides, or wedge-shaped aggregates, or cracks due to high clay content
Vitric properties	Apply to layers with volcanic glass and other primary minerals derived from volcanic ejecta, which contain a limited amount of short-range-order minerals or organometallic complexes
Diagnostic materials	
Artifacts	Artifacts solid or liquid substances that are created or substantially modified or brought to the surface by human activity from a depth, where they were not influenced by surface processes
Calcaric soil material	Soil material that contains more than 2% calcium carbonate equivalent and shows strong effervescence with 10% HCl in most of the fine earth
Colluvic material	Formed by sedimentation through human-induced erosion
Fluvic soil material	Refers to fluviatile, marine, and lacustrine sediments that receive fresh material at regular intervals or have received it in the recent past
Gypsiric soil material	Mineral soil material, which contains 5% or more gypsum (by volume)
Limnic material	Organic and mineral materials that are deposited in water by precipitation or through action of aquatic organisms, such as diatoms and other algae
Mineral material	The soil properties are dominated by mineral components
Organic soil material	Organic debris that accumulates at the surface and in which the mineral component does not significantly influence soil properties
Ornithogenic material	Material with strong influence of bird excrement
Sulfidic soil material	Waterlogged deposit containing sulfur, mostly sulfides, and not more than moderate amounts of calcium carbonate
Technic hard rock	Consolidated material resulting from an industrial process, with properties substantially different from those of natural materials
Tephric soil material	Unconsolidated, non or only slightly weathered products of volcanic eruptions, with or without admixtures of material from other sources

briefly a limited number of diagnostic horizons, properties, or materials. In the current WRB key, the RSGs are allocated to sets on the basis of dominant identifiers, that is, the soil-forming factors or processes that most clearly condition the soil formation. The sequencing of the groups is done according to the following principles:

1. First, organic soils key out to separate them from mineral soils (*Histosols*).
2. The second major distinction in the WRB is to recognize human activity as a soil-forming factor, hence the position of the *Anthrosols and Technosols* after the Histosols; it also appears logical to key out the newly introduced Technosols close to the beginning of the key, for the following reasons: One can almost immediately key out soils that should not be touched (toxic soils that should be handled by experts); a homogeneous group of soils in strange materials is obtained; politicians and decision-makers, who consult the key will immediately encounter these problematic soils.
3. Next are the soils with a severe limitation to rooting (*Cryosols and Leptosols*).
4. Then comes a group of RSGs that are or have been strongly influenced by water: *Vertisols, Fluvisols, Solonetz, Solonchaks,* and *Gleysols.*
5. The following set of soil groups are the RSGs in which iron (Fe) and/or aluminum (Al) chemistry plays a major role in their formation: *Andosols, Podzols, Plinthosols, Nitisols,* and *Ferralsols.*
6. Next comes the set of soils with perched water: *Planosols* and *Stagnosols.*
7. The next grouping comprises soils that occur predominantly in steppe regions and have humus-rich topsoils and a high base saturation: *Chernozems, Kastanozems,* and *Phaeozems.*
8. The next set comprises soils from the drier regions with accumulation of gypsum (*Gypsisols*), silica (*Durisols*), or calcium carbonate (*Calcisols*).
9. Then comes the soils with clay-rich subsoil: *Albeluvisols, Alisols, Acrisols, Luvisols,* and *Lixisols.*
10. Finally, relatively young soils or soils with very little or no profile development, or very homogenous sands, are grouped together: *Umbrisols, Arenosols, Cambisols,* and *Regosols.*

The key to the RSGs of the WRB is presented in Table 32.8. An example for the classification in the WRB system is presented in Figure 32.1.

The dark, cracking, clayey heavy soil satisfies the criteria of the mollic, the vertic, and the calcic horizons. Therefore, this soil will key out in the Vertisols reference group of the WRB. From the qualifiers, the ones indicated with bold in the figure apply. According to the rules of classification (IUSS Working Group WRB, 2007), the name of the soil is *Calcic Mollic Vertisol* (*Pellic*).

TABLE 32.8   The Key to the RSGs of the WRB

Soils having organic material, either

1. Ten centimeters or more thick starting at the soil surface and immediately overlying ice, continuous rock, or fragmental materials, the interstices of which are filled with organic material or
2. Cumulatively within 100 cm of the soil surface either 60 cm or more thick if 75% (by volume) or more of the material consists of moss fibers or 40 cm or more thick in other materials and starting within 40 cm of the soil surface

Histosols (HS)

Other soils having

1. Either a hortic, irragric, plaggic, or terric horizon 50 cm or more thick or
2. An anthraquic horizon and an underlying hydragric horizon with a combined thickness of 50 cm or more

Anthrosols (AT)

Other soils having

1. Twenty percent or more (by volume, by weighted average) artifacts in the upper 100 cm from the soil surface or to continuous rock or a cemented or indurated layer, whichever is shallower or
2. A continuous, very slowly permeable to impermeable, constructed geomembrane of any thickness starting within 100 cm of the soil surface or
3. Technic hard rock starting within 5 cm of the soil surface and covering 95% or more of the horizontal extent of the soil

Technosols (TC)

Other soils having

1. A cryic horizon starting within 100 cm of the soil surface or
2. A cryic horizon starting within 200 cm of the soil surface and evidence of cryoturbation in some layer within 100 cm of the soil surface

Cryosols (CR)

Other soils having

1. One of the following:
   a. Limitation of depth by continuous rock within 25 cm of the soil surface or
   b. Less than 20% (by volume) fine earth averaged over a depth of 75 cm from the soil surface or to continuous rock, whichever is shallower and
2. No calcic, gypsic, petrocalcic, petrogypsic, or spodic horizon

Leptosols (LP)

Other soils having

1. A vertic horizon starting within 100 cm of the soil surface
2. After the upper 20 cm have been mixed, 30% or more clay between the soil surface and the vertic horizon throughout and
3. Cracks that open and close periodically

Vertisols (VR)

Other soils having

1. Fluvic material starting within 25 cm of the soil surface and continuing to a depth of 50 cm or more or starting at the lower limit of a plow layer and continuing to a depth of 50 cm or more and
2. No argic, cambic, natric, petroplinthic, or plinthic horizon starting within 50 cm of the soil surface and
3. No layers with andic or vitric properties with a combined thickness of 30 cm or more within 100 cm of the soil surface and starting within 25 cm of the soil surface

Fluvisols (FL)

TABLE 32.8 (continued)   The Key to the RSGs of the WRB

Other soils having a natric horizon starting within 100 cm of the soil surface

Solonetz (SN)

Other soils having

 1. A salic horizon starting within 50 cm of the soil surface and
 2. No thionic horizon starting within 50 cm of the soil surface

Solonchaks (SC)

Other soils having

 1. Within 50 cm of the mineral soil surface a layer, 25 cm or more thick, that has reducing conditions in some parts and a gleyic color pattern throughout and
 2. No layers with andic or vitric properties with a combined thickness of either
    a. Thirty centimeters or more within 100 cm of the soil surface and starting within 25 cm of the soil surface or
    b. Sixty percent or more of the entire thickness of the soil when continuous rock or a cemented or indurated layer is starting between 25 and 50 cm from the soil surface

Gleysols (GL)

Other soils having

 1. One or more layers with andic or vitric properties with a combined thickness of either
    a. Thirty centimeters or more within 100 cm of the soil surface and starting within 25 cm of the soil surface or
    b. Sixty percent or more of the entire thickness of the soil when continuous rock or a cemented or indurated layer is starting between 25 and 50 cm from the soil surface and
 2. No argic, ferralic, petroplinthic, pisoplinthic, plinthic, or spodic horizon (unless buried deeper than 50 cm)

Andosols (AN)

Other soils having a spodic horizon starting within 200 cm of the mineral soil surface

Podzols (PZ)

Other soils having either

 1. A plinthic, petroplinthic, or pisoplinthic horizon starting within 50 cm of the soil surface or
 2. A plinthic horizon starting within 100 cm of the soil surface and, directly above, a layer 10 cm or more thick, that has in some parts reducing conditions for some time during the year and in half or more of the soil volume, single or in combination
    a. A stagnic color pattern or
    b. An albic horizon

Plinthosols (PL)

Other soils having

 1. A nitic horizon starting within 100 cm of the soil surface
 2. Gradual to diffuse horizon boundaries between the soil surface and the nitic horizon
 3. No ferric, petroplinthic, pisoplinthic, plinthic, or vertic horizon starting within 100 cm of the soil surface and
 4. No gleyic or stagnic color pattern starting within 100 cm of the soil surface

Nitisols (NT)

TABLE 32.8 (continued)   The Key to the RSGs of the WRB

Other soils having

 1. A ferralic horizon starting within 150 cm of the soil surface
 2. No argic horizon that has, in the upper 30 cm, 10% or more water-dispersible clay unless the upper 30 cm of the argic horizon has one or both of the following:
    a. Geric properties or
    b. 1.4 % or more organic carbon

Ferralsols (FR)

Other soils having

 1. An abrupt textural change within 100 cm of the soil surface and, directly above or below, a layer 5 cm or more thick that has in some parts reducing conditions for some time during the year and in half or more of the soil volume, single or in combination
    a. A stagnic color pattern or
    b. An albic horizon and
 2. No albeluvic tonguing starting within 100 cm of the soil surface

Planosols (PL)

Other soils having

 1. Within 50 cm of the mineral soil surface in some parts reducing conditions for some time during the year and in half or more of the soil volume, single or in combination
    a. A stagnic color pattern or
    b. An albic horizon and
 2. No albeluvic tonguing starting within 100 cm of the soil surface

Stagnosols (ST)

Other soils having

 1. A mollic horizon
 2. A Munsell chroma, moist, of 2 or less from the soil surface to a depth of 20 cm or more or having this chroma directly below any plow layer that is 20 cm or more deep
 3. A calcic horizon or concentrations of secondary carbonates starting within 50 cm below the lower limit of the mollic horizon and, if present, above a cemented or indurated layer and
 4. A base saturation (by 1 M NH_4OAc) of 50% or more from the soil surface to the calcic horizon or the concentrations of secondary carbonates throughout

Chernozems (CH)

Other soils having

 1. A mollic horizon
 2. A calcic horizon or concentrations of secondary carbonates starting within 50 cm below the lower limit of the mollic horizon and, if present, above a cemented or indurated layer and
 3. A base saturation (by 1 M NH_4OAc) of 50% or more from the soil surface to the calcic horizon or the concentrations of secondary carbonates throughout

Kastanozems (KS)

Other soils having

 1. A mollic horizon and
 2. A base saturation (by 1 M NH_4OAc) of 50% or more throughout to a depth of 100 cm or more from the soil surface or to continuous rock or a cemented or indurated layer, whichever is shallower

Phaeozems (PH)

TABLE 32.8 (continued)   The Key to the RSGs of the WRB

Other soils having

1. A petrogypsic horizon starting within 100 cm of the soil surface or
2. A gypsic horizon starting within 100 cm of the soil surface and no argic horizon unless the argic horizon is permeated with gypsum or calcium carbonate

Gypsisols (GY)

Other soils having a petroduric or duric horizon starting within 100 cm of the soil surface

Durisols (DU)

Other soils having

1. A petrocalcic horizon starting within 100 cm of the soil surface or
2. A calcic horizon starting within 100 cm of the soil surface and no argic horizon unless the argic horizon is permeated with calcium carbonate

Calcisols (CL)

Other soils having an argic horizon starting within 100 cm of the soil surface with albeluvic tonguing at its upper boundary

Albeluvisols (AB)

Other soils having

1. An argic horizon, which has a CEC (by 1 M NH_4OAc) of 24 $cmol_c$ kg^{-1} clay or more throughout or to a depth of 50 cm below its upper limit, whichever is shallower, either starting within 100 cm of the soil surface or within 200 cm of the soil surface if the argic horizon is overlain by loamy sand or coarser textures throughout and
2. A base saturation (by 1 M NH_4OAc) of less than 50% in the major part between 50 and 100 cm

Alisols (AL)

Other soils having

1. An argic horizon that has a CEC (by 1 M NH_4OAc) of less than 24 $cmol_c$ kg^{-1} clay in some part to a maximum depth of 50 cm below its upper limit, either starting within 100 cm of the soil surface or within 200 cm of the soil surface if the argic horizon is overlain by loamy sand or coarser textures throughout, and
2. A base saturation (by 1 M NH_4OAc) of less than 50% in the major part between 50 and 100 cm

Acrisols (AC)

Other soils having an argic horizon with a CEC (by 1 M NH_4OAc) of 24 $cmol_c$ kg^{-1} clay or more throughout or to a depth of 50 cm below its upper limit, whichever is shallower, either starting within 100 cm of the soil surface or within 200 cm of the soil surface if the argic horizon is overlain by loamy sand or coarser textures throughout

Luvisols (LV)

Other soils having an argic horizon, either starting within 100 cm of the soil surface or within 200 cm of the soil surface if the argic horizon is overlain by loamy sand or coarser textures throughout

Lixisols (LX)

Other soils having an umbric or mollic horizon

Umbrisols (UM)

TABLE 32.8 (continued)   The Key to the RSGs of the WRB

Other soils having

1. A weighted average texture of loamy sand or coarser, if cumulative layers of finer texture are less than 15 cm thick, either to a depth of 100 cm from the soil surface or to a petroplinthic, pisoplinthic, plinthic, or salic horizon starting between 50 and 100 cm from the soil surface
2. Less than 40% (by volume) of gravels or coarser fragments in all layers within 100 cm of the soil surface or to a petroplinthic, pisoplinthic, plinthic, or salic horizon starting between 50 and 100 cm from the soil surface
3. No fragic, irragric, hortic, plaggic, or terric horizon and
4. No layers with andic or vitric properties with a combined thickness of 15 cm or more

Arenosols (AR)

Other soils having

1. A cambic horizon starting within 50 cm of the soil surface and having its base 25 cm or more below the soil surface or 15 cm or more below any plow layer
2. An anthraquic, hortic, hydragric, irragric, plaggic, or terric horizon
3. A fragic, petroplinthic, pisoplinthic, plinthic, salic, thionic, or vertic horizon starting within 100 cm of the soil surface or
4. One or more layers with andic or vitric properties with a combined thickness of 15 cm or more within 100 cm of the soil surface

Cambisols (CM)

Other soils

Regosols (RG)

Source: IUSS Working Group WRB. 2007. World reference base for soil resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome, Italy. Available online at http://www.fao.org/nr/land/soils/soil/wrb-documents/en/

32.5 The French Systems of Soil Classification

32.5.1 Introduction

The *Commission de Pédologie et de Cartographie des Sols* (CPCS, 1967) issued the French soil classification, building on previous work published by Aubert and Duchaufour (1956), which has been the basis for many soil surveys during the 1970s and 1980s, not only in France but also in many of the former French colonies, notably in Africa. It was replaced by the pedological reference base (PRB; *Référentiel Pédologique Français*) (AFES–INRA, 1990, 1992).

32.5.2 The 1967 CPCS System

The CPCS soil classification system comprises four main levels: the class (*classe*), the subclass (*sous-classe*), the group (*groupe*), and the subgroup (*sous-groupe*); followed by four minor levels: the family (*famille*), the series (*série*), the type (*type*), and the phase (*phase*). However, due to the limited knowledge at the time of design of the system, the four minor levels have not been developed for all classes.

The class comprises soils that have main characteristics in common, such as a certain degree of profile development, weathering mode, composition and distribution of organic matter, and
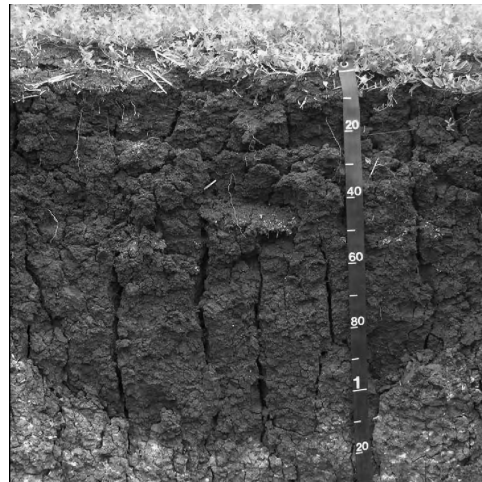
Diagnostics of the example profile:

Dark, high base surface horizon with sufficient structure and organic carbon to satisfy the *mollic horizon*

Vertic subsurface horizon with slickensides and wedge-shaped aggregates and deep, wide cracks

Calcic horizon with sufficient amount (≥15%) of secondary CaCO$_3$

Key to the RSGs (according to Table 32.8, Five RSGs Are Not Satisfied before Getting to the Vertisols in the Key)	**Prefix Qualifiers**	**Suffix Qualifiers**
Other soils having 1. A *vertic* horizon starting within 100 cm of the soil surface 2. After the upper 20 cm have been mixed, 30 percent or more clay between the soil surface and the *vertic* horizon throughout 3. Cracks that open and close periodically Vertisols (VR)	Grumic Mazic Technic Endoleptic Salic Gleyic Sodic Stagnic Mollic Gypsic Duric Calcic Haplic	Thionic Albic Manganiferric Ferric Gypsiric Calcaric Humic Hyposalic Hyposodic Mesotrophic Hypereutric Pellic* Chromic Novic

*Pellic: refers to very dark (Munsell value, moist ≤3.5 and a chroma, moist ≤1.5) surface horizon

Full classification and name: *Calcic Mollic Vertisol* (*Pellic*)
(The applying prefix qualifier names are put before the RSG name and applying suffix qualifier names are placed between brackets following the RSG name).

FIGURE 32.1 Example for the classification in the WRB (IUSS WG WRB, 2007).

predominant soil-forming factors (e.g., wetness). The subclass differentiation is related to criteria resulting from climatic factors, which influence, among others, the pedoclimate. The groups are defined according to morphological characteristics corresponding to soil development, while the subgroup is differentiated either on degree of intensity of the fundamental evolutionary characteristics or on the presence of important secondary soil-forming processes.

In total, 12 classes are distinguished, namely, nondeveloped mineral soils (sols minéraux bruts), slightly developed soils (sols peu évolués), Vertisols, Andosols, Ca/Mg-saturated soils (sols calcimagnésiques), humus-rich soils (sols isohumiques), brunified soils (sols brunifiés), podzolized soils (sols podzolisés), soils rich in Fe (sols sesquioxides de fer), ferralitic soils, hydromorphic soils, and sodic soils.

Nondeveloped mineral soils (*sols minéraux bruts*) show very little trace of soil development apart from some accumulation of organic matter at the surface. They are characterized by an (A)C, (A)R, or R horizon sequence. Included in this class are eroded soils (Lithosols, Régosols), alluvial, colluvial and eolian accumulations, volcanic deposits, manmade soils, nondeveloped soils in the arctic regions (Cryosols), and nondeveloped desert soils.

The class of *slightly developed (sols peu évolués)* soils has higher organic matter content than the nondeveloped soils and is characterized by an AC or AR horizon sequence. No B horizon is permitted in this class. Included are soils from the arctic regions, with or without permafrost, soils with high organic matter content directly overlying hard rock (rankers, soils over limestone, and slightly weathered soils on volcanic ashes), slightly developed desert soils, and soils resulting from erosion and deposition.

Vertisols are described as clayey soils, which are homogenized or irregularly differentiated as a result of internal movement and which are dominated by swell/shrink clays. The normal horizon sequence is A(B)C, A(B)gC, or A(B)Cg. Subdivision into subclasses and groups is based on external drainage factors and type of structure in the surface horizon (rounded or angular).

Andosols are defined as soils in which the mineral fraction is dominated by poorly crystalline minerals and/or metal humus complexes, associated with variable, but usually high amounts of organic matter.

The *Ca/Mg-saturated soils* (*sols calcimagnésiques*) have an exchange complex, 90% of which is saturated with Ca and/or Mg, and have a pH above 6.8. They are generally associated with calcareous or basic rocks, and have an AR, AC, A(B)R, or A(B)C horizon sequence. They are subdivided into (1) carbonate-rich soils (rendzinas, cryptorendzinas, and brown calcareous soils), (2) saturated soils (soils containing only traces of primary $CaCO_3$ in the fine earth fraction), and (3) gypsiferous soils.

The class of *humus-rich soils* (*sols isohumiques*) comprises soils characterized by a moderate to high accumulation of well-humified, polymerized organic matter. Normally, base saturation is high with Ca as dominant cation, followed by Mg and, sometimes, Na. If the base saturation is only moderate (50%–80%) in the upper part of the soil, it increases with depth. Profile evolution is slight to moderate, with an A(B)C or ABC, rarely AC horizon sequence. Subdivision of these soils (Brunizems, Chernozems, Chestnut soils, Brown soils Sierozems), is based at subclass level on pedoclimatic characteristics.

Brunified (*sols brunifiés*) soils are well-developed soils with an A(B)C or ABC horizon sequence and are characterized by the presence of dominantly mull-type humus. They may have a structural or textural B horizon. Subdivision is based on climate (humid temperate, continental temperate, boreal, or tropical) and the morphology of the profile, giving rise to the groups of brown soils (with a structural B horizon), eluvial soils (with a textural B horizon), gray wooded soils, dernopodzolic soils, eluvial boreal soils, and eutrophic brown tropical soils.

The class of *podzolic soils* (*sols podzolisés*) is characterized by the processes of alteration and destruction of the silicate minerals by fulvic acids and complexation of liberated Fe and Al. These processes result morphologically in a strongly depleted and light-colored eluvial horizon and an illuvial horizon, which has a higher organic matter content than the eluvial horizon and a sesquioxide content, which is higher than the original material. Division at subclass level is based on climatic or pedoclimatic characteristics.

Soils rich in Fe (*sols sesquioxydes de fer*) have an ABC or A(B)C profile characterized by the presence of Fe and/or Mn (hydr) oxides giving the soils characteristic red, yellow, rusty brown, or even black (in the case of Mn) colors, an SiO_2/Al_2O_3 ratio of >2, a base saturation of >50%, and a low amount of organic matter. Two subclasses are recognized, one in which the role of sesquioxides is dominant (ferruginous tropical soils) and another in which the behavior of the clay fraction predominates (fersiallitic soils).

The ferralitic soils are characterized by complete weathering of primary minerals, residual enrichment of resistant minerals (quartz, rutile, etc.), loss of nutrients, and the presence of neo-formations such as kaolinite, gibbsite, goethite, hematite, etc. Subclasses are distinguished on the degree of leaching.

The *class of hydromorphic soils* comprises both organic and mineral soils. The three subclasses are separated on organic matter content. The first subclass (hydromorphic organic soils) is defined as having >30% organic matter over a depth of at least 40 cm if the mineral component is clayey or >20% if the mineral component is sandy. The second subclass (moderately organic hydromorphic soils) has between 8% and 30% organic matter over at least 20 cm depth; while the third subclass, the mineral soils (weakly organic hydromorphic soils) have <8% organic matter over a depth of at least 20 cm. The organic hydromorphic soils are at group level separated on decomposition rate (weak or fibric, moderate or hemic, or strong or sapric), while the two other subclasses are divided on the character of hydromorphism (gley or stagnogley) and accumulation of Fe, $CaCO_3$, or gypsum.

The sodic soils comprise both soils that have a high amount of soluble salts as well as soils with a high exchangeable sodium percentage (ESP). This difference is used to separate the subclasses into sodic soils with a nondegraded structure, comprising the saline soils (solonchak) with an AC horizon sequence, and sodic soils with a degraded structure, having an A(B)C or ABC horizon sequence and comprising alkaline saline soils (high ESP), sodic soils with a textural B horizon (solonetz), and so-called solodized solonetz, which are acid at the surface.

32.5.3 The Pedological Reference Base (*Référentiel Pédologique Français*)

This differs basically from the older CPCS system, which it replaces, in that it is being presented as a reference system, not a hierarchical classification (AFES–INRA, 1992). It considers the soil mantles as objects of study for which three sets of data are required: (1) the composition of the soil mantle (mineral, organic, etc.), (2) the internal arrangement of the individual constituents (e.g., structure), and (3) the soil dynamics (e.g., evolution over time).

At the highest level, the PRB recognizes the pedological system, which comprises several associated horizons grouped in a 3D space pattern. A horizon is defined as a part of the soil mantle, which can be considered homogenous. Because dimensions of horizons and pedological systems are not infinite, vertically and laterally, they merge into other systems (e.g., bedrock or other pedological systems).

The authors of the PRB have tried to design a system, which, at the same time, is both scientific and practical as well as precise but flexible. An example of this is the depth indications used in the descriptions of the horizons; the PRB starts off with the tolerance limits concerning depth or thickness requirements, for example, 10 cm must be considered as 5–15 cm, 40 cm means 30–50 cm, etc. Therefore, only two categories are distinguished—the references and the types, the latter being indicated by one or more qualifiers. The reference horizons form the basis of the system. So far, 50 have been proposed (AFES–INRA, 1992), defined, and described by several of the following: (1) morphological characteristics (constituents, pedological features, etc.), (2) analytical data, (3) pedogenetic significance, (4) major possible variations of the characteristics, and (5) most common positions within soil mantles. A succession or combination of reference horizons identifies a diagnostic solum and permits the association of such a solum with a reference. The PRB has proposed some 90 references (Table 32.9). Several of

TABLE 32.9 Major Groupings of References (MGR) and References of the AFES–INRA Pedological Reference Base

MGR	Reference	Brief Description
Alocrisols	Typic Alocrisols	Very acid, brown or yellow soils with a high amount of exchangeable aluminum ($2–8\,cmol_c\,kg^{-1}$ fine earth and Al saturation of 20%–50%)
	Humic Alocrisols	Very acid, brown or yellow soils with a thick, dark colored surface horizon high in organic matter and a high amount of exchangeable aluminum
Alu-Andisols	Humic Alu-Andisols	Non-allophanic soils in weathered volcanic deposits having a thick surface horizon rich in organic matter
	Typic Alu-Andisols	Soils in volcanic deposits or strongly weathered ferralitic material having an aluminum-rich surface horizon and an allophane-dominated subsurface horizon
Andosols	Humic Andosols	Allophane-rich soils having a thick surface horizon rich in organic matter
	Eutric Andosols	Allophane-rich soils having a surface horizon with base saturation > 50%
	Dystric Andosols	Allophane-rich soils having a surface horizon with base saturation < 50%
	Perhydrated Andosols	Allophane-rich soils having a high irreversible water content
Anthroposols	Transformed Anthroposols	Soils modified by intensive or long-continued human activities
	Artificial Anthroposols	Man-made soils consisting of non-soil material (mine refuse, urban debris, etc.)
	Reshaped Anthrosols	Man-made soils consisting of transported soil material
Arenosols	Arenosols	Deep (>120 cm) sandy soils
Brunisols	Saturated Brunisols	Non-calcareous soils with a structural B horizon and 80%–100% base saturation
	Meso-saturated Brunisols	Non-calcareous soils with a structural B horizon and 50%–80% base saturation
	Oligo-saturated Brunisols	Non-calcareous soils with a structural B horizon and 20%–50% base saturation
	Resaturated Brunisols	Non-calcareous soils with a structural B horizon and >80% base saturation due to cultivation
Calcarisols	Calcarisols	Soils with a calcic horizon at least 10 cm thick, starting within 20 cm depth
Calcisols	Calcisols	Soils with non-calcareous, base-saturated (mainly Ca^{2+}) A and B horizons
Calcosols	Calcosols	Soils with calcareous A and B horizons ($CaCO_3$ > 5%)
Castanosols	Castanosols	Soils with a moderately thick to thick, dark colored, base-saturated surface horizon rich in organic matter
Chernosols	Chernosols	Soils with a thick, very dark colored, base-saturated surface horizon rich in organic matter
Colluviosols	Colluviosols	Soils in colluvial deposits
Cryosols	Histic cryosols	Soils with permafrost within 1 m depth and a histic surface horizon
	Mineral cryosols	Soils with permafrost within 2 m depth lacking a histic surface horizon
Dolomitosols	Dolomitosols	Soils with dolomitic A and B horizons (molar ratio of $CaCO_3/MgCO_3$ < 1.5)
Ferrallisols	Soft Ferrallisols	Strongly weathered soils with a ferralitic or oxidic mineralogy
	Nodular Ferrallisols	Strongly weathered soils with a high amount of sesquioxide nodules
	Petroxydic Ferrallisols	Strongly weathered soils with indurated sesquioxide layers (e.g., cuirasses)
Fersialsols	Carbonated Fersialsols	Calcareous soils with significant amounts of 2:1 clays and "free iron"
	Saturated Fersialsols	Base-saturated soils with significant amounts of 2:1 clays and "free Fe"
	Desaturated Fersialsols	Desaturated soils with significant amounts of 2:1 clays and "free Fe"
	Xanthic Fersialsols	Yellowish soils with significant amounts of 2:1 clays and "free Fe"
Fluviosols and thalassosols	Raw Fluviosols	Soils in fluviatile deposits lacking any horizon development
	Typical Fluviosols	Soils in fluviatile deposits with one or more not fully developed reference horizons
	Brunified Fluviosols	Soils in fluviatile deposits with a well-developed structural B horizon
	Thalassosols	Non- or weakly developed soils in marine or fluvio-marine deposits
Gypsosols	Gypsosols	Soils with accumulation of gypsum
Histosols	Leptic Histosols	Shallow organic soils with consolidated or unconsolidated rock within 40 cm
	Fibric Histosols	Organic soils with weakly decomposed organic material more than 60 cm thick
	Mesic Histosols	Organic soils with moderately decomposed organic material more than 40 cm thick
	Sapric Histosols	Organic soils with strongly decomposed organic material more than 40 cm thick
	Composite Histosols	Organic soils without dominance of either fibric, mesic, or sapric materials between 40 and 120 cm
	Covered Histosols	Organic soils with a cover of mineral soil material 10–40 cm thick
	Floating Histosols	Organic soils on water occurring between 40 and 160 cm depth
Lithosols	Lithosols	Shallow soils (<10 cm) over continuous hard rock or indurated layer
Luvisols	Neoluvisols	Soils with a moderately developed eluvial and well developed textural B horizon
	Typic Luvisols	Soils with a well-developed eluvial and textural B horizon
	Degraded Luvisols	Soils with a well-developed, partially light colored and hydromorphic eluvial horizon penetrating a gleyed textural B horizon

TABLE 32.9 (continued)   Major Groupings of References (MGR) and References of the AFES–INRA Pedological Reference Base

MGR	Reference	Brief Description
	Dernic Luvisols	Soils with a well-developed, partially light colored eluvial horizon penetrating a textural B horizon
	Truncated Luvisols	Soils with a textural B horizon but lacking an eluvial horizon
Magnesisols	Magnesisols	Soils with non-calcareous, base-saturated ($Ca^{2+}/Mg^{2+} < 2$) A and B horizons
Organosols	Calcaric Organosols	Well-drained, organic matter rich (>8% organic C), calcareous soils directly overlying an unconsolidated or consolidated substratum
	Saturated Organosols	Well-drained, organic matter rich (>8% organic C), base-saturated ($Ca^{2+}/Mg^{2+} > 5$) soils directly overlying an unconsolidated or consolidated substratum
	Undersaturated Organosols	Well-drained, organic matter rich (>8% organic C), undersaturated (BS < 80%) soils directly overlying an unconsolidated or consolidated substratum
	Tangelic Organosols	Well-drained, organic matter rich (>8% organic C), base-saturated soils with a thick, greasy horizon consisting of soil animal casts ("tangel horizon")
Pelosols	Typic Pelosols	Clay-rich, slightly weathered soils lacking coloration in the B horizon
	Brunified Pelosols	Clay-rich, slightly weathered soils with a brown colored B horizon
	Differentiated Pelosols	Clay-rich, slightly weathered soils with a clear eluvial horizon
Peyrosols	Stony Peyrosols	Soils which have throughout the upper 50 cm 40% or more stones plus 20% or more other coarse fragments
	Gravelly Peyrosols	Soils which have throughout the upper 50 cm 60% or more gravel, stones and boulders, but less than 40% stones
Planosols	Typic Planosols	Soils with abrupt textural change and temporary perched watertable within 50 cm
	Distal Planosols	Soils with abrupt textural change and temporary perched watertable below 50 cm
	Structural Planosols	Soils with a temporary perched watertable within 50 cm of the surface caused an impermeable layer which is not texturally induced (e.g., fragipan, duripan)
Podzosols	Duric Podzosols	Soils with an eluvial horizon and a cemented podzol B horizon
	Humo-Duric Podzosols	Soils with an indurated podzol B horizon, lacking an eluvial horizon
	Soft Podzosols	Soils with an eluvial horizon and a soft podzol B horizon
	Placic Podzosols	Soils with podzol B horizon and a placic horizon
	Ochric Podzosols	Soils with a weakly developed humic podzol B horizon, lacking an eluvial horizon
	Humic Podzosols	Soils with a soft humic podzol B horizon, lacking an eluvial horizon
	Post_Podzosols	Man-modified soils in which (remnants of) the podzol B horizon can be recognized
	Eluvial Podzosols	Soils lacking a podzol B horizon, but having lateral linkage to a podzol B horizon
Rankosols	Rankosols	Soils with a moderately thick A horizon with non-calcareous coarse fragments overlying consolidated or unconsolidated rock
Reductisols and redoxisols	Typic Reductisols	Hydromorphic soils conditioned by saturation of fluctuating groundwater table
	Stagnic Reductisols	Hydromorphic soils conditioned by a perched water table
	Duplex Reductisols	Hydromorphic soils conditioned by groundwater and stagnating surface water
	Redoxisols	Soils with a textural discontinuity and a perched water table
Regosols	Regosols	Shallow soils (<10 cm) over unconsolidated material or only slightly coherent rock
Rendisols	Rendisols	Soils with non-calcareous, base-saturated (mainly Ca^{2+}) A horizon over consolidated or unconsolidated calcareous rock
Rendosols	Rendosols	Soils with deep (>30 cm) calcareous A horizon ($CaCO_3$ > 5%) over consolidated or unconsolidated calcareous rock
Salisols	Chloridi-Sulfatic Salisols	Neutral soils with a high amount of sodium, magnesium or calcium salts
	Carbonatic Salisols	Alkaline soils with a high amount of carbonate/bicarbonates
Sodisols	Undifferentiated Sodisols	Alkaline soils with a high amount of exchangeable sodium
	Solonetzic Sodisols	Soils with clay illuviation and moderate leaching of sodium in the upper part
	Solodic Sodisols	Soils with clay illuviation and strong leaching of sodium in the upper part
Sulfatosols	Sulfatosols	Very acid soils with jarosite within 50 cm depth
Thiosols	Thiosols	Waterlogged soils containing sulfide minerals, rapidly acidifying upon aeration
Veracrisols	Veracrisols	Soils with a thick (50–150 cm), acid, dark colored surface horizon with a high biological activity overlying a slowly permeable horizon (e.g., textural B horizon)
Vertisols	Topovertisols	Deep, clayey soils in level, low-lying positions which crack and show gilgai microrelief, slick
Vitrandisols	Vitrandisols	Soils in slightly weathered pyroclastic material
Yermosols	Yermosols	Hot desert soils

Source: AFES–INRA (Association Française pour l'Etude du Sol) (Institut National de la Recherche Agronomiqie). 1992. Référentiel Pédologique, principaux sols d'Europe. INRA, Paris, France.

these are closely associated with each other because, for example, they may have the same reference horizons. Such references are described together as major groupings of references to avoid duplication and to associate the references with traditional pedological concepts. For example, the Podzosols major grouping of references comprises seven references characterized by a process of podzolization. However, major groupings of references do not form part of the PRB.

32.6 The Soil Classification System of Russia

32.6.1 The Soil Classification System of the Former USSR

32.6.1.1 Introduction

This system was available for soil survey from 1967 as a technical document. An official revised version was published in Russian in 1977 and in English in 1986 (VASKhNIL, 1986). This classification serves as a manual for soil examination and survey in many countries of the former USSR, for governmental assessment of land records and evaluation, as well as a reference book on agricultural and industrial planning for agronomists, land use planners, reclamation specialists, and others. It was amended (Shishov and Sokolov, 1990) to correct obvious errors and to integrate new knowledge and data. The taxonomic levels of the classification system are expanded and new names have been introduced for a number of soils.

32.6.1.2 Structure of the System

The soils are not defined as sequences of diagnostic horizons with fixed frames of properties, like in many other systems. Every soil type has a description of a "central concept," a typical profile. Profound knowledge of the soil-forming processes is required in order to recognize and classify the soils. The higher levels are known as types, subtypes, genera, and species. At subtype level, apart from subtypes with overlapping soil-forming process, the facies modifier may be added to indicate the thermal regime. Twenty-seven of these facies are recognized, ranging from arctic permafrost to subtropical hot nonfreezing.

32.6.1.3 Brief Description of the Types

Some 71 soil types are distinguished at the highest level and are characterized as follows:

Podzolic soils are characterized by either downward movement of organic acids facilitating decomposition of primary and secondary minerals and removal of weathering products or downward movement of clay particles.

Bog-podzolic soils have a combination of downward movement of organic acids or clay particles with either stagnating water conditions in the upper part of the soil or groundwater affecting the bottom part, resulting in a

bleached eluvial horizon and hydromorphic properties in both the eluvial and illuvial horizons. Some of these soils have a peaty layer at the surface.

Sod-calcareous soils have a dark-colored, base-saturated, and humus-rich surface horizon overlying calcareous parent material.

Sod-gley soils are poorly drained with a dark, humus-rich surface horizon and hydromorphic features at shallow depth.

Gray forest soils have a dark-colored, humus-rich surface horizon of variable thickness with a bleached horizon or conspicuous white powdery spots overlying a clay-illuviated subsurface horizon.

Gley gray forest soils are similar to those above but with distinct hydromorphic features below the surface horizon.

Brown forest soils (*or Burozems*) have well-developed nut-like structure and intensive brownish and yellow color in B horizon combined with certain enrichment with organic matter in the surface horizon(s).

Gley brown forest soils (*or Gley burozems*) are similar to those above except for clear hydromorphic features below the surface horizon.

Podzolic brown forest soils (*or Podzolic burozems*) have a clearly developed eluvial horizon, clay-illuvial horizon, and weak hydromorphism due to seasonal surface waterlogging.

Gley podzolic brown forest soils (*or Gley podzolic burozems*) are similar to those above, but with seasonal wetness more pronounced. The process of acidic hydrolysis may take place in the upper part of these soils.

Bleached meadow soils (*Podbels*) are seasonally waterlogged with a bleached horizon near the surface, in which segregation of Fe in concretions is the main ongoing process.

Meadow chernozem-like soils have a thick, dark-colored, and humus-rich surface horizon and distinct features of hydromorphism (gray and rusty colors, Fe/Mn concretions, white powdery coatings) in the lower part of the soils, with no evidence of secondary carbonates.

Chernozem-like dark meadow soils are waterlogged with a peaty or mucky surface horizon overlying a dark-colored, humus-rich mineral horizon with rust colored mottles.

Chernozems are well-drained, base-saturated, or only slightly undersaturated soils with a thick, dark-colored surface horizon rich in organic matter and accumulation of illuvial secondary carbonates.

Meadow-chernozem soils are similar to those above, but with some features of wetness in the lower part of the solum.

Chestnut soils are well-drained and base-saturated with a dark-colored surface horizon, which is less thick and less rich in organic matter than in chernozems. The lower part of the solum often contains accumulations of calcium carbonate and/or gypsum.

Meadow chestnut soils are similar as the soils above, but with distinct features of wetness in the lower part of the solum.

Meadow soils are conditioned by a brief period of surface waterlogging and a longer period of saturation by groundwater, resulting in humus-rich surface horizons overlying a gleyed subsoil.

Semidesert brown soils have an accumulation of calcium carbonate, possibly overlying accumulations of gypsum and a crusty surface horizon.

Semidesert meadow brown soils are similar to those above, but with a higher organic matter content, weak signs of hydromorphism, and deeper $CaCO_3$ accumulations.

Desert gray brown soils are calcareous with a low organic matter content and a variable degree of salinization.

Desert takyr-like soils are weakly developed with a friable porous surface crust.

Takyrs have a hard, porous but crusted surface horizon cracking into polygonal patterns.

Desert sandy soils are coarse textured with little horizon differentiation apart from some accumulation of organic matter and segregation of calcium carbonate at depth.

Meadow desert soils are poorly differentiated, characterized by enrichment with organic matter at the surface and signs of hydromorphism in the subsoil.

Serozems have a shallow gray humus-enriched horizon and a calcareous illuvial layer in the subsoil.

Meadow-serozem soils are similar to those above, but signs of wetness occur in the deeper subsoil.

Semidesert and desert meadow soils have periodic or permanent wetness through capillary rise reaching the surface, giving rise to a well-developed, humus-rich surface horizon and a gleyed subsoil, in which carbonate concentrations are linked to the groundwater level.

Irrigated soils include a variety of types, which are all related to the original soil (irrigated serozems, irrigated brown soils, irrigated meadow brown soils, irrigated gray brown soils, irrigated takyr-like soils, irrigated meadow desert soils, irrigated meadow soils, and irrigated bog soils), in which irrigation has caused considerable modification including enhancement of biological activity, leaching, accumulation of sediments from irrigation water, enrichment in carbonates and soluble salts, etc.

Gray-cinnamon brown soils have a well-developed surface horizon with a low amount of organic matter and a clay-enriched subsurface horizon. The soils are calcareous throughout and differ from chestnut soils in that they do not freeze during wintertime.

Meadow gray-cinnamon brown soils are similar to those above, but with clear indications of increased wetness in the subsoil.

Cinnamon brown soils are deep with a high amount of organic matter, a clay-enriched subsurface horizon with a characteristic cinnamon brown color, and secondary carbonates accumulation.

Meadow cinnamon brown soils are similar to those above, but with clear indications of increased wetness in the subsoil.

Gray meadow forest soils have a thick, humus-enriched surface horizon and hydromorphic features starting at or near to the surface.

Zheltozems are leached subtropical soils with no or only weak textural differentiation and rich in sesquioxides.

Gley zheltozems are similar to those above, but with pronounced gleying throughout the profile.

Podzolic–zheltozem soils are leached subtropical soils with clear textural differentiation and a high content in sesquioxides. Gley features are common in the transition zone between the eluvial and illuvial horizons.

Podzolic–zheltozem gley soils are similar to those above, but with pronounced gleying throughout the profile, resulting in the accumulation of Fe in the illuvial horizon.

Krasnozems are strongly weathered subtropical soils with a high amount of sesquioxides. The clay fraction mainly contains kaolinite, halloysite, goethite, and gibbsite.

Peat-bog soils are a group that includes two soil types, high peat-bog soils and low peat-bog-soils. Both are waterlogged organic soils, the first type forming in upland positions and fed with rainwater, the second one—in lowland positions and fed with groundwater.

Reclaimed peat soils include corresponding two types that are drained peat soils with a plow layer.

Meadow-bog soils are waterlogged mineral soils with or without a shallow organic layer at the surface.

Bog soils of the semideserts and deserts have shallow groundwater (usually <50 cm) and an organic matter-rich surface horizon in desert or semidesert conditions.

Solods are degraded solonetzes and solonetzic soils of which the upper horizons are acidified, resulting in a well-differentiated soil with a humus-rich surface horizon, a white eluvial horizon, and a brownish colored, compact illuvial B horizon.

Solonetzes conform a group of types that have a high amount of exchangeable Na, an (near) absence of readily soluble salts, resulting in a well-expressed eluvial horizon and a compact illuvial horizon. There are three types in this group: automorphic (related to the parent material), semihydromorphic, and hydromorphic (related to groundwater influence) solonetzes.

Solonchaks have a high amount of soluble salts. Two soil types, automorphic and hydromorphic solonchaks, are distinguished.

Alluvial soils are divided into three main groups: alluvial soils with deep groundwater and only a brief period of flooding (sod alluvial soils), alluvial soils influenced by both flooding and groundwater at moderate (1–2 m)

depth (meadow alluvial soils), and alluvial soils that are conditioned by a long period of flooding or shallow groundwater in combination with surface flooding (bog alluvial soils). In addition, the soil reaction, related to soil zone, is used to further subdivide these groups into types.

Mountain meadow soils occur in cold and moist high mountains, which receive a large excess of moisture resulting in a strongly leaching regime, with an acid soil reaction and a considerable accumulation of organic matter in the surface horizon.

Chernozem-like mountain meadow soils occur in high mountains, which, although receiving an excess in moisture, have only a moderately leaching regime, resulting in a weakly acid to weakly alkaline soil reaction as well as a considerable accumulation of organic matter in the surface horizon.

Mountain meadow-steppe soils develop under similar conditions as above but have a much lower exchange capacity than the chernozem-like mountain meadow soils.

An overview of the types and subtypes distinguished in the 1986 USSR soil classification is given in Table 32.10.

32.6.2 The Soil Classification System of the Russian Federation

32.6.2.1 Introduction

The 1977 classification and diagnostics of soils of USSR (VASKhNIL, 1986) did not satisfy both scientists and practical experts. Just 5 years after its publication, the Dokuchaev Soil Science Institute initiated work on a new version of soil classification. A complete version was published in 1997; an English translation is also available, in an adapted and improved form (Arnold, 2001). The discussion of the classification resulted in its further revision, and soon after, a new version was published (Shishov et al., 2004; see also Krasilnikov et al. 2009) for detailed description and correlation. In the Russian Federation, this system is currently introduced and used along with the older soil classification of the USSR.

32.6.2.2 Structure of the System

Like most actual soil classifications, the Russian taxonomy uses the concept of diagnostic horizons; a unique sequence of genetic diagnostic horizons comprises a soil type. In the Russian classification, the designated horizons are mutually exclusive, which provides more order to the taxonomy; for example, this classification has a key for diagnostic horizons that is impossible in other classifications. Soil types are grouped in sections and trunks. On the lower level, the types are subdivided on the basis of overlapping pedogenetic processes, particular properties, texture, and origin of parent material. The Russian classification stresses the importance of agricultural transformation of soils: soils having minor agrogenic impact are defined as subtypes of

TABLE 32.10 Types and Subtypes of the 1986 USSR Soil Classification

Types	Subtypes
Podzolic soils	Gley-podzolic soils
	True podzolic soils
	Sod-podzolic soils
Bog-podzolic soils	Surface gleyed peaty podzolic soils
	Surface gleyed soddy-podzolic soils
	Surface gleyed mucky-podzolic soils
	Subsoil gleyed peaty-podzolic soils
	Subsoil gleyed soddy-podzolic soils
	Subsoil gleyed mucky-podzolic soils
Sod-calcareous soils	Typical sod-calcareous soils
	Leached sod-calcareous soils
	Podzolized sod-calcareous soils
Sod-gley soils	Sod surface gleyey soils
	Mucky surface gley soils
	Sod subsurface gleyey soils
	Mucky subsurface gley soils
Gray forest soils	Light gray forest soils
	Gray forest soils
	Dark gray forest soils
Gley gray forest soils	Surface gleyey (and surface meadow gray forest) soils
	Subsurface gleyey gray forest soils
	Subsurface gley gray forest soils
Brown forest soils	Acid mor brown forest soils
	Acid mor podzolized brown forest soils
	Acid brown forest soils
	Acid podzolized brown forest soils
	Slightly unsaturated brown forest soils
	Slightly unsaturated podzolized brown forest soils
Gley brown forest soils	Podzolized surface gleyey brown forest soils
	Podzolized surface gley brown forest soils
	Gleyey brown forest soils
	Gley brown forest soils
Podzolic brown forest soils	Unsaturated podzolic brown forest soils
	Slightly unsaturated podzolic brown forest soils
Gley podzolic brown forest soils	Surface gleyey podzolic brown forest soils
	Surface gley podzolic brown forest soils
	Gleyey podzolic brown forest soils
	Gley podzolic brown forest soils
Bleached meadow soils	Podzolized bleached meadow soils
	Podzolized gley bleached meadow soils
Meadow chernozem-like soils	Meadow chernozem-like soils (surface-wet)
Chernozem-like dark meadow soils	Dark meadow prairie soils
	Dark moist-meadow prairie soils
Chernozems	Podzolized chernozems
	Leached chernozems
	Typical chernozems
	Ordinary chernozems

TABLE 32.10 (continued) Types and Subtypes of the 1986 USSR
Soil Classification

Types	Subtypes
	Southern chernozems
	Mountain chernozems
Meadow-chernozem soils	Meadowy chernozemic soils
	Meadow-chernozem soils
Chestnut soils	Dark chestnut soils
	Chestnut soils
	Light chestnut soils
	Mountain chestnut soils
Meadow chestnut soils	Meadowy chestnut soils
	Meadow chestnut soils
Meadow soils	Meadow soils
	Moist-meadow soils
Semi-desert brown soils	Semi-desert brown soils
Semi-desert meadow brown soils	Semi-desert meadowy brown soils
	Semi-desert meadow brown soils
Desert gray brown soils	Desert gray brown soils
Desert takyr-like soils	Desert takyr-like soils
Takyrs	Takyr
Desert sandy soils	Desert sandy soils
Meadow desert soils	Meadow desert (meadowy takyr-like) soils
	Meadow desert (meadow takyr-like) soils
	Meadow desert soils with complementary surface moistening
	Gray brown meadow desert soils
	Sandy meadow desert soils
Serozems light colored	Typical serozems
	Dark serozems
Meadow-serozem soils	Meadowy serozem
	Meadow serozem
Semi-desert and desert meadow soils	Semi-desert and desert meadow
	Semi-desert and desert moist-meadow soils
Irrigated serozems	Irrigated light colored serozem soils
	Irrigated typical serozem soils
	Irrigated dark serozem soils
	Old irrigated serozem soils
Irrigated meadow serozems	Irrigated meadow serozem soils
	Irrigated serozem meadow soils
Irrigated brown soils	
Irrigated meadow brown soils	
Irrigated gray brown soils	
Irrigated takyr-like soils	Irrigated takyr-like soils
	Ancient irrigated takyr-like soils
Irrigated meadow desert soils	Irrigated meadow desert soils
	Ancient irrigated meadow desert soils
Irrigated meadow soils	Irrigated meadow soils
	Irrigated moist meadow soils
	Ancient irrigated meadow soils
Irrigated bog soils	
Gray-cinnamon brown soils	Dark gray-cinnamon brown soils

TABLE 32.10 (continued) Types and Subtypes of the 1986 USSR
Soil Classification

Types	Subtypes
	Common gray-cinnamon brown soils
	Light gray-cinnamon brown soils
Meadow gray-cinnamon brown	Surface-meadowy gray-cinnamon brown soils
	Meadowy gray-cinnamon brown soils
	Meadow gray-cinnamon brown soils
Cinnamon brown soils	Leached cinnamon brown soils
	Typical cinnamon brown soils
	Calcareous cinnamon brown soils
Meadow cinnamon brown soils	Surface-meadowy cinnamon brown soils
	Meadowy cinnamon brown soils
	Meadow cinnamon brown soils
Gray meadow-forest soils	Gray meadow-forest soils
	Gray wet meadow-forest soils
Zheltozems	Unsaturated zheltozems
	Weakly unsaturated zheltozems
	Unsaturated podzolized zheltozems
	Weakly unsaturated podzolized zheltozems
Gley zheltozems	Surface gleyey zheltozems
	Gleyey zheltozems
	Gley zheltozems
Podzolic–zheltozem soils	Unsaturated podzolic–zheltozem soils
	Slightly unsaturated podzolic–zheltozem soils
Podzolic–zheltozem gley soils	Podzolic–zheltozem surface-gleyey soils
	Podzolic–zheltozem gleyey soils
	Podzolic–zheltozem gley soils
Krasnozems	Typical krasnozems
	Podzolized krasnozems
Upland peat-bog soils	Upland peat-gley bog soils
	Upland peat-bog soils
Lowland peat-bog soils	Depleted lowland peat-gley bog soils
	Lowland (typical) peat-gley bog soils
	Depleted lowland peat-bog soils
	Lowland (typical) peat-bog soils
Reclaimed upland peat soils	
Reclaimed lowland peat soils	Reclaimed depleted lowland peat-gley soils
	Reclaimed depleted lowland peat soils
	Reclaimed lowland muck-gley soils
	Reclaimed lowland mucky-peat soils
Meadow bog soils	Mucky meadow-bog soils
	Clayey meadow-bog soils
Bog soils (semi-desert/deserts)	Peat-bog soils
	Clayey-bog soils
Solods	Meadow-steppe solods
	Solods meadow
	Meadow-bog solods
Automorphic solonetzes	Chernozemic solonetzes
	Chernozemic solonchak-solonetzes
	Solonchakic chernozemic solonetzes
	Deep-solonchakic chernozemic solonetzes
	Deep-salinized chernozemic solonetzes
	Chestnut solonetzes
	Semi-desert solonetzes

Types	Subtypes
Semihydromorphic solonetzes	Meadow-chernozemic solonetzes
	Meadow-chernozemic solonchak-solonetzes
	Solonchakic meadow-chernozemic solonetzes
	Deep-solonchakic meadow-chernozemic solonetzes
	Meadow-chestnut solonetzes
	Meadow semi-desert solonetzes
	Semihydromorphic cryogenic solonetzes
Hydromorphic solonetzes	Chernozemic-meadow solonetzes
	Chestnut-meadow solonetzes
	Meadow-bog solonetzes
	Cryogenic-meadow solonetzes
Automorphic solonchaks	Typical automorphic solonchaks
	Takyrized automorphic solonchaks
Hydromorphic solonchaks	Typical solonchaks
	Meadow-solonchaks
	Bog-solonchaks
	Sor-solonchaks
	Mud-volcanic solonchaks
	Hummocky solonchaks
Sod acidic alluvial soils	Primitive stratified sod acidic alluvial soils
	Stratified sod acidic alluvial soils
	Typical sod acidic alluvial soils
	Podzolized sod acidic alluvial soils
Saturated sod alluvial soils	Primitive stratified saturated sod alluvial soils
	Stratified saturated sod alluvial soils
	Typical stratified sod alluvial soils
	Saturated steppe sod alluvial soils
Calcareous sod desertified alluvial soils	Primitive stratified calcareous sod desertified alluvial soils
	Stratified calcareous sod desertified alluvial soils
	Typical calcareous sod desertified alluvial soils
Meadow acid alluvial soils	Primitive stratified acid meadow alluvial soils
	Stratified acid meadow alluvial soils
	Typical acid meadow alluvial soils
Meadow saturated alluvial soils	Primitive stratified saturated meadow alluvial soils
	Stratified saturated meadow alluvial soils
	Typical saturated meadow alluvial soils
	Dark colored saturated meadow alluvial soils
Calcareous meadow alluvial soils	Stratified calcareous meadow alluvial soils
	Tugai calcareous meadow alluvial soils
	Typical calcareous meadow alluvial soils
Meadow-bog alluvial soils	Typical meadow-bog alluvial soils
	Peaty meadow-bog alluvial soils

TABLE 32.10 (continued) Types and Subtypes of the 1986 USSR Soil Classification

Types	Subtypes
Clayey-peat bog alluvial soils	Clayey-peat-gley bog alluvial soils
	Clayey-peat bog alluvial soils
Mountain meadow soils	Alpine mountain meadow soils
	Subalpine mountain meadow soils
Chernozem-like mountain meadow soils	Typical chernozem-like mountain meadow soils
	Leached chernozem-like mountain meadow soils
	Calcareous chernozem-like mountain meadow soils
Mountain meadow-steppe soils	Alpine mountain meadow-steppe soils
	Subalpine mountain meadow-steppe soils

Source: VASKhNIL (V.V. Dokuchaev Institute of Soil Science). 1986. Classification and diagnostics of soils of the USSR. Translated from Russian by S. Viswanathan. Amerind Publishing Co., New Delhi, India.

superficially disturbed soils, those under intensive cultivation enter "agrotypes" within the same section (e.g., *agropodzolic soils* vs. *podzolic soils*), and those completely transformed by agricultural practices are grouped in a special section of Agrozems.

32.6.2.3 Brief Description of the Sections

The classification includes 227 soil types, which are grouped in 27 sections and 3 trunks, defined as follows:

Postlithogenic soils: *a trunk of soils*—bringing together soils where soil-formation processes occur on a previously formed parent material and modern accumulation of matter on the surface is negligible. Includes the following soil sections:

Texture differentiated soils: soils with distinct clay illuviation, expressed both as the difference in clay content between the topsoil and B horizon and as clay coating in the illuvial horizons

Al–Fe–humus soils: soils having illuviation of aluminum and iron in complexes with organic matter

Iron metamorphic soils: soils having a B horizon, altered in situ. The alteration is expressed mainly as iron (hydr) oxides crystallization

Structural metamorphic soils: soils having a B horizon, altered in situ. The alteration is expressed mainly as soil structure modification comparing with that of parent material

Cryometamorphic soils: soils having a B horizon, altered in situ. The alteration is related mainly to the presence of permafrost and the effect of freezing–thawing cycles in the topsoil

Pale metamorphic soils: soils having structural alteration, combined with secondary calcium carbonates and (provisionally) soluble salts accumulation. These soils are geographically related to extracontinental cryoarid regions of Eastern Siberia

Cryoturbated soils: soils formed in the presence of permafrost and affected with soil material cryoturbation due to freezing–thawing cycles in the topsoil

Gleyic soils: soils affected with excessive moisture due to high groundwater level

Humus-accumulating soils: soils having a deep dark, well-structured humus-accumulative topsoil horizon

Light-humus carbonate-accumulating soils: soils having relatively poor and light topsoil, combined with secondary calcium carbonates accumulation

Alkaline clay-differentiated soils: soils with alkaline reaction and with a clay-illuvial horizon

Halomorphic soils: soils with toxic concentrations of soluble salts

Hydrometamorphic soils: soils affected with excessive moisture due to high groundwater level but having no typical "gleyic" (bluish-green) colors due to high organic matter content and/or alkalinity

Organic matter–accumulating soils: soils with no other diagnostic horizons than humus-accumulative one

Eluvial soils: soils with a distinct eluvial (clay and/or iron depleted) horizon but without evidences of clay and iron illuviation beneath

Lithozems: shallow soils (less than 30 cm) with a humus-accumulative horizon

Weakly developed soils: immature soils either in sandy or clayey loose materials or over consolidated rock

Abrazems: strongly eroded soils with exposed B or BC horizon

Agrozems: soils mixed in the course of agricultural management down to the B or BC horizon

Agroabrazems: cultivated soils, previously affected with intensive erosional processes

Turbated soils: soils mixed down to the depth of more than 1 m in the process of land improvement

Sinlithogenic soils: *a trunk of soils*—gathering together the soils, where soil formation occurs in the condition of periodical or continuous accumulation of new material on the surface. It includes the following soil sections:

Alluvial soils: soils formed in actual river valleys in recent alluvial deposits

Volcanic soils: soils formed under influence of recent periodic volcanic ash deposition

Stratozems: include all the soils with periodical eolian or fluvial accumulation of sediments (including long-term irrigation) with no or weak development of soil profile on the surface

Weakly developed soils: soils formed in fresh eolian or fluvial sediments, with incipient pedogenesis

Organogenic soils: *a trunk of soils*—including the soils formed in organic material, mainly in peat. There are two sections within this trunk, namely:

Peat soils: soils formed in organic materials

Torfozems: drained peat soils, in places deeply cultivated and ameliorated with mineral ground

32.7 The Chinese Soil Taxonomic Classification

32.7.1 Introduction

Until 1949, the Chinese soil classification was based on that of the United States but then was replaced by that of the USSR geographical classification (ISS-AS, 1990), and integrating locally important soils, such as long-continued cultivated soils, paddy soils among others. In 1994, a drastically renewed first proposal for a new *Chinese Soil Taxonomic Classification* (CSTC) based largely on *Soil Taxonomy* was issued (Gong, 1994), which is still under review and subject to modification. In developing the CSTC system, many elements of the *Legend of the Soil Map of the World* (FAO–UNESCO, 1974) and other soil classifications were incorporated.

32.7.2 Structure

The CSTC is a hierarchical system with seven categories: order, suborder, group, subgroup, genus, species, and variety. The first four levels are used to construct mapping units for small-scale maps, the lower three for more detailed maps.

The order is based on soil properties, which result from or reflect major soil-forming processes. The suborder is defined according to soil properties, which either control recent soil-forming processes or reflect dominant limiting factors. At group level, intensities of major or secondary soil-forming processes are used to differentiate among the soils, while the subgroup reflects the deviation from the central concept of the group.

The system presently has defined 13 orders, 33 suborders, 78 groups, and 301 subgroups. The nomenclature used is a mixture of older and more recent names, as well as local names for typical Chinese soils. The names of the order and suborder are linked, the group and subgroup nomenclature is different, to avoid names becoming too long. An overview of the orders, suborders, and groups of the Chinese Soil Taxonomic System is given in Table 32.11.

32.7.3 Diagnostic Horizons and Characteristics

Like many other soil classification systems, the Chinese Soil Taxonomic System uses diagnostic horizons and characteristics to identify the soil. A number of the diagnostic horizons are directly taken from other systems, such as the argillic horizon from *Soil Taxonomy*, while others are newly defined to suit Chinese conditions. Thirty diagnostic horizons are defined, 8 surface horizons or epipedons, 10 subsurface horizons, and 12 horizons, which may occupy any position in the soil profile.

Three categories of diagnostic surface horizons are recognized: histic, humus (isohumic, umbrihumic, and ochrihumic epipedons), and anthropic epipedons. Particularly, the latter category has a number of diagnostic epipedons, which do not occur in other systems, namely, the warpic (finely stratified epipedon resulting from long-continued irrigation), cumulic (thick

TABLE 32.11 Order, Suborder, and Groups of the Chinese Soil Taxonomy

Order	Suborder	Group
Histosols	Permagelic Histosols	Foli-Permagelic Histosols
		Fibri-Permagelic Histosols
		Hemi-Permagelic Histosols
	Orthic Histosols	Foli-Orthic Histosols
		Fibri-Orthic Histosols
		Hemi-Orthic Histosols
		Sapri-Orthic Histosols
Anthrosols	Stagnic Anthrosols	Gleyi-Stagnic Anthrosols
		Fe-Leachi-Stagnic Anthrosols
		Fe-Accumuli-Stagnic Anthrosols
		Hapli-Stagnic Anthrosols
	Orthic Anthrosols	Fimi-Orthic Anthrosols
		Siltigi-Orthic Anthrosols
		Mud-Cumuli-Orthic Anthrosols
		Earth-Cumuli-Orthic Anthrosols
Spodosols	Humic Spodosols	Hapli-Humic Spodosols
	Orthic Spodosols	Hapli-Orthic Spodosols
Andosols	Cryic Andosols	Geli-Cryic Andosols
		Hapli-Cryic Andosols
	Vitric Andosols	Usti-Vitric Andosols
		Udi-Vitric Andosols
	Udic Andosols	Humi-Udic Andosols
		Hapli-Udic Andosols
Ferralosols	Udic Ferralosols	Rhodi-Udic Ferralosols
		Xanthi-Udic Ferralosols
		Hapli-Udic Ferralosols
Vertosols	Aquic Vertosols	Calci-Aquic Vertosols
		Hapli-Aquic Vertosols
	Ustic Vertosols	Calci-Ustic Vertosols
		Hapli-Ustic Vertosols
	Udic Vertosols	Humi-Udic Vertosols
		Calci-Udic Vertosols
		Hapli-Udic Vertosols
Aridosols	Cryic Aridosols	Calci-Cryic Aridosols
		Gypsi-Cryic Aridosols
		Argi-Cryic Aridosols
		Hapli-Cryic Aridosols
	Orthic Aridosols	Calci-Orthic Aridosols
		Sali-Orthic Aridosols
		Gypsi-Orthic Aridosols
		Argi-Orthic Aridosols
		Hapli-Orthic Aridosols
Halosols	Alkalic Halosols	Takyri-Alkalic Halosols
		Aqui-Alkalic Halosols
		Hapli-Alkalic Halosols
	Orthic Halosols	Aridi-Orthic Halosols
		Aqui-Orthic Halosols
Gleyosols	Permagelic Gleyosols	Histi-Permagelic Gleyosols
		Hapli-Permagelic Gleyosols
	Stagnic Gleyosols	Histi-Stagnic Gleyosols
		Hapli-Stagnic Gleyosols

TABLE 32.11 (continued) Order, Suborder, and Groups of the Chinese Soil Taxonomy

Order	Suborder	Group
	Orthic Gleyosols	Histi-Orthic Gleyosols
		Molli-Orthic Gleyosols
		Hapli-Orthic Gleyosols
Isohumosols	Lithomorphic Isohumosols	Phosphi-Lithomorphic Isohumosols
		Black-Lithomorphic Isohumosols
	Ustic Isohumosols	Cryi-Ustic Isohumosols
		Cumuli-Ustic Isohumosols
		Pachi-Ustic Isohumosols
		Calci-Ustic Isohumosols
		Hapli-Ustic Isohumosols
	Udic Isohumosols	Stagni-Udic Isohumosols
		Argi-Udic Isohumosols
		Hapli-Udic Isohumosols
Ferrosols	Ustic Ferrosols	Argi-Ustic Ferrosols
		Hapli-Ustic Ferrosols
	Perudic Ferrosols	Carbonati-Perudic Ferrosols
		Alliti-Perudic Ferrosols
		Hapli-Perudic Ferrosols
	Udic Ferrosols	Carbonati-Udic Ferrosols
		Hi-Weatheri-Udic Ferrosols
		Alliti-Udic Ferrosols
		Argi-Udic Ferrosols
		Hapli-Udic Ferrosols
Argosols	Boric Argosols	Albi-Boric Argosols
		Molli-Boric Argosols
		Hapli-Boric Argosols
	Ustic Argosols	Carbonati-Ustic Argosols
		Calci-Ustic Argosols
		Ferri-Ustic Argosols
		Hapli-Ustic Argosols
	Perudic Argosols	Carbonati-Perudic Argosols
		Ali-Perudic Argosols
		Hapli-Perudic Argosols
	Udic Argosols	Albi-Udic Argosols
		Carbonati-Udic Argosols
		Claypani-Udic Argosols
		Ali-Udic Argosols
		Acidi-Udic Argosols
		Ferri-Udic Argosols
		Hapli-Udic Argosols
Cambosol	Gelic Cambosol	Permi-Gelic Cambosol
		Aqui-Gelic Cambosols
		Matti-Gelic Cambosols
		Molli-Gelic Cambosols
		Umbri-Gelic Cambosols
		Hapli-Gelic Cambosols
	Aquic Cambosols	Litteri-Aquic Cambosols
		Shajiang-Aquic Cambosols
		Dark-Aquic Cambosols
		Ochri-Aquic Cambosols

Order	Suborder	Group
	Ustic Cambosols	Siltigi-Ustic Cambosols
		Ferri-Ustic Cambosols
		Endorusti-Ustic Cambosols
		Molli-Ustic Cambosols
		Hapli-Ustic Cambosols
	Perudic Cambosols	Bori-Perudic Cambosols
		Stagni-Perudic Cambosols
		Carbonati-Perudic Cambosols
		Ali-Perudic Cambosols
		Acidi-Perudic Cambosols
		Hapli-Perudic Cambosols
	Udic Cambosols	Bori-Udic Cambosols
		Carbonati-Udic Cambosols
		Purpli-Udic Cambosols
		Ali-Udic Cambosols
		Ferri-Udic Cambosols
		Acidi-Udic Cambosols
		Hapli-Udic Cambosols
Primosols	Anthric primosols	Turbi-Anthric Primosols
		Silti-Anthric Primosols
	Sandic primosols	Geli-Sandic Primosols
		Aqui-Sandic Primosols
		Aridi-Sandic Primosols
		Usti-Sandic Primosols
		Udi-Sandic Primosols
	Alluvic primosols	Geli-Alluvic Primosols
		Aqui-Alluvic Primosols
		Aridi-Alluvic Primosols
		Usti-Alluvic Primosols
		Udi-Alluvic Primosols
	Orthic primosols	Loessi-Orthic Primosols
		Purpli-Orthic Primosols
		Rougi-Orthic Primosols
		Geli-Orthic Primosols
		Aridi-Orthic Primosols
		Usti-Orthic Primosols
		Udi-Orthic Primosols

Source: Gong, Z.T., and G.L. Zhang. 2007. Pedogenesis and soil taxonomy. Science Press, Beijing, China (in Chinese).

epipedon resulting from additions of manure and organic-rich soil material), mellowic (epipedon resulting from long and intensive cultivation and applying night soil, organic trash, and manure), and hydragric (epipedon resulting from cultivating the soil under wet conditions) epipedons.

The 10 diagnostic subsurface horizons are the albic (white), weathering B, humilluvic (illuvial organic matter), argillic (clay illuviated), clayific (in situ clay accumulation through weathering of primary minerals), claypan (slowly permeable, clay-rich horizon), alkalic (clay-illuviated horizon with high Na percentage),

spodic (illuviation of amorphous organic compounds in combination with Al and/or Fe), agric (illuviated humus clay or humus silt clay under conditions of cultivation), and hydragric–redoxic (oxidation–reduction horizon related to wet cultivation) horizons.

The other diagnostic horizons comprise the calcic, hypercalcic, gypsic, hypergypsic, salic, hypersalic, sulfuric, phosphic, and gleyic horizons and the calci-, gypsi-, and salipans.

Twenty-three diagnostic characteristics (characters, properties, features, contacts, saturation, materials, regimes, or layers) are defined. These are lithologic characters (features inherited from the parent material), vertic features (shrink/swell phenomena), desertic features (surface and topsoil features related to arid conditions), takyric features (cracking of surface soil in arid conditions induced by wetting and drying), redoxic features (hydromorphism), frost/thaw features, regosolic features (<35% coarse fragments), skeletic features (between 35% and 70% coarse fragments), lithic features (>70% coarse fragments), lithic contact, paralithic contact, siallic properties (weathering B horizon of which the clay fraction is dominated, 2:1 or 2:1:1 clays), fersiallic properties (weathering B horizon of which the clay fraction is dominated 2:1 or 2:1:1 clays and which has ≥2% free Fe oxides), ferralic properties (weathering B horizon of which the clay fraction is dominated 1:1 clays and which has ≥2% free Fe oxides), andic properties (related to weathering of pyroclastic deposits), base saturation, Al saturation, calcaric properties (≥1% $CaCO_3$), humic properties (high organic matter content), organic soil materials, soil moisture regimes, soil temperature regimes, and permafrost layer (perennial temperature at or below 0°C).

When important and obvious properties of soil horizon do not fulfill the combination required for a diagnostic horizon, the CSTC uses the term evidence. They are used to identify soil taxa, particularly at subgroup level. Twelve diagnostic evidences are defined so far, namely, histic, warpic, cumulic, mellowic, clayific (all meeting the requirements of the related diagnostic horizons apart from thickness), alkalic (lower ESP), spodic (lower ratio of pyrophosphate extractable Al + Fe to clay), calcic (either shallower thickness or lower $CaCO_3$ content), gypsic (either shallower thickness or lower gypsum content), salic (lower salt content), gleyic (discontinuous gleyic parts), and vertic (less wider cracks and a clayey subhorizon within 50 cm depth) evidences.

32.8 The Australian Soil Classification

32.8.1 Introduction

In 1996, a new soil classification system for Australia was issued (Isbell, 1996). Until then, two different soil classification systems were widely used (Stace et al., 1968; Northcote, 1979). A soil classification committee has worked over 15 years to develop the new system, taking advantage of the previously used systems, the numerous soil surveys carried out all over the country, and the large database containing data of some 14,000 soil profiles, which was constructed during the time of the committee's work.

32.8.2 Structure and Nomenclature

The new system is multicategorical, comprising orders, suborders, great groups, subgroups, and families. In total, 14 orders are recognized (Anthroposols, Calcarosols, Chromosols, Dermosols, Ferrosols, Hydrosols, Kandosols, Kurosols, Organosols, Podosols, Rudosols, Sodosols, Tenosols, and Vertosols), distinguished by fairly straightforward criteria, which can easily be recognized in the field. Nomenclature is often based on Latin or Greek roots, as with many other modern soil classification systems, but the names are clearly different from *Soil Taxonomy*, the *FAO–UNESCO Legend*, or the *World Reference Base for Soil Resources*, in order to avoid confusion.

Unlike many other soil classification systems, there is no fixed soil depth for classification purposes in the new Australian system. The concept of pedologic organization (McDonald et al., 1990) is used to define the soil to be classified. Isbell (1996) describes this as "… a broad concept used to include all changes in soil material resulting from the effect of the physical, chemical and biological processes that are involved in soil formation." Consequently, soil studies for classification purposes may need to go to considerable depth. At family level, this is recognized since one of the family criteria, soil depth, includes a class Giant for those soils that are deeper than 5 m.

Division of seven orders (Chromosols, Dermosols, Ferrosols, Kandosols, Kurosols, Sodosols, and Vertosols) into suborders is based on color criteria (e.g., red, brown, yellow, gray, and black). Suborders of Anthroposols are distinguished on the kind and nature of the human activity, in Calcarosols, the degree and kind (shells, gypsum, $CaCO_3$) of the accumulation are used for the subdivision, suborders of the Hydrosols reflect the type of tidal inundation, the degree of salinity, and the reduction–oxidation regime, Organosols are classically divided according to the degree of decomposition of the organic material, suborders of the Podosols reflect the degree of wetness, Rudosols are divided according the nature of the material, while, in Tenosols, the type of surface horizon, the occurrence of a bleached horizon, and the presence or absence of a weakly developed B horizon are used to characterize the suborders.

Subdivision of the great groups and subgroups is still incomplete for a number of suborders. For those suborders that are developed into great groups and subgroups, division is generally made on the presence or absence of important diagnostic horizons and materials and on the degree of importance of certain characteristics.

Criteria considered at family level are A horizon thickness (not in Organosols, Rudosols, and Vertosols), gravel content of the surface and A1 horizon (not in Organosols), A1 horizon texture (not in Organosols and Vertosols), B horizon maximum texture (not in Organosols and Rudosols), soil depth (not in Organosols), thickness of the soils above the upper boundary of the Bk horizon (in Calcarosols only), nature of the uppermost organic materials and cumulative thickness of the organic materials (in Organosols only), and clay content of the upper 0.1 m (in Vertosols only).

The full soil name is constructed as follows: subgroup, great group, suborder, order, and family. A unique coding system for all levels is provided for recording the classification of soil profiles. Included in this coding system is the opportunity to indicate a confidence level to the soil classification.

32.8.3 Description of the Orders

The 14 orders now recognized in the Australian soil classification, and their characteristics are as follows, in alphabetical order:

Anthroposols (Gr. anthropos, man): soils resulting from human activities

Calcarosols (L. calcis, lime): soils having pedogenetic carbonate accumulations throughout the solum or at least directly below the A1 or Ap horizon

Chromosols (Gr. chroma, color): soils having a clear or abrupt textural B horizon in which the major part of the upper 0.2 m of the B2 horizon is not strongly acid $[pH_{H_2O} (1:5) \geq 5.5$ and $pH_{CaCl_2} (1:5) \geq 4.6]$

Dermosols (L. dermis, skin): soils having a B2 horizon, which has a structure more developed than weak throughout the major part of the horizon

Ferrosols (L. ferrum, iron): soils having a B2 horizon in which the major part contains >5% free Fe oxide in the fine earth fraction

Hydrosols (Gr. hydor, water): soils that are saturated with water in the major part of the solum for at least 2–3 months in most years. Reducing conditions are not an essential requirement

Kandosols (kandite, 1:1 clay minerals): soils having a well-developed B2 horizon of which the major part is massive or has only a weak grade of structure, and having a maximum clay content in some part of the B2 horizon >15%

Kurosols (no root): soils having a clear or abrupt textural B horizon in which the major part of the upper 0.2 m of the B2 horizon is strongly acid $[pH_{H_2O} (1:5) < 5.5$ and $pH_{CaCl_2} (1:5) < 4.6]$

Organosols (no root): soils having more than 0.4 m of organic materials within the upper 0.8 m or having organic materials from the surface to a minimum depth of 0.1 m overlying either hard rock or other hard layers, partially weathered or decomposed rock or fragmental material with interstices (partially) filled

Podosols (R. pod, under, and zola, ash): soils having a Bs, Bhs, or Bh horizon (horizons with illuvial accumulations of amorphous organic matter Al and Al silica complexes, with or without Fe)

Rudosols (L. rudimentum, a beginning): soils having negligible (or rudimentary) pedological organization apart from the minimal development of an A1 horizon or the presence of less than 10% of B horizon material in fissures in the parent rock or saprolite

Sodosols (E. sodium): soils having a clear or abrupt textural B horizon, in which the major part of the upper

0.2 m of the B2 horizon is sodic (ESP of the fine earth
soil material ≥6)

Tenosols (L. tenius, weak, slight): soils having weak expres-
sion of pedological organization

Vertosols (L. vertere, to turn): soils having ≥35% clay and
developing open cracks, which are at least 5 mm wide
and extend upward to the surface or to the base plow
layer at some time in most years self-mulching hori-
zon, or a thin, surface crust, and having slickensides or
lenticular peds

32.9 The German Soil Classification

32.9.1 Introduction

The latest version of the German soil classification system was
published in 1998 (Bodensystematik, 1998) and updated in 2005
(Ad-hoc-AG Boden, 2005). It is based on the system laid down by
Kubiëna (1953) and Mückenhausen (1962). Older concepts from
the nineteenth and the beginning of the twentieth century using
different approaches will not be discussed in this book. During
recent history, West Germany and East Germany had slightly
differing systems, which were brought together again in 1994.
Responsible for the German classification system is a Working
Group of the German Soil Science Society.

32.9.2 Structure and Nomenclature

The German system is hierarchical with 6 categorical levels
(English terms according to Wittmann, 1997): 4 orders, 21 sub-
orders, 56 types, and 230 subtypes are defined. For the lowest
levels, the varieties and subvarieties, additional (often quantita-
tive) attributes are provided that have to be added to the sub-
type name according to specific rules. For common discussions
among soil scientists, the type or subtype level is used. A key is
available to classify soils down to the type level. The German
system is designed as a natural system, whose structure reflects
the mutual relationships of the units and the inherent architec-
ture of the whole (Kubiëna, 1953).

In addition to classify soils, a system to classify substrates is
provided characterizing texture, coarse fraction, carbonates,
coal, lithogenic carbon, and consolidated and unconsolidated
rock. Usually, soil and substrate classification are given together.

The orders are defined by the water regime: The Terrestrial
soils are dominated by a descending water regime and form
by far the largest order. The Semiterrestrial soils are character-
ized by groundwater influence starting within the first 40 cm
or by groundwater fluctuations over greater depth ranges. The
Semisubhydric and Subhydric soils comprise the soils in tidal
regions and the soils covered by water not exceeding 2 m. The
last order is formed by fens and bogs.

The suborders reflect the stage of soil formation and the
degree of horizon differentiation. They will be described later.

The types are differentiated according to horizons and hori-
zon sequences reflecting specific pedogenic processes and the

properties resulting from these processes (morphogenetic
approach). The allocation of soils requires the knowledge of all
major soil characteristics and the full horizon sequence. Specific
rules apply to fens and bogs and to poorly developed soils,
where characteristics of the parent material are included in the
definition.

Three kinds of subtypes are recognized: The norm sub-
types, the deviation subtypes, and the transition subtypes.
A transition subtype shows some properties of another type.
The names of both types are combined using the name of the
dominant type last. A deviation subtype has some additional
characteristics that are not part of the definition of the type
but fit well within the type. The additional characteristics
are indicated by a prefix added to the name of the soil type.
Finally, the norm subtype lacks both, transition features and
additional features.

Types and subtypes are defined by horizon sequences. A hori-
zon denomination consists of a capital letter as master symbol
followed by one or more lower case letters indicating proper-
ties resulting from soil-forming processes. Lower case letters set
before the master symbol describe geogenic and anthropogenic
characteristics. Most soil characteristics required for the hori-
zon denomination can be recognized in the field.

32.9.3 Description of the Suborders

The 21 suborders recognized in the German soil classification
system are described briefly:

Order Terrestrial Soils

O/C-Böden: soils having an organic horizon overlying
directly continuous rock or coarse material, the inter-
stices of which are filled with organic material

Terrestrische Rohböden: soils having an initial accumula-
tion of organic matter and an initial chemical weather-
ing of the geological parent material

Ah/C-Böden: soils having a fully developed mineral
topsoil horizon enriched with organic matter and
some chemical weathering of the geological parent
material

Schwarzerden: soils having a thick dark mineral topsoil
horizon rich in organic matter with high base satura-
tion and intensive bioturbation and not having a pro-
nounced subsoil horizon

Pelosole: soils originating from parent materials rich in
swelling clays and having a well-developed aggregate
structure without pronounced chemical weathering

Braunerden: soils having a well-developed subsoil horizon
with neoformation of oxides and clay minerals

Lessivés: soils having an accumulation of illuviated clay in
the subsoil

Podsole: soils having an accumulation of illuviated Fe, Al,
and organic matter in the subsoil

Terrae Calcis: soils having a residual accumulation of large
amounts of clay and oxides after limestone weathering

Fersiallitische und Ferrallitische Paläoböden: soils having a kaolinite-rich subsoil; in Germany only as relics from old (mostly Tertiary) weathering

Stauwasserböden: soils having a bleached topsoil or bleached aggregate surfaces and an Fe oxide accumulation in the interiors of the aggregates (if aggregates are present) due to redox processes caused by temporal water stagnation

Reduktosole: soils having a redistribution of Fe oxides due to redox processes caused by reducing gases like CO_2 or CH_4 (which maybe natural or anthropogenic)

Terrestrische Anthropogene Böden: soils having been changed by human activity (like deep cultivation or transportation of natural soil material) to such an extent that the natural horizon sequence is no longer present

Order Semiterrestrial Soils

Auenböden: soils originating from alluvial sediments that are periodically or episodically flooded and that have groundwater fluctuations over greater depth ranges

Gleye: soils having bleached horizons with Fe oxide accumulations on aggregate surfaces (if aggregates are present) or at root channels due to redox processes caused by groundwater that at least temporally reaches a depth above 40 cm below the mineral soil surface

Marschen: soils originating from tidal sediments and having groundwater influence

Strandböden: soils originating from sandy tidal sediments that are periodically or episodically flooded and show an initial accumulation of organic matter

Order Semisubhydric and Subhydric Soils

Semisubhydrische Böden: soils affected by regular tidal fluctuation

Subhydrische Böden: underwater soils covered by water not exceeding 2 m

Order Fens and Bogs

Naturnahe Moore: natural or only slightly altered fens and bogs

Erd- und Mulmmoore: drained and used fens and bogs showing (especially structural) changes toward a terrestrial soil formation

32.10 Classification Systems of Brazil, Canada, England and Wales, New Zealand, and South Africa

32.10.1 Introduction

The soil classification systems of Brazil, Canada, England and Wales, New Zealand, and South Africa have in common that they are developed in one, or at the most, two major world ecological regions, and therefore, focus on the characteristics important in that particular region. The Brazilian soil classification specializes in tropical soils, the Canadian system of soil classification focuses mainly on soils of the (sub)arctic and drier boreal regions, that of England and Wales aims at classifying soil conditions characteristic of the humid temperate regions, as does that of New Zealand, but with focus on volcanic regions, while the South African system is designed to classify soils in dry subtropical areas with emphasis on land use and management aspects.

32.10.2 The Brazilian System

Work on soil inventory of Brazil started in 1947 (Jacomine and Camargo, 1996). However, detailed descriptions and building of classification system started after 1964. The construction started "from top to bottom," filling the taxa with empirical content and developing lower taxonomic levels. Since the 1960s, the classification has gradually improved and became more descriptive and detailed (Jacomine and Camargo, 1996). The latest available version was published in 1999 (EMBRAPA, 1999). The structure and some diagnostic elements are very similar to the U.S. *Soil Taxonomy* (Soil Survey Staff, 1996); however, moisture and temperature regimes are not included as diagnostic criteria in the upper levels of the Brazilian taxonomy. The nomenclature, especially the names of higher categories, is closer to that of the WRB (IUSS Working Group WRB, 2006).

Brazilian soils are divided into 14 orders, at the first level. There are 150 great groups divided into 580 subgroups, which are defined according to qualitative variations in soil properties. The orders are defined as follows:

Alissolos: soils with evident active clay-enriched B horizon with high concentration of exchangeable aluminum

Argissolos: soils with B horizon enriched with low-activity (kaolinitic) clay

Cambissolos: soils with moderate alteration of the structure or color of the B horizon

Chernossolos: soils with dark humus-enriched topsoil horizon

Espodossolos: soils having evidences of iron and aluminum migration in organic complexes

Gleissollos: excessively humid soils affected with high groundwater level

Latossolos: deeply weathered soils

Luvissolos: soils with evident active clay-enriched B horizon with high base saturation

Neossolos: soils with moderate development

Nitososolos: deeply weathered clayey soils with well-developed aggregates with shiny surfaces

Organossolos: soils formed in organic parent material (peat)

Planossolos: soils with clay-depleted topsoil, where water stagnation causes prolonged reductive conditions

Plintossolos: soils having a discrete or continuous layer cemented with iron (hydr)oxides, formed by alternating reductive and oxidative conditions around the groundwater limit

Vertissolos: clayey soils with predominantly swelling clays, which expand in humid periods and shrink in dry seasons

32.10.3 The Canadian System

The Canadian system (Canada Soil Survey Committee, Subcommittee on Soil Classification, 1978) comprises a number of taxa, which are defined on the basis of soil properties; however, the system is genetically biased with respect to the definition of the higher categories. The system is differentiated as follows: orders, suborders, great groups, subgroups, families, and series. The following nine orders are recognized as follows:

Brunosolic: weak to moderately developed soils

Chernozemic: soils with a dark-colored, high base-saturated (dominantly Ca) A horizon, which is at least 10 cm thick and which have an organic carbon content between 1% and 17% (depending on clay content)

Cryosolic: soils that have permafrost within 1 m of the surface, or 2 m if they are strongly cryoturbated

Gleysolic: soils permanently or temporarily saturated with water and which experience reducing conditions

Luvisolic: soils with a subsurface horizon showing evidence of clay accumulation

Organic: soils dominated by organic horizons (horizons >17% organic C)

Podzolic: soils having a podzolic B horizon (horizon with accumulation of amorphous materials)

Regosolic: soils showing little or no soil development

Solonetzic: soils having a solonetzic B horizon (subsurface horizon with prismatic or columnar structures and a Ca/Na ratio ≥ 10)

32.10.4 The Soil Classification System of England and Wales

The differentiating criteria, which build the soil classification for England and Wales (Avery, 1980), comprise the composition of the soil material, the presence or absence of particular horizons or sequences of horizons, or other specified differentiating criteria. Ten MSGs are recognized, which are further subdivided in soil groups and soil subgroups. The 10 MSGs are as follows:

Brown soils: soils having a weathered (cambic) or argillic B horizon

Groundwater gley soils: soils influenced by the presence of a shallow fluctuating groundwater table

Lithomorphic soils: shallow soils (mineral substratum starting within 40 cm depth) with a distinct, humose, or peaty topsoil

Manmade soils: soils having a thick manmade A horizon or which are profoundly disturbed to depths exceeding 40 cm

Peat soils: soils with >40 cm organic material, or >30 cm if directly overlying bedrock, and lacking overlying non-humose mineral horizons extending below 30 cm

Pelosols: slowly permeable clayey soils

Podzolic soils: soils having a podzolic B horizon (subsurface horizon with accumulations of organic matter and Al, Fe, or both)

Raw gley soils: soils with a gleyed subsurface horizon lacking a distinct topsoil

Surface-water gley soils: soils having a gleyed subsurface horizon, which can be attributed to saturation by surface water

Terrestrial raw soils: soils consisting of little altered mineral material having no diagnostic surface or subsurface horizons

32.10.5 The New Zealand System

The newly published New Zealand soil classification (Hewitt, 1998) comprises 15 orders, which are subdivided into groups and subgroups. They are distinguished from each other by the presence or absence of diagnostic horizons and other differentia, some of which are unique to New Zealand. An important element in the classification is a listing of accessory properties of the order, some of which are directly related to soil management and conservation. The 15 orders are as follows:

Allophanic soils: soils strongly influenced by minerals with short-range order (especially, allophane, imogolite, and ferrihydrite)

Anthropic soils: soils formed by the direct action of people

Brown soils: soils having a weathered B, argillic, or cutanic horizon (horizon containing translocated material, which, however, fails to meet the requirements for argillic or Bh horizon), generally having a low base status

Gley soils: poorly and very poorly drained soils

Granular soils: clayey soils dominated by kaolinite-group minerals

Melanic soils: soils with highly base-saturated, well-structured, very dark A horizons

Organic soils: soils occurring in partly decomposed remains of wetland plants or forest litter

Oxidic soils: soils containing low-activity clays and secondary oxides, which give rise to variable charge properties

Pallic soils: soils with moderate to high base status and low contents of secondary Fe oxides

Podzols: acid soils with low base saturation and subsurface accumulation of organo-Al complexes, with or without Fe

Pumice soils: soils having properties dominated by a pumiceous and glassy skeleton with a low content of clay

Raw soils: soils lacking distinct topsoil development

Recent soils: soils showing only incipient marks of soil-forming processes

Semiarid soils: soils having a high base status and soil water deficit during the growing season

Ultic soils: acid soils with clayey and/or organic illuvial features

32.10.6 The South African System

The South African taxonomic system of soil classification (Soil Classification Working Group, 1991) comprises only two categories, soil forms and soil families. Soil forms are defined by a unique vertical sequence of diagnostic horizons and materials, while the soil families are separated on the basis of other properties. Nomenclature is South African in the sense that both soil forms and soil families bear local, geographical names (e.g., Escourt form, Haarlem family). So far, 73 soil forms are defined, divided into 406 soil families. The definitions of the diagnostic horizons and materials fit the unique South African conditions and are very difficult to correlate with other existing classification systems. Important elements for classification of the soil forms are the various topsoil horizons (organic O, humic A, vertic A, melanic A, and orthic A horizons), the pedality (apedal, structural grade weaker than moderate, vs. structured B horizons), various clay-enriched horizons (prismacutanic, pedocutanic, lithocutanic, and neocutanic B horizons), signs of wetness (e.g., soft and hard plinthic B horizons), carbonate accumulation (neocarbonate B horizon, soft carbonate horizon, and hardpan carbonate horizon), and nature of the underlying material (regic sand, stratified alluvium, saprolite, hard rock). Distinguishing criteria at soil family level comprise the thickness, colors, base status structure, consistency, and wetness of horizons.

References

Ad-hoc-AG Boden. 2005. Bodenkundliche Kartieranleitung, herausgegeben von der Bundesanstalt für Geowissenschaften und Rohstoffe und den Geologischen Landesämtern in der Bundesrepublik Deutschland, Hannover, 5 Auflage.

AFES–INRA (Association Française pour l'Etude du Sol) (Institut National de la Recherche Agronomique). 1990. Référentiel Pédologique Français. 3éme proposition. AFES, Plaisir, France.

AFES–INRA (Association Française pour l'Etude du Sol) (Institut National de la Recherche Agronomiqie). 1992. Référentiel Pédologique, principaux sols d'Europe. INRA, Paris, France.

Arnold, R.W. (ed.). 2001. Russian soil classification system. Smolensk, Oykumena, Russia.

Asiamah, R.D., J.K. Senayah, T. Adjei-Gyapong, and O.C. Spaargaren. 1997. Ethno-pedology surveys in the semi-arid savanna zone of northern Ghana. Soil Research Institute, Kwadaso-Kumasi, Ghana and ISRIC, Wageningen, the Netherlands.

Aubert, G., and P. Duchaufour. 1956. Projet de classification des sols. *In* Rapp. VIᵉ Congr. Int. Sc. Sol. 6:597–604.

Avery, B.W. 1980. Soil classification for England and Wales. Technical monograph no. 14. Soil Survey for England and Wales, Harpenden, U.K.

Bodensystematik, A.K. 1998. Systematik der Böden und der bodenbildenden Substrate Deutschlands, herausgegeben vom Arbeitskreis für Bodensystematik der Deutschen Bodenkundlichen Gesellschaft. Mitteilungen der Deutschen Bodenkundlichen Gesellschaft, Band 86.

Canada Soil Survey Committee, Subcommittee on Soil Classification. 1978. The Canadian system of soil classification. Canada Department of Agriculture Publication No. 1646. Supply and Services Canada, Ottawa, Canada.

CPCS (Commission de Pédologie et de Cartographie des Sols). 1967. Classification de sols. Mimeogr. Thivernal-Grignon, France.

Dudal, R. 1990. Progress in IRB preparation, p. 69–70. *In* B.G. Rozanov (ed.) Soil classification, Reports of the International Conference on Soil Classification. Centre for International Projects, USSR State Committee for Environmental Protection, Moscow, Russia.

EMBRAPA. 1999. Sistema Brasileiro de Clasificação de Solos. Embrapa Produção de Informação, Brasília—Embrapa Solos. Rio de Janeiro, Brazil.

FAO–ISRIC–ISSS. 1998. World reference base for soil resources. World Soil Resources Report No. 84. Food and Agriculture Organization, Rome, Italy. Available at: http://www.fao.org/docrep/W8594E/W8594E00.htm

FAO–UNESCO. 1974. Soil map of the world 1:5,000,000. Volume I, Legend. United Nations Educational, Scientific and Cultural Organization, Paris, France.

FAO–UNESCO. 1988. FAO–UNESCO soil map of the world, revised legend. World Soil Resources Report No. 60. Food and Agriculture Organization, Rome, Italy.

Gong, Z. 1994. Chinese soil taxonomic classification (first proposal). Institute of Soil Science. Academia Sinica, Nanjing, China.

Gong, Z.T., and G.L. Zhang. 2007. Pedogenesis and soil taxonomy. Science Press, Beijing, China (in Chinese).

Hewitt, A.E. 1998. New Zealand soil classification. Landcare research science series 1. Manaaki Whenua Press, Lincoln, New Zealand.

Isbell, R.F. 1996. The Australian soil classification. Australian soil and land survey handbook. CSIRO, Collingwood, Australia.

ISS-AS (Institute of Soil Science-Academia Sinica). 1990. Soils of China. Science Press, Beijing, China.

ISSS–ISRIC–FAO. 1994. World reference base for soil resources. Draft. Compiled and edited by O.C. Spaargaren. International Society of Soil Science, International Soil Reference and Information Centre and the Food and Agriculture Organization of the United Nations, Wageningen, the Netherlands.

IUSS Working Group WRB. 2006. World reference base for soil resources 2006. World Soil Resources Reports No. 103. FAO, Rome, Italy. Available online at http://www.fao.org/nr/land/soils/soil/wrb-documents/jp/

IUSS Working Group WRB. 2007. World reference base for soil resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome, Italy. Available online at http://www.fao.org/nr/land/soils/soil/wrb-documents/en/

IUSS Working Group WRB. 2010. Guidelines for constructing small-scale map legends using the world reference base for soil resources. Addendum to the world reference base for soil resources. FAO, Rome, Italy. Available online at http://www.fao.org/nr/land/soils/soil/wrb-documents/en/

Jacomine, P.T.K., and M.N. Camargo. 1996. Classificaçao pedologica nacional em vigor', p. 675–689. *In* V.H. Alvarez, L.E.F. Fontes, and M.P.F. Fontes (eds.) O solo nos grandes dominios morfoclimaticos do Brasil o e desenvolvimento sustentado. SBCS–UFV, Viçosa, Brasil.

Krasilnikov, P.V., J.J. Ibáñez Martí, R.W. Arnold, and S.A. Shoba. 2009. Handbook of soil terminology, correlation and classification. Earthscan Publications, London, U.K.

Kubiëna, W.L. 1953. Bestimmungsbuch und Systematik der Böden Europas. Verlag Enke, Stuttgart, Germany.

McDonald, R.C., R.F. Isbell, J.G. Speight, J. Walker, and M.S. Hopkins. 1990. Australian soil and land survey field handbook. 2nd edn. Inkata Press, Melbourne, Australia.

Mückenhausen, E. 1962. Entstehung, Eigenschaften und Systematik der Böden der Bundesrepublik Deutschland. DLG-Verlag, Frankfurt, Germany.

Northcote, K.H. 1979. A factual key for the recognition of Australian soils. 4th edn. Rellim Technical Publications, Glenside, South Australia.

Shishov, L.L., and I.A. Sokolov. 1990. Genetic classification of soils in the USSR, p. 77–93. *In* B.G. Rozanov (ed.) Soil classification. Reports of the international conference on soil classification. USSR State Committee for Environmental Protection, Moscow, Russia.

Shishov, L.L., V.D. Tonkonogov, I.I. Lebedeva, and M.I. Gerasimova. 2004. Classification and diagnostics of soils of Russia. Oykumena, Smolensk, Russia (in Russian).

Soil Classification Working Group. 1991. Soil classification. A taxonomic system for South Africa. Memoirs on the agricultural natural resources of South Africa 15. Department of Agricultural Development, Pretoria, South Africa.

Soil Survey Staff. 1975. Soil taxonomy. A basic system of soil classification for making and interpreting soil surveys. Agriculture handbook. SCS–USDA, Washington, DC.

Soil Survey Staff. 1996. Keys to soil taxonomy. 7th edn. USDA–NRCS, Washington, DC.

Stace, H.C.T., G.D. Hubble, R. Brewer, K.H. Northcote, J.R. Sleeman, M.J. Mulcahy, and E.G. Hallsworth. 1968. A handbook of Australian soils. Rellim Technical Publications, Glenside, South Australia.

VASKhNIL (V.V. Dokuchaev Institute of Soil Science). 1986. Classification and diagnostics of soils of the USSR. Translated from Russian by S. Viswanathan. Amerind Publishing Co., New Delhi, India.

Wittmann, O. 1997. Soil classification of the Federal Republic of Germany. Mitteilungen der Deutschen Bodenkundlichen Gesellschaft, Band 84.

33

Classification of Soils

Olafur Arnalds
Agricultural University of Iceland

Fredrich H. Beinroth
University of Puerto Rico

J.C. Bell
University of Minnesota

J.G. Bockheim
University of Wisconsin-Madison

Janis L. Boettinger
Utah State University

M.E. Collins
University of Florida

R.G. Darmody
University of Illinois

Steven G. Driese
Baylor University

Hari Eswaran
United States Department of Agriculture

Delvin S. Fanning
University of Maryland

D.P. Franzmeier
Purdue University

C.T. Hallmark
Texas A&M University

Willie Harris
University of Florida

Wayne H. Hudnall
Texas Tech University

Randall K. Kolka
United States Department of Agriculture-Forest Service

David J. Lowe
University of Waikato

Paul A. McDaniel
University of Idaho

D.G. McGahan
Tarleton State University

H. Curtis Monger
New Mexico State University

Lee C. Nordt
Baylor University

Chien-Lu Ping
University of Alaska

Martin C. Rabenhorst
University of Maryland

Paul F. Reich
United States Department of Agriculture

Randall Schaetzl
Michigan State University

Joey N. Shaw
Auburn University

Christopher W. Smith
United States Department of Agriculture

Randal J. Southard
University of California

David Swanson
United States Department of Agriculture

C. Tarnocai
Agriculture and Agri-Food Canada

Goro Uehara
University of Hawaii at Manoa

Larry T. West
United States Department of Agriculture

Larry P. Wilding
Texas A&M University

33.1 Introduction: General Characteristics of Soil Orders and Global Distributions

Larry P. Wilding

Larry T. West

The purpose of this chapter is to present a discussion of the distribution, attributes, pedogenesis, classification, use, and management of the 12 soil orders recognized by *Soil Taxonomy* (Soil Survey Staff, "1999"). In this chapter, soil orders have been arrayed in a sequence to consider the organic soils first, followed next by soils that are weakly expressed or lack subsoil diagnostic horizons, and closing with soils exhibiting the greatest evidence of pedogenic weathering and/or accumulation of sesquioxides and translocated phyllosilicates. Only general statements can be made at this categorical level, but the information content becomes more precise as one descends to lower taxa (Sections 33.2 through 33.13). Because *Soil Taxonomy* is a dynamic classification system, the content is subject to change. For example, the most recent edition of the *Keys to Soil Taxonomy* (Soil Survey Staff, 2010) includes newly defined taxa for subaqueous soils (soils permanently inundated by shallow water) in the Histosols and Entisols (Sections 33.2 and 33.4).

Because of space limitations, data sets verifying the physical, chemical, mineralogical, and biological attributes of given orders are abridged. However, access to complete soil characterization databases, engineering properties, pedon descriptions, soil interpretations, and other information for different soil taxa are available online at the following address: http://soils.usda.gov/. For example, analytical data for more than 27,000 pedons in the United States and 1,100 pedons from other countries can

be accessed from this Web site. Standard morphological pedon descriptions are also available in most cases. Further, this Web site includes databases with listings of published soil surveys in the United States, Puerto Rico, Virgin Islands, Trust Territories, and some foreign countries; areal extent and classification of soil series; descriptions of ongoing soil survey investigations; various documents that define the standards used for soil surveys in the United States including field and laboratory methods; and links to Web Soil Survey and other soil surveys available in digital format.

The global distribution of soil orders at 1:1 million scale is illustrated in Figure 33.1—while Table 33.1 provides the areal extent of suborders associated with these orders. Tables 33.2 and 33.3 present proportional percentages of soil orders occurring in different soil temperature (Table 33.2) and soil moisture (Table 33.3) regimes as defined in *Soil Taxonomy* (Soil Survey Staff, 1999). Broad geographical distribution patterns (Figure 33.1) reflect regional conditions of climate, vegetation, geology, and topography, which function interactively over time. Soils found at any point on the land surface are considered products of multiple sets of pedogenic processes with state factors serving as controls (Chapter 30). Generally, soils are developed along polygenetic pathways, on dynamically evolving landforms under the influence of paleoclimates, in nonuniform parent materials. Hence, rarely can bio-, litho-, climo-, chrono-, and topo-sequence functions be developed with scientific rigor. Clearly, the distribution of soil orders in which local conditions strongly govern pedogenic dynamics over short-range distances is not well illustrated in Figure 33.1. In particular, such comments are germane to distributions of Histosols, Spodosols, Vertisols, and Gelisols. In these cases, the distributions presented in sections of this chapter may differ somewhat from those presented herein (Figure 33.1). This uncertainty will also impact the accuracy of

Global soil regions

Robinson projection
Scale 1:130,000,000

Soil orders

Alfisols	Entisols	Inceptisols	Spodosols	Rocky land
Andisols	Gelisols	Mollisols	Ultisols	Shifting sand
Aridisols	Histosols	Oxisols	Vertisols	Ice/Glacier

FIGURE 33.1 Global distribution of soil orders. (Courtesy of the USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC, 2006.)

TABLE 33.1 Global Areas and Percentages of Suborders and Miscellaneous Land Units Based on Ice-Free Land Area

Order	Suborder		Area (km² × 10³)	Proportion (%)
Alfisols	Aqualfs		1,029	0.8
	Cryalfs		2,531	1.9
	Ustalfs		6,024	4.6
	Xeralfs		893	0.7
	Udalfs		2,678	2.0
		Subtotal	13,156	10.1
Andisols	Cryands		251	0.2
	Torrands		1	<0.1
	Xerands		32	<0.1
	Vitrands		281	0.2
	Ustands		62	<0.1
	Udands		275	0.2
	Gelands		62	<0.1
		Subtotal	963	0.7
Aridisols	Cryids		1,036	0.8
	Salids		1,287	1.0
	Gypsids		680	0.5
	Argids		4,678	3.6
	Calcids		4,887	3.7
	Cambids		2,919	2.2
		Subtotal	15,487	11.9
Entisols	Aquents		109	0.1
	Psamments		4,447	3.4
	Fluvents		3,056	2.3
	Orthents		15,834	12.1
		Subtotal	23,446	17.9
Gelisols	Histels		1,006	0.8
	Turbels		5,065	3.9
	Orthels		5,692	4.4
		Subtotal	11,764	9.0
Histosols	Folists		—	<0.1
	Fibrists		196	0.1
	Hemists		981	0.8
	Saprists		331	0.3
		Subtotal	1,507	1.2
Inceptisols	Aquepts		3,657	2.8
	Anthrepts		450	0.3
	Cryepts		2,598	2.0
	Ustepts		2,230	1.7
	Xerepts		683	0.5
	Udepts		4,102	3.1
	Gelepts		6,043	4.6
		Subtotal	19,764	15.1
Mollisols	Albolls		7	<0.1
	Aquolls		118	0.1
	Rendolls		261	0.2
	Xerolls		923	0.7

TABLE 33.1 (continued) Global Areas and Percentages of Suborders and Miscellaneous Land Units Based on Ice-Free Land Area

Order	Suborder		Area (km² × 10³)	Proportion (%)
	Cryolls		2,464	1.9
	Ustolls		3,937	3.0
	Udolls		1,263	1.0
	Gelolls		156	0.1
		Subtotal	9,128	7.0
Oxisols	Aquox		322	0.2
	Torrox		31	<0.1
	Ustox		3,115	2.4
	Perox		1,167	0.9
	Udox		5,233	4.0
		Subtotal	9,868	7.6
Spodosols	Aquods		167	0.1
	Cryods		2,574	2.0
	Humods		57	<0.1
	Orthods		649	0.5
	Gelods		1,111	0.8
		Subtotal	4,558	3.5
Ultisols	Aquults		1,285	1.0
	Humults		380	0.3
	Udults		5,540	4.2
	Ustults		3,340	2.6
	Xerults		19	<0.1
		Subtotal	10,563	8.1
Vertisols	Aquerts		5	<0.1
	Cryerts		17	<0.1
	Xererts		98	0.1
	Torrerts		894	0.7
	Usterts		1,770	1.4
	Uderts		383	0.3
		Subtotal	3,167	2.4
Miscellaneous	Salt		145	0.1
	Shifting sands		5,341	4.1
	Rock		1,803	1.4
	Ice		2,210	
		Subtotal	7,288	5.6
Total			**130,658**	**100.0**

Source: Courtesy of USDA-NRCS Soil Survey Division, Washington, DC, 2010.

global distributions and areal data given for Histosols, Andisols, and Inceptisols found interspersed with Gelisols.

The human impact on soils is recognized in some taxa of Entisols and Inceptisols, but, in general, *Soil Taxonomy* does not adequately handle the anthropogenic effects on soils. For example, Mollisols can be transformed into Alfisols when subject to severe erosion; paddy soils can undergo secondary salinization and waterlogging under long-term rice culture; soils in urban/ industrial environments can be markedly modified by landfills,

TABLE 33.2 Extent of Soil Orders Found in Different Temperature Regions

| Soil Order | Soil Temperature Regions[a] | | | |
	Tropical	Temperate	Boreal	Tundra
Alfisols	38	39	23	0
Andisols	49	23	28	0
Aridisols	12	74	14	<1
Entisols	28	68	4	0
Gelisols	0	0	0	100
Histosols	21	8	71	0
Inceptisols	47	42	11	0
Mollisols	4	50	46	0
Oxisols	98	2	0	0
Spodosols	<1	18	82	0
Ultisols	69	31	<1	0
Vertisols	47	52	1	0

Source: Modified from data supplied by USDA-NRCS, Soil Survey Division, 1998.

[a] Derived from soil temperature regimes defined in *Keys to Soil Taxonomy* (Soil Survey Staff, 2010).

Tropical: isomesic, isothermic, and isohyperthermic (MAST ≥ 8°C in which the difference in mean summer and mean winter soil temperatures is <6°C at 50 cm depth).

Temperate: mesic, termic, and hyperthermic (MAST ≥ 8°C at 50 cm depth).
Boreal: frigid, isofrigid, and cryic (MAST is 0°C–8°C at 50 cm depth).
Tundra: Gelic (MAST <0°C or <1°C in Gelisols).

TABLE 33.3 Extent of Soil Orders Found in Different Soil Moisture Regimes

| Soil Order | Soil Moisture Regimes[a] | | | |
	Aridic	Xeric	Ustic	Udic[b]
Alfisols	0	8	56	36
Andisols	4	3	30	63
Aridisols	95	1	3	1
Entisols	60	4	22	14
Gelisols	N/A[c]	N/A	N/A	N/A
Histosols	8	1	23	68
Inceptisols	0	6	42	52
Mollisols	0	13	66[d]	22
Oxisols	<1	0	35	65
Spodosols	<1	1	9	90
Ultisols	0	<1	42	58
Vertisols	28	3	56	13

Source: Modified from data supplied by USDA-NRCS, Soil Survey Division, 1998.

[a] Soil moisture regimes defined in *Keys to Soil Taxonomy* (Soil Survey Staff, 2010).

[b] Udic soil moisture regime includes perudic moisture regime and soils with aquic conditions at a shallow depth (soil horizons saturated, reduced, and with redoximorphic features).

[c] N/A—not appropriate for the Gelisol order.

[d] A small percentage of Mollisols in the Ustic moisture regime may in fact have an Aridic soil moisture regime, especially in steppe regions of Eurasia that lie along the border between Ustic and Aridic soil moisture regimes.

landfarming, earth movement, and heavy metal contamination; and drastically, disturbed soils are common in regions where precious metals, rock aggregate, and fossil fuels have been mined. The International Committee on Anthropic Soils (ICOMANTH) where ANTH stands for anthropogenic is currently addressing this matter. For information on this activity, refer to the ICOMANTH Web site, http://clic.cses.vt.edu/icomanth/

Soil Taxonomy currently does not accommodate systematic cataloging of paleosols formed under remarkably different paleoenvironments. With increasing geomorphic age, properties of soils commonly reflect a welding of contemporaneous and paleoenvironments. In particular, when paleosols are well preserved, they are valuable proxies for biological and physicochemical evolution of the Earth. This is an area for *Soil Taxonomy* development that will likely gain greater attention in the future. With multiple uses of soil surveys for connectivity between the land surface and geological substrata (e.g., for environmental, hydrological, archaeological, and biogeoscience synergisms), there are increasing driving forces to observe soils beyond the 2 m depth limit currently in vogue. This is likely to enhance interest in revising *Soil Taxonomy* to better accommodate recognition and cataloging of paleosols as important morphogenetic markers of paleoenvironments (Chapter 29).

In perusing the global distributions of soil orders (Figure 33.1) and aerial extents given in Tables 33.1 through 33.3, several noteworthy relationships become apparent. These are summarized in a cursory overview in the following paragraphs. The organic-rich soils [Histosols (Section 33.2)], which comprises about 1.2% of the Earth's ice-free land surface (Table 33.1), generally occur in cool, humid, high latitude, boreal regions or in microhabitats such as seasonally ponded areas that favor the balance of organic matter accumulation over decomposition (Tables 33.2 and 33.3). This is because of saturated and reduced conditions, biological inhibitors, and/or cool, short summer periods with low evapotranspiration rates. Histosols are important wildlife and wetland habitats. Because they contain a disproportionally high level of soil carbon in relation to their aerial extent, Histosols are an important component of the global C cycle. These soils can be important sources of methane in the atmosphere, and accelerated organic matter decomposition resulting from more frequent drought and/or elevated temperatures associated with climate change not only potentially reduces their global abundance but also makes them a potential source of atmospheric carbon dioxide. Histosols are used extensively for cranberry, citrus, vegetable, rice, and sugarcane production and are a major fuel source in many regions. Major constraints to their use are subsidence, nutrient deficiencies, and wind erosion. Effective water-table management is critical in minimizing subsidence. These land resources are also critically limiting to construction activities because of their low soil strength and load bearing capacities.

Entisols (Section 33.4) comprise about 17.9% of the Earth's ice-free land surface (Figure 33.1). Over two-thirds of the Entisols are in the temperate region while the remainder mostly occur in the tropics (Table 33.2). Sixty percent of the Entisols

have aridic soil moisture regimes with ustic and udic being the next most extensive (Table 33.3). These are mineral soils, which do not have expression of diagnostic subsurface horizons within a specified depth of the soil surface, generally 1–2 m. Subsoil materials of these soils reflect the nature of the geological materials from which they were derived. These weakly expressed soils are commonly found on geomorphic surfaces, which are unstable because of frequent flooding, erosion/truncation, or human impact (drastically disturbed lands). They are formed also in coarse-textured, resistant mineral parent materials (e.g., quartzose sands) that are subject to little pedogenic development over time. Entisols are common along flood plains of river and stream valley systems, sand dunes in desert regions, on high-gradient mountainous terrain, and associated with recently mined or disturbed lands.

Andisols (Section 33.3) comprise <0.7% of the Earth's ice-free land surface (Table 33.1), but these soils are so unique physically, chemically, and mineralogically that a separate soil order was established for them. They are found along the tectonically active Pacific Ring of Fire, Central Atlantic Ridge, North Atlantic rift, the Caribbean, and the Mediterranean regions where volcanic or pyroclastic deposits are common (Figure 33.1). About half of them are located in tropical regions with the remainder split between temperate and boreal climates (Table 33.2). About two-thirds occur under udic soil moisture regimes and one-third in ustic regions (Table 33.3). These weakly developed soils are texturally undifferentiated and characterized by short-range order (amorphous) aluminosilicates that have not been translocated from upper to lower horizons. Andisol land resources are used extensively for crop production in developing countries. They commonly have favorable physical properties for plant growth while in many areas, high native fertility is present because of frequent additions of tephra that renew soil nutrients.

Inceptisols (Section 33.5), which comprise about 15.1% of the Earth's ice-free land surface (Figure 33.1), occur indiscriminately on a global basis because they lack a sharply defined central concept. Over 90% occur in tropical and temperate climates (Table 33.2) under ustic and udic soil moisture regimes (Table 33.3). Inceptisols serve to make *Soil Taxonomy* fully bifurcated. Soils classified as Inceptisols generally contain light-colored surface horizons with weakly expressed diagnostic subsoil horizons (cambic) but, under certain circumstances, may contain more strongly developed subsoil diagnostic horizons (e.g., petrocalcic, petrogypsic, duripan, and fragipan). Where these soils reflect youthfulness, they occur in high-gradient mountainous regions, along major river systems as terraces and deltaic/fluvial plains, and as soils developed from carbonate-rich bedrocks or sediments in Mediterranean environments. They also occur interspersed among more strongly developed Alfisols, Ultisols, and Oxisols in tropical and subtropical regions. Properties of Inceptisols vary widely as does their potential for plant production and engineering use.

Gelisols, the newest of the soil orders (Section 33.6), comprise about 9.0% of the Earth's ice-free land surface (Figure 33.1). These are soils of the high-latitude polar regions underlain by permafrost (materials wet/dry that remain below 0°C for 2 consecutive years) at depths of 1–2 m depending on cryoturbation (frost churning) activity (Table 33.2). Cryoturbation in Gelisols is driven by the physical volume change from water to ice and subsequent moisture migration along thermal, hydrostatic, chemical, and electrical gradients. Cryopedogenic processes include freezing/thawing, cryoturbation, frost heaving, cryogenic sorting, thermal cracking, and ice segregation. Remarkable spatial variability occurs in soil properties over distances of a few meters or less giving rise to pattern ground in the form of stone stripes, stone circles, high and low center polygons, and barren frost boils. Because Gelisols are large C sinks, warming in polar regions may result in melting of permafrost, increased organic matter decomposition, and increased CO_2 release to the atmosphere. Gelisols present management challenges because of volume changes associated with cryopedogenic processes. In addition, melting of ice following disturbance associated with construction or agricultural production leads to subsidence or thermokarst. Only through careful management of subsurface ice and maintenance of a negative thermal balance can the integrity of structures be preserved.

Vertisols (Section 33.7) comprise about 2.4% of the Earth's ice-free land surface (Figure 33.1). About half the Vertisols are found in tropical and about half in temperate environments (Table 33.2). Over 50% of the Vertisols occur in regions with ustic soil moisture regimes where seasonal desiccation is common. The remainder are found primarily under aridic and xeric environments; here, the oscillation from wet to dry soil conditions is less frequent (Table 33.3) and gilgai expression less extreme. These clayey, shrink/swell soils are highly diverse in physical, chemical, biological, and mineralogical properties. Although concentrated in subtropical and tropical environments, they do occur across broad climatic, vegetative, topographic, and geologic regions of Africa, India, Australia, China, and North and South America (Figure 33.1). Swelling caused by wetting dry Vertisols often results in soil failure when swelling pressures exceed shear strength. Soil failure is usually expressed as slickensides occurring as diagonal, polished, and grooved slip surfaces, wedge-shaped structural units, and microtopography in the form of gilgai. Commonly, these soils have self-swallowing and self-mulching surface features where strongly aggregated surface materials fill in vertical structural cracks. While inversion by turbation is commonly invoked as a major pedogenic processes, most movement in these soils is believed to be diagonally along inclined major slickenside fault planes. Major limitations for using these highly productive soils for agriculture, especially in developing countries, are their high energy requirements for tillage and narrow favorable soil moisture range for workability. Construction activities are constrained by their propensity for soil failure and high shrink/swell activity.

Mollisols (Section 33.8) comprise about 7% of Earth's ice-free land surface (Figure 33.1). Essentially, all the Mollisols are approximately equally split between temperate and boreal regions (Table 33.2). About two-thirds of the Mollisols are in

ustic soil moisture regimes (Ustolls) with most of the remainder occurring in regions with a udic moisture regime (Udolls) and the rest in climates with winter precipitation and summer periods of soil moisture deficit (Xerolls) (Table 33.3). A few Mollisols may occur under aridic soil moisture regimes in the southern steppe zones of Eurasia but their extent is uncertain. Mollisols are the dark-colored, high base status soils commonly found under prairie grass vegetation in mid-latitudes of North and South America and Eurasia. The distribution pattern of these soils in Eurasia is in east-west trending belts between Alfisols to the north and Aridisols and Inceptisols to the south. This distribution reflects decreasing precipitation and increasing temperature gradients from north to south. In contrast, in North America, the precipitation isohyets and isotherms are set normal to each other with precipitation decreasing from east to west and temperature increasing from north to south. These climatic controls result in Mollisols being bordered on the east by Alfisols and Ultisols, and to the west by Aridisols, Entisols, and drier Alfisols. Mollisols are commonly associated with calcareous or high base status glacial drift, limestones, and loess. Mollisols (especially Udolls) are the "bread basket" soils of the Americas and Eurasia because their high native fertility, occurrence in favorable climates, and excellent physical and chemical properties result in high crop production potentials.

Spodosols (Section 33.9) comprise about 3.5% of Earth's ice-free land area (Figure 33.1) with over 80% occurring in boreal environments (Table 33.2) with udic soil moisture regimes (Table 33.3) between Mollisols of mid-to-high latitudes and Alfisols in lower latitudes. Spodosols are also commonly associated with imperfectly drained sediments in coastal regions such as the Atlantic Coast of the southeast United States. Spodosols commonly have prominent eluvial E horizons overlying black to red spodic and placic horizons. The spodic horizons have formed from translocated illuvial organo-metal (Al and Fe) complexes. These soils have many properties in common with Andisols. The major difference is that Spodosols have formed for a sufficient period of time for the weathered organic–metal complexes to be translocated to subsoils. In Andisols, little or no translocation of these complexes has taken place. Formation of Spodosols is favored by environments with abundant acid litter, coarse-textured resistant parent materials, and strong leaching potentials. These soils are very strongly acid and extremely phosphorus deficient. Native vegetation is composed largely of acid tolerant plants. They are little used for agricultural production because of low nutrient and water retention and high management requirements. The major land use is forests for lumbering and recreation.

Aridisols (Section 33.10) which comprise about 11.9% of the Earth's ice-free land area (Figure 33.1) occur mostly in hot temperate deserts with aridic soil moisture regimes, dry coastal regions, or on rain-shadow plains leeward of high mountains (Tables 33.2 and 33.3). Locally, they also occur in more humid regions where salts have concentrated at or near the soil surface through evaporative pumping from shallow saline ground waters. Aridisols range from base-rich to base-poor soils and exhibit a wide diversity in physical, chemical, and mineralogical properties. Commonly, they represent contemporaneous xerophytic environments, but some Aridisols have well-expressed subsoil diagnostic properties such as argillic, natric, calcic, petrocalcic horizons. These likely reflect more pluvial paleoclimates, and the diagnostic horizons would be, in part, relict. Aridisols are used extensively for rangelands, military bases, and nomadic agrarian activities. Although these soils are not generally considered arable land resources without irrigation, where an adequate quantity of high-quality subsurface water is available they become suitable for intensive agriculture and urban development. Irrigation and drainage must be managed for a favorable salt balance in Aridisols.

Alfisols (Section 33.11) and Ultisols (Section 33.12) are closely related soil orders that comprise, respectively, about 10.1% and 8.1% of the Earth's ice-free land area (Figure 33.1). These soils, developed mostly under deciduous forested or savannah environments, occur extensively in mid-to-lower latitudes. Alfisols are distributed approximately equally across tropical, temperate, and boreal environments whereas the distribution of Ultisols is skewed toward tropical conditions (Table 33.2). For both Alfisols and Ultisols, most of their extent is associated with ustic and udic soil moisture regimes (Table 33.3). Alfisols and Ultisols are differentiated from other orders on the basis of textural differentiation between surface and subsoil horizons. The textural differentiation partly defines argillic and kandic horizons and has often resulted from translocation and/or in situ neoformation of clays. Alfisols differ from Ultisols in having higher base saturation (>35%) at a specified depth in the subsoil. Rationale for this differentiation was that Alfisols could sustain traditional slash/burn subsistence agriculture without significant external fertilizer inputs, while Ultisols would require applied nutrients for crop production. Both orders are used extensively for forest and crop production. Major limitations are nutrient deficiencies, wind and water erosivity, and seasonal soil moisture deficits, especially for those in xeric and ustic regions.

Sesquioxide-rich Oxisols (Section 33.13) comprise about 7.6% of the Earth's ice-free land surface occurring in equatorial regions intermixed with Ultisols and Inceptisols (Figure 33.1). Nearly all the Oxisols occur in tropical environments (Table 33.2) with two-thirds under udic and one-third under ustic soil moisture regimes (Table 33.3). Oxisols contain low activity clays and are nutrient poor, weakly buffered systems that have little or no horizon differentiation. This reflects (1) their high weathering intensities in low latitude, humid environments, (2) their geomorphic stability and age associated with mafic-rich, Precambrian platforms, or (3) development from polcycled, pre-weathered parent materials derived from highly weathered terrestrial source areas. These soils are often considered residual carcasses, but many are very productive for agriculture and forestry if properly managed for water conservation, soil erosion, and nutrient inputs. With high nutrient inputs and management to conserve soil and water, Oxisols can be very productive agricultural soil systems.

References

Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd ed. USDA-SCS Agric. Handb. 436. US Government Printing Office, Washington, DC.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th ed. USDA-NRCS. US Government Printing Office, Washington, DC.

33.2 Histosols

Randall K. Kolka

Martin C. Rabenhorst

David Swanson

33.2.1 Introduction

While most soils of the world comprise primarily mineral materials, a small but important group of soils are formed from organic materials derived from plants, or less frequently, from animals. *Organic soil materials* contain a minimum of 12%–18% organic carbon, depending on the particle size of the mineral component (Soil Survey Staff, 2010). Generally speaking, soils with at least 40 cm of the upper 80 cm that are organic materials, and which do not have permafrost within 1 m of the soil surface, are Histosols. Prior to 1997, organic soils with permafrost were included in the Histosol order; they are now placed in the Histel suborder of Gelisols (Soil Survey Staff, 1998). Following the separation of Histosols and permafrost soils, Histosols occupy about 1% of the global land area while the Histel suborder of Gelisols occupies about 0.8% (Buol et al., 2003). Organic soil materials are commonly referred to as *peat*, and land covered by Histosols or Histels is known as *peatland*. The term *mire* is a synonym of peatland that is more commonly used in Europe. Histosols also include a narrowly distributed group of soils, the Folists, that consist of well-drained organic soil materials that directly overly bedrock or coarse fragments with little or no intervening fine soil. The peat layer in Folists may be (and usually is) thinner than the 40 cm required for other Histosols.

Because of their high organic C content, many Histosols have been utilized as a combustible energy resource. Mankind has mined and burned peat since prehistoric times, and peat is still an important fuel in a number of northern countries, although it has a lower energy rating than oil or coal. In Russia, Germany, and Ireland, peat is not only utilized for domestic heating but is also used on a large scale in electricity generation. In 2009, peat burning was the source of 10% of Ireland's electricity (Public communication, 2009a), and in 2003, about 7% of Finland's electricity was a result of burning peat (Kirkinen et al., 2007). Sweden, Belarus, Latvia, Estonia, and Lithuania also utilize peat for energy; and Canada is currently investigating the possibility of using peat for power production in Ontario and Labrador (Public communication, 2009b, 2009c). In addition to being mined as an energy source, peat is mined for use as a soil amendment in agriculture and horticulture. Moreover, the agricultural value of Histosols has long been recognized. Provided that the water tables can be effectively managed, high yields of vegetables and other specialty crops can be produced on Histosols in such different climatic regions as Michigan and Florida (Lucas, 1982).

Because most Histosols occur in wetlands,* their utilization as agricultural and energy resources has come under intense scrutiny. Histosols perform many of the beneficial functions of wetlands, and they are negatively impacted by mining, drainage, or other practices associated with agriculture; thus, there are benefits derived from preservation of Histosols in their natural state. Histosols are perhaps uniquely fragile and are highly vulnerable to degradation. When drained or dry, organic soil material is highly susceptible to wind and water erosion (Lucas, 1982; Parent et al., 1982). Organic soils also have very low strength, are highly compressible (MacFarlane, 1969; MacFarlane and Williams, 1974), and gradually subside by decomposition if drained (Gesch et al., 2007). Furthermore, within the framework of current discussions of global climate change and carbon budgeting, Histosols contribute significantly to the terrestrial carbon pool. On an areal basis, C storage in Histosols is often greater than mineral soils by an order of magnitude (Rabenhorst, 1995).

33.2.2 Distribution

Histosols occur at all latitudes, but are most prevalent in the boreal forest regions of northern North America, Europe, and Asia (Figure 33.2). The world's largest expanses of Histosols occur in the West Siberian lowland (Walter, 1977) and the Hudson Bay lowland of central Canada (Sjors, 1963; Canada Committee on Ecological (Biophysical) Land Classification, National Wetlands Working Group, 1988). At lower latitudes, Histosols occur locally on humid coastal plains, notably southeast Asia and Indonesia (Anderson, 1983).

Histosols in the United States are most widespread in lowlands of the Great Lakes region, the northeast, the Atlantic Coastal plain and Florida, the Pacific Northwest, and Alaska (Figure 33.2). The largest expanses of Histosols in the continental United States are on the Lake Agassiz plain in north-central Minnesota (Wright et al., 1992). Coastal and estuarine areas inundated by tidal water are also sites for Histosol formation, most notably along the Atlantic and Gulf coastlines. Drained Histosols are widely used in agriculture in the Great Lakes region, southern Florida, and the Sacramento–San Joaquin delta region of California. In the semiarid Great Plains and mountainous west, Histosols are very rare and occur only in areas of steady groundwater discharge (Mausbach and Richardson, 1994) or humid areas at high elevations (Cooper and Andrus, 1994). Organic soils are widespread in lowlands throughout Alaska, although most of the organic soils in central and northern Alaska have permafrost and hence are classified as Histels in the Gelisol soil order rather than Histosols.

---

* The National Technical Committee on Hydric Soils (USA) has included "All Histosols except Folists" within the criteria for hydric soils (USDA-NRCS, 1996).
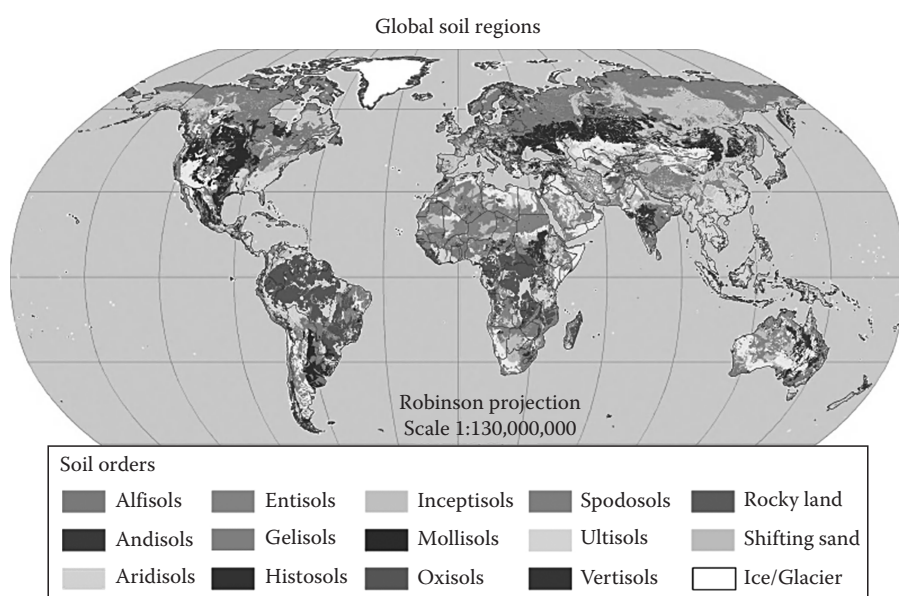
FIGURE 33.2 Worldwide distribution of Histosols. (Courtesy of USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC, 2010.)

33.2.3 Formation of Histosols

33.2.3.1 Parent Material

In contrast to the wide variety of mineral materials, which may serve as parent materials for other soils, the parent materials from which Histosols are formed are organic in nature. The conditions that cause the accumulation of organic parent materials are very closely tied to the processes, which form various organic soil horizons. The unique properties of Histosols result from the nature of the organic parent material.

Some of the factors, which affect the nature of organic parent materials, include hydroperiod, water chemistry, and vegetation type and will be discussed in more detail in the following sections. The net accumulation of organic materials occurs when rates of additions (usually as primary plant production) exceed rates of decomposition. In natural soils of most ecosystems, a steady state exists between these two processes, which maintains the quantity of organic C in the surface horizons somewhere between 0.5% and 10%, although some forest soils have relatively thin layers of organic soil materials (O horizons). In Histosols, the rates of decomposition are slowed and organic matter accumulates to the degree that the organic materials amass to a significant depth. In most cases, this is caused by saturation in soils leading to anaerobic conditions, which causes organic matter decomposition to be less efficient than under aerobic conditions. Occasionally, usually under cool and moist conditions, organic parent materials may accumulate without prolonged saturation, producing soils in the unique Folist suborder of Histosols. Under certain conditions and landscapes, mineral materials can be added to accumulating organic parent materials, typically by alluviation or eolian deposition. In such cases, the balance between mineral inputs and organic accumulation will determine whether a Histosol or mineral soil will form.

The accumulation of organic parent materials often occurs over long periods of time and under changing conditions.

Thus, the stratigraphy of a bog may reflect many thousands of years of organic matter accumulation. Rarely does vegetation remain constant over such long periods. Microscopic examination of the partially decomposed peat or evaluation of pollen or plant microfossils can provide information concerning the types of plants that have contributed to the organic parent material during various stages of accumulation (e.g., Wieder et al., 1994).

The organic parent material of Histosols is a major source of acidity. Acids produced by the partial decomposition of organic matter cause the organic horizons of Histosols to be highly acidic unless the acids are neutralized by bases that were dissolved from mineral soils or rocks and transported into the peat by groundwater. Highly acidic peatlands are called *bogs* and commonly described as *ombrotrophic* (rain-fed, because all nutrients are derived from atmospheric sources). The less acid Histosols that receive base-rich groundwater or runoff are called *fens* and usually described as *minerotrophic* (fed with mineral-derived nutrients). The term *swamp* is often applied to forested systems on both fens and on mineral soil wetlands. Peatlands are also sometimes divided by their acidity into *oligotrophic* (very acid and mineral-poor), *mesotrophic* (intermediate), and *eutrophic* (weakly acidic to neutral and mineral-rich) classes (Moore and Bellamy, 1974; Gore, 1983).

33.2.3.2 Climate

Histosol formation is favored by wet or cold climates. Wetness and cold favor Histosol formation by hindering decomposition of organic matter. The northern Histosol-dominated regions of North America and Eurasia (Figure 33.2) have temperate and boreal climates in which average annual precipitation exceeds annual potential evapotranspiration (Trewartha, 1968). Further to the north, in Greenland and the islands bordering the Arctic Ocean, for example, Histosols are rare because the growing season is so short that there is little production of organic matter. To the south of the boreal zone, Histosol formation is apparently

constrained by rapid decomposition of soil organic matter. At lower latitudes, Histosols are restricted mostly to coastal plains with very flat topography, high annual precipitation, and no dry season, or to those areas where a high water table is maintained by tidal waters, leading to coastal marshes and mangrove systems (Anderson, 1983).

Climate affects not only where Histosols occur, but also the chemistry of the resulting soils. Where average annual precipitation exceeds annual potential evapotranspiration, Histosols in suitable settings can remain saturated by rainfall alone (Ivanov, 1981). In such regions, both the highly acidic Histosols of bogs and the less acid Histosols of groundwater-fed fens may form. In less humid regions, where groundwater is required to maintain saturation of soils, only the less acid, minerotrophic Histosols can form.

Climate change scenarios suggest more frequent extreme events such as heat waves, droughts, and high precipitation events (IPCC, 2007). Increases in drought and heat will likely lead to a higher frequency of fire in the boreal zone resulting in significant losses of C from Histosols. Currently, disturbances, mainly fire, reduce the net C uptake of continental boreal peatlands by about 85% (Turetsky et al., 2002). Predicted increases in fire frequency and intensity could lead to an even greater proportion of C uptake being balanced by fire, and boreal peatlands may actually become a net C source to the atmosphere (Turetsky et al., 2002).

33.2.3.3 Topography

While climate controls the occurrence of Histosols on a regional scale, topography controls where they occur on a given landscape. The classical soil-forming factor of topography is considered broadly here to incorporate relief, geomorphic setting, and hydrologic setting. Histosols occur where the setting facilitates concentration of runoff, discharge of groundwater, or retention of precipitation (Figure 33.3a through c). These conditions are most often satisfied in topographic depressions or very flat areas. On plains with low-permeability substrates in the boreal zone, Histosols may cover entire interfluves (Sjors, 1963; Walter, 1977), and in extremely humid climates, such as the British Isles and southeastern Alaska, Histosols may also occur on gentle slopes (Moore and Bellamy, 1974; Sjors, 1985; Ward et al., 2007). Histosols of floodplains receive suspended mineral matter during floods, while other Histosols obtain only dissolved material from source waters or inputs of eolian materials.

In lower latitude coastal areas, rising sea level has caused brackish or saline waters to engulf drowned river valleys or to extend over formerly upland soils. This has led to the formation of coastal marsh Histosols in several geomorphic settings (Darmody and Foss, 1979). Figure 33.4 shows a schematic of a cross section of a submerged upland type marsh where organic soil materials are accumulating over what were previously upland soils, forming Histosols. As sea level continues to rise, the margin of the marsh is pushed landward, and the organic materials continue to thicken such that older and deeper Histosols generally exist closer to the open water (Rabenhorst, 1997).
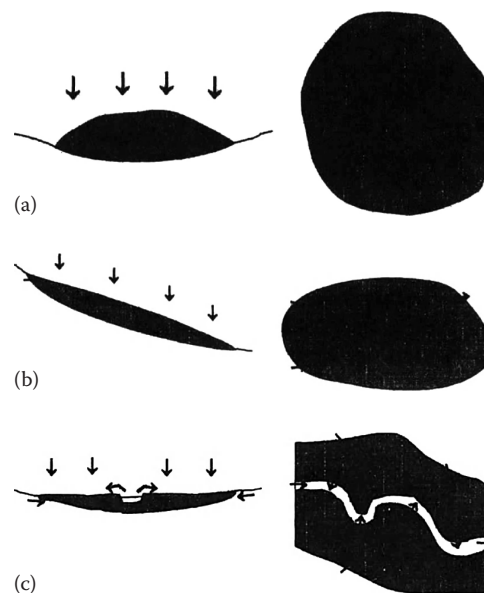
FIGURE 33.3 Settings of peatlands. Peat is shown in gray, and arrows indicate the direction of water movement (a) Bog. The domed surface of the bog precludes input of runoff or groundwater containing bases dissolved from minerals. All water is derived from the atmosphere and evaporates or runs radially off the bog. (b) Fen or swamp. Runoff and groundwater from mineral soils surrounding the peatland supplement precipitation on the peatland. (c) Floodplain fen or swamp. The peatland receives precipitation and runoff or groundwater from mineral soils adjacent to the peatland. Water seeps from the peatland into the stream during periods of low water, while during floods suspended mineral matter is deposited on the peatland, increasing the content of mineral matter in the peat.
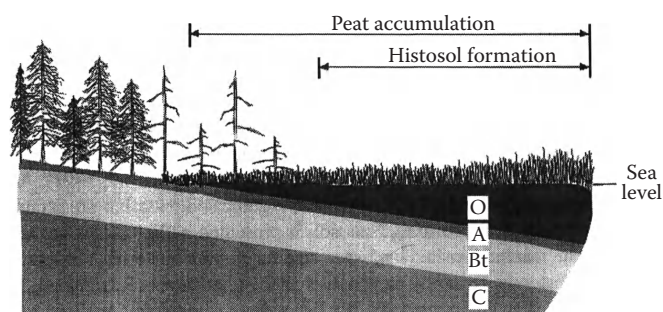
FIGURE 33.4 A schematized cross section of a submerged-upland type marsh showing development of O horizons and a Histosol (probably a Sulfihemist) over what was formerly an upland soil with an argillic (Bt) horizon.

The formation of Histosols, like other soils, is a function of topography, but Histosols are unique among soils in that their formation also modifies the topography. Accumulation of organic soil material can fill depressions and create gentle topographic highs where the topography was once level or depressional (Figure 33.5). The development of a topographic high due to peat accumulation can prevent base-rich water from moving onto the peatland and thereby facilitate formation of a highly
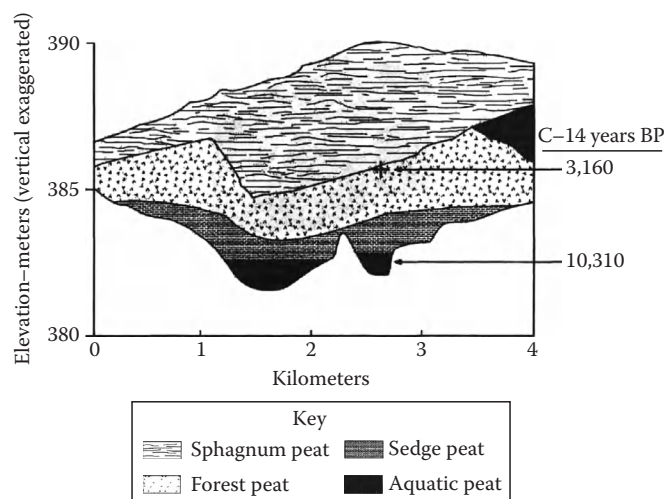
FIGURE 33.5 Stratigraphic cross section of a bog in the Myrtle Lake peatland, northern Minnesota. (Modified from Heinselman, M.L. 1970. Landscape evolution, peatland types and the environment in the Lake Agassiz Peatlands Natural Area, Minnesota. Ecol. Monogr. 40:235–261.)

FIGURE 33.6 Typical bog vegetation on a highly acidic Histosol in boreal North America. Trees are stunted black spruce (*Picea mariana*); the largest trees visible are about 4 m tall. Understory plants include ericaceous dwarf shrubs, sedges (*Carex* sp.), and a continuous cover of *Sphagnum* sp. moss. The soil is a Typic Borohemist. Toivola peatland, northeastern Minnesota.

acidic bog. Peat accumulation also produces intriguing microtopography on some peatlands, notably a pattern of ridges (or *strings*) and pools (Foster et al., 1983; Seppälä and Koutaniemi, 1985; Swanson and Grigal, 1988) (or *flarks*), the cause of which is still debated (Washburn, 1980).

Histosols frequently form by accumulation of organic matter in basins of lakes or ponds, a process known as *lakefill* or *terrestrialization*. Histosols may also form by *paludification*, the expansion of wetland onto what was originally drier soils. Both processes operated in the formation of the peatland depicted in Figure 33.5. Folists, unlike other Histosols, generally occur in mountainous regions (Reiger, 1983; Wakeley et al., 1996). Mountainous settings provide the high precipitation and underlying bedrock or fragmental material required for Folist formation.

33.2.3.4 Vegetation

The flora of Histosols varies widely as a result of the great range in climates over which they occur. Moreover, while most Histosols owe their existence to saturation of the soil by water, the depth at which saturation occurs is variable and this has a major effect on the vegetation. The vegetation of bogs in the worldwide circumboreal zone is remarkably uniform, apparently due to the limited number of plants that can tolerate the poor nutrient conditions, cold climate, and high water table of these soils (Figure 33.6). *Sphagnum* mosses cover the ground, along with scattered sedges (*Carex*) and cotton sedges (*Eriophorum*). Low shrubs from the family Ericaceae are common, and trees are usually present but stunted; black spruce (*Picea mariana*) is most widespread in North America and Scot's pine (*Pinus sylvestris*) in Eurasia. The vegetation of bogs in more southerly climates includes different species, but shares with northern bogs the *Sphagnum* moss and prevalence of nutrient-conserving evergreen plants (Anderson, 1983; Hofstetter, 1983).

The less-acid conditions of groundwater-fed Histosols permit a greater diversity of plants, many of which also occur on wet mineral soils. On minerotrophic peatlands with the water table continuously near the surface (fens), most vascular plants are *aerenchymous* grass-like plants (mainly sedges, family Cyperaceae) that can transport oxygen downward to their roots in hollow stems, allowing root metabolism in anaerobic soils (Crawford, 1983). Fens that exhibit an aerobic surface horizon present during at least part of the growing season, allow growth of nonaerenchymous plants, including woody plants (Gill, 1970; Kozlowski, 1984). Though nutrient conditions for tree growth on these Histosols are superior to those of bogs, trees may be stunted due to short duration of aerobic conditions or thinness of the aerobic rooting zone.

Many of the plants that occur on Histosols are highly adapted to specific conditions of pH/nutrients and high water table; thus, plants are useful indicators of these conditions (Heinselman, 1963, 1970; Sjors, 1963; Jeglum, 1971; Vitt and Slack, 1975; Vitt and Bayley, 1984; Andrus, 1986; Janssens and Glaser, 1986; Swanson and Grigal, 1989, 1991; Glaser, 1992; Janssens, 1992). Because the high water table restricts rooting of most plants to near-surface soil horizons, vegetation is a useful indicator of pH/nutrient conditions near the surface but not at depth.

Because the vegetation actually creates most of the soil material in Histosols, composition of the vegetation that formed the soil exerts strong control over its physical and chemical properties. Peats are commonly divided into three broad groups on the basis of botanical composition: moss, herbaceous (sedge), and woody peat (Kivinen, 1977). The botanical origin of organic soil material may be determined by examination of plant remains (Birks and Birks, 1980; Janssens, 1983; Levesque et al., 1988).

The changes in soil drainage and trophic conditions on peatlands that accompany peat accumulation affect the vegetation and hence botanical composition of the peat. As peat

TABLE 33.4   Rates of Peat Accretion/Accumulation in Various Organic Soils

Location	Site Characteristics	Peat Accretion–Accumulation Rate (mm year^{-1})	References
Alberta, Canada	Fen, last 50 years	3.2–6.4	Vitt et al. (2009)
Northwest Territories, Canada	Rich fen	0.38	Robinson (2006)
	Poor fen	0.41	
	Bog	0.47	
West Siberia, Russia	Forested bog	0.40–0.70	Peregon et al. (2007)
	Shrub bog	0.25–0.60	
Western Canada	Bog	3.0	Turetsky et al. (2007)
British Columbia, Canada	Sloping, open peatland	0.94–1.7	Asada and Warner (2004)
Southern Sweden	Bog	5.2–5.9	Belyea and Malmer (2004)
West Siberia, Russia	River valley fen	0.84	Borren et al. (2004)
	Raised bog	0.67	
	Forested fen	0.42	
	Bog	0.37	
	Flow-through fen	0.38	
	Flow-through fen	0.35	
	Bog	0.42	
	Bog	1.13	
North America		0.58	Gorham et al. (2003)
North Dakota, USA			
Minnesota, USA		0.41–0.79	
Quebec, Canada		0.59–0.83	
Maine, USA		0.47–0.56	
New Brunswick, Canada		0.49–0.79	
Nova Scotia, Canada		0.40–0.48	
Newfoundland, Canada		0.39–1.05	
Alaska, USA		0.18–0.38	
West Siberia, Russia	Bog	0.57	Glebov et al. (2002)
Southeastern Norway	Bog	3–28	Ohlson and Okland (1998)
Bergslagen, Sweden	Raised bog	0.3–1.0	Derived from Almquist-Jacobson and Foster (1995)
Slave Lake, Canada	*Sphagnum* peat	0.3–0.6	Kuhry and Vitt (1996)
Eastern United States	Last 100 years		Wieder et al. (1994)
	Minnesota bog	2.4	
	Pennsylvania fen	1.4	
	Maryland fen	1.9	
	West Virginia bog 1	3.1	
	West Virginia bog 2	2.3	
West Greenland	Nearshore peat	0.43	Bennike (1992)
Subarctic and boreal Canada	Based on 138 basal ^{14}C dates	0.31–0.54	Gorham (1991)
S. Sweden and N. Germany		0.70	Tolonen (1979) (after Gorham, 1991)
S. and Central Finland		0.75	Tolonen (1979) (after Gorham, 1991)
N. Europe		0.60	Aaby (1986) (after Gorham, 1991)
Boreal USSR	Raised bogs	0.6–0.8	Botch et al. (1983)
Siberian USSR	Palsa province	0.2–0.4	Botch and Masing (1983)
Eurasia		0.52	Zurek (1976) (after Gorham, 1991)
Maine, USA		0.35–0.75	Tolonen et al. (1988) (after Gorham, 1987)
Minnesota, USA	Red Lake, Minnesota	0.85–1.15	Gorham (1987)
Los Angeles, USA	Coastal marsh		Derived from Nyman and DeLaune (1991)
	Fresh	6.5–8.5	
	Brackish	5.9–9.5	
	Saline	7.5–7.6	

TABLE 33.4 (continued)   Rates of Peat Accretion/Accumulation in Various Organic Soils

Location	Site Characteristics	Peat Accretion–Accumulation Rate (mm year^{-1})	References
Chesapeake Bay, MD	Coastal marsh, Brackish	3.3–7.8	Derived from Kearney and Stevenson (1991)
Chesapeake Bay, MD	Coastal marsh, Brackish	3.5–7.5	Griffin and Rabenhorst (1989)
Chesapeake Bay, MD	Coastal marsh, Brackish	1.4–3.2	Hussein (1996)
	Current (^{210}Pb)	0.5–1.1	
	Long term (^{14}C)		
Louisiana, USA	Barataria basin, Coastal marsh	7–13	Hatton et al. (1983)
Massachusetts, USA	Barnstable coastal Marsh	1.1–2.6	Redfield and Rubin (1962)

accumulates, the rooting zone often becomes more and more isolated from mineral nutrient sources, and portions of the peatland may become drier as the ground surfaces rises. A cross section of a bog in Minnesota shows how a lake was filled in with aquatic peat (i.e., limnic material; Figure 33.4). Then, the peatland expanded onto what was originally dry land of the lakeshore as minerotrophic sedge peat was followed by minerotrophic forest peat as the surface rose and became drier. By about 3000 years ago, the center of the peatland became isolated from minerotrophic water, allowing accumulation of highly acidic *Sphagnum* moss peat, which subsequently expanded over the entire peatland and continues to accumulate today. Such changes in vegetation over time in peatlands make it difficult to predict subsurface peat properties from surface vegetation (Swanson and Grigal, 1989).

33.2.3.5 Time

Essentially all extant Histosols have formed since the end of the Pleistocene epoch. Most northern latitude Histosols occupy regions which were covered by glaciers during the last ice age and have formed following the glacial retreat. Reported average rates of peat accumulation in northern bogs and fens have been as high as >3 mm year^{-1}, but more typically fall in the range of 0.2–0.7 mm year^{-1} (Table 33.4). These average rates usually are based on basal ^{14}C dates, and actual rates may have been higher or lower during particular periods.

Although distant from the glacial activity, coastal Histosols at lower latitudes were also impacted by the glaciation. During the glacial maximum (approximately 20,000 years ago), sea level worldwide was approximately 150 m below the present level when large quantities of water were tied up in the glacial ice. Melting of the ice and concurrent ocean warming caused sea level to rise at such a rapid rate (10–20 mm year^{-1}) that initially vegetation could not colonize the tidal regions. Approximately 3000–5000 years ago, sea level rise slowed to a more modest pace such that marsh vegetation could become established and organic parent materials began to accumulate (Bloom and Stuvier, 1963; Redfield, 1972). As sea level has continued to rise, organic materials have accumulated in Histosols, and coastal marshes and mangrove systems generally have been thought to have accreted at approximately the rate of sea level rise.

In addition to the eustatic sea level rise, sediment in transgressing coastal regions is subsiding (e.g., along the Atlantic and Gulf coasts of the United States). The combination of rising sea level (presently estimated at 1 mm year^{-1} worldwide) and coastal subsidence can be joined to yield an apparent sea level rise, which is substantially greater. Estimates of peat accretion in coastal areas generally range from 3 to 8 mm year^{-1}, which are much higher than in noncoastal regions, with even higher rates reported in rapidly subsiding areas (Table 33.4). Current evidence suggests that the highest rates of sea level rise may be too great for marsh systems to maintain, and that some of these areas are suffering marsh loss (Kearney et al., 2002).

33.2.4 Morphological Properties of Histosols

Most organic soil material is derived from terrestrial plants, and soil particles initially resemble the plants from which they were derived. As decomposition progresses, the organic matter is converted into a homogenous, dark-colored mass. Some organic soil materials are derived from aquatic plants and animals that accumulate on the bottom of water bodies, producing *limnic materials* (Finney et al., 1974; Soil Survey Staff, 2010).

The most obvious morphological properties of organic soil horizons are related to their degree of decomposition. Master horizons Oi, Oe, and Oa are used to designate fibric, hemic, and sapric horizons, respectively (Table 33.5) and are defined by the portion of the soil material, which retains discernable plant fibers after rubbing and by the color of a Na pyrophosphate extracting solution (Soil Survey Staff, 2010). Other methodologies and rating scales have been developed and utilized for evaluating degree of decomposition; the one most broadly used in Europe is the Von Post scale, which ranks organic materials on a scale of decomposition from 1 to 10 based on soil color, quantity of recognizable fibers, and the proportion of material remaining in one's hand when the sample is squeezed. Von Post scale numbers 1–4 correspond approximately to fibric material, 5–7 to hemic, and 8–10 sapric (Von Post and Granlund, 1926; Clymo, 1983).

There are a number of soil properties which are related to the degree of decomposition of the organic materials, including color and a variety of physical and chemical properties. Field moist colors of sapric organic soil material are usually

TABLE 33.5   Defining Morphological Criteria for Histosol Organic Horizons

Horizon Designation	Type of Material	Common Descriptor	Volumetric Rubbed Fiber (RF) Content	Color of Pyrophosphate Extract (Value/Chroma)
Oi	Fibric	Peat	RF > 3/4 or RF > 2/5	7/1, 7/2, 8/1, 8/2, 8/3
Oe	Hemic	Mucky peat	2/5 > RF > 1/6	Does not otherwise qualify for either fibric or sapric materials
Oa	Sapric	Muck	RF < 1/6	Below or right of line drawn to exclude blocks 5/1, 6/2, 7/3

Source: Soil Survey Staff. 2010 Keys to soil taxonomy. 11th edn. USDA-NRCS, U.S. Government Printing Office, Washington, DC.

nearly black. Fibric peat is lighter colored and often reddish in hue, while hemic peat has an intermediate color (Table 33.6). The relation between the degree of decomposition and various physical properties is discussed later under the section on physical properties.

Undrained Histosols typically lack pedogenic structure as it is usually defined, although flattened plant remains and stratification commonly produces a plate-like structure (Lee and Manoch, 1974). This sedimentary structure often results in Histosols having different shear strength and hydraulic conductivity in horizontal and vertical directions (MacFarlane and Williams, 1974; Rycroft et al., 1975). Drained Histosols may develop pedogenic structure in the man-made aerobic zone, such as granular structure due to earthworm casts and blocks or prisms due to cycles of wetting and drying (Lee and Manoch, 1974).

The mineral soil material that occurs under the peat in Histosols is typically chemically reduced as a result of saturation by water and the abundance of organic matter above it. Where paludification has occurred, soil horizons of prior mineral soils may be buried beneath the peat. In some cases, the preceding mineral soil pedogenesis may facilitate paludification by producing low-permeability horizons such as placic horizons (Ugolini and Mann, 1979; Klinger, 1996). The depth to mineral soil underlying drained Histosols has been measured with some success by ground-penetrating radar (Shih and Doolittle, 1984; Collins et al., 1986; Sheng et al., 2004; Kettridge et al., 2008).

The morphology of Folists differs from that of other Histosols in that the peat layers are thinner and underlain by fragmental material or bedrock (Witty and Arnold, 1970; Everett, 1971; Lewis and Lavkulich, 1972). Folists drain freely and as a result are less saturated than other Histosols (not saturated for >30 cumulative days per year; Soil Survey Staff, 2010).

TABLE 33.6   Color of Organic Soil Material in Relation to Degree of Decomposition[a]

Degree of Decomposition	Median Munsell Color (Hue Value/Chroma)	N
Sapric	10 YR (2/1)	69
Hemic	10 YR (3/2)	49
Fibric	7.5 YR (3/2)	18

[a] Data include organic horizons of Histosols in the USDA Natural Resources Conservation Service, National Soil Survey Laboratory characterization database.

33.2.5 Micromorphology of Histosols

Micromorphological observations can provide a direct examination of the structural integrity of plant fragments and components in organic soil materials. Levesque and Dinel (1982) and Fox (1985) have summarized the characteristics of organic soil materials at various stages of decomposition (fibric, hemic, and sapric materials). Fibric materials mainly show unaltered or slightly altered plant tissues without appreciable darkening and with little organic fine material. The plant fragments, which are only slightly decomposed, appear to be loosely arranged with a porous and open structure. Partially decomposed (hemic) materials also possess a fibrous appearance, and most fragments show incomplete degradation. The development of brown or black colors in the plant tissues is typical. Fine organic material is also present intermixed with, or adhering to, the coarser fragments of plant tissue (Figure 33.7). Fecal pellets, which are evidence of faunal activity, can also be common. In the most highly decomposed (sapric) materials, organic fragments are sufficiently darkened and decomposed, so that identification of botanical origin is not possible. Fine organic material is usually the dominant component although faunal excrement is also common.

The effects of draining organic soils can sometimes be seen during microfabric examination. When an undrained sphagnum

FIGURE 33.7   Thin section showing organic material from hemic Oe horizon (80–88 cm) of a Typic Sulfihemist in a coastal marsh of Chesapeake Bay; the organic material reflects an intermediate degree of decomposition with some discernable plant structures and cell components intermixed with decomposed organic material; frame length = 5 mm; PPL.

peat profile in Ireland was compared with those which had been drained for between 10 and 100 years, the drained profiles had undergone substantial alteration and decomposition leaving little of the original tissue structures (Hammond and Collins, 1983). The fine organic material was dominant, showing some biological granulation (Pons, 1960; Lee, 1983). The change in microfabric materials directly corresponded to an increase in density of the material.

In another study, Lee and Manoch (1974) concluded that 50 years of drainage and cultivation of organic soils led to significant decomposition and the development of pedogenic structure in the subsoil, whereas in the lower portion of the profile where the soil remained saturated, sedimentary structure persisted and the peat was more highly fibrous and less decomposed. The activity of soil fauna in the drained portions of the soil contributed to biological granulation and the formation of two distinct types of surface horizons. The *moder* mostly consists of faunal excrement and usually forms in oligotrophic peats, while the *mull* is formed by an intense mixing and binding of organic with mineral particles by larger organisms such as earthworms, and usually forms in mesotrophic or eutrophic peats.

While not widely reported, following drainage and cultivation of organic soils, illuvial humus termed *humilluvic material* (Soil Survey Staff, 2011) may accumulate in the lower parts of acid organic soils (Van Heuveln et al., 1960). Both the lower pH of the oligotrophic peat and the disturbance by cultivation apparently contribute to the dispersion of the organic fraction, which can then be translocated within the soil, and accumulate within the lower horizons of the peat, at the peat–mineral soil contact, or within the underlying mineral soil material.

33.2.6 Classification of Histosols

The definition of *organic materials* for saturated soils requires a minimum of 12% OC if there is no clay, and a minimum of 18% OC if the soil contains 60% or more clay, with a sliding scale for intermediate textures. Those soils which are not saturated must contain at least 20% OC to be considered organic soil materials (Soil Survey Staff, 1998, 2011). For a soil to be classified as a Histosol, at least 40 cm of the upper 80 cm must comprise organic materials, and it must not have permafrost within 1 m of the surface. However, if the organic materials are especially low in density (<0.1 g cm⁻¹), then at least 60 cm of the upper 100 cm must be organic materials. Histosols may be buried by as much as 40 cm of overlying mineral soil materials and still be considered Histosols.

The types of differentiating characteristics used to discriminate between classes of soils at the various categorical levels are presented in Table 33.7. Basically, organic soils that are not saturated for extended periods are classified as Folists, while the saturated organic soils are differentiated according to the degree of decomposition of the organic materials in the subsurface tier (the zone approximating 40–100 cm) into Fibrists, Hemists, or Saprists. Within the United States, some 314 soil series have been established for Histosols. Table 33.8 shows the distribution of those 314 series among the classes of the various

TABLE 33.7 Criteria Utilized in the Classification of Histosols

Suborder	Great Group	Subgroup	Family
Degree of saturation with water	Soil temperature regime	Thickness of organic materials (Terric vs. Typic)	Particle size and mineralogy (used only for Terric subgroups or for those containing Limnic materials)
Degree of decomposition of the subsurface tier	Special components (sphagnum fibers, sulfidic materials or sulfuric horizon, humilluvic materials)	Underlying materials Special materials contained (Limnic) Intergrades to other great groups (Cryic and Sphagnic)	Reaction (pH in 0.01 M CaCl₂) Temperature regime Soil depth (only used if <50 cm deep)

TABLE 33.8 Number of Soil Series in the United States That Are Classified into Particular Taxonomic Groups

Suborders		Great Groups		Subgroups	
Formative Element	Number of Series	Formative Element	Number of Series	Formative Element	Number of Series
Fibrists	26	Cryo	58	Fluvaquentic	14
Folists	35	Haplo	198	Limnic	19
Hemists	82	Sphagno	7	Lithic	37
Saprists	171	Sulfi	17	Terric	124
		Udi	19	Typic	102
		Usti	7	Hemic	7
		Others	8	Others	11
Total	314	Total	314	Total	314

taxonomic categories. The number of series that exist within a particular class may result from many factors and should not be taken to represent the areal extent of those soils. Histosols classified at the family level are differentiated into classes based upon particle size and mineralogy (used only for terric subgroups or for those containing limnic materials), reaction (pH in 0.01 M CaCl$_2$), temperature regime, and soil depth (only used if <50 cm deep).

33.2.7 Biological and Chemical Properties

Organic carbon contents are generally higher and nitrogen contents lower in less-decomposed peats than highly decomposed peats (Table 33.9). As a result, C:N ratios are generally higher for less-decomposed peats (Table 33.9; Lee et al., 1988). The ash content (mineral component) is also higher for more highly decomposed peats (Table 33.9; Lévesque et al., 1980; Lee et al., 1988). Thus, more highly decomposed peats are generally more fertile than less-decomposed peats. Some drained, sapric peats may supply nitrogen in excess of crop requirements without fertilization. However, nutrients derived from minerals, such as phosphorus, potassium, and most micronutrients, are frequently deficient in Histosols (Lucas, 1982; Yefimov, 1986).

The chemical and physical properties of peats are also related to their botanical composition. Peats derived mostly from *Sphagnum* mosses tend to be more acid, less decomposed, contain less ash, have lower cation exchange capacity (CEC), and have lower bulk density than woody peats; sedge peats typically have intermediate properties (Farnham and Finney, 1965; Lévesque et al., 1980).

33.2.7.1 Soil Carbon

Interest in the global C cycle has focused attention on the high proportion (3/4) of terrestrial C stored in soils (Lal et al., 1995). As a group, wetland soils maintain a disproportionately high level of soil carbon, and Histosols, which are composed largely of organic matter, clearly store the largest quantities of soil

carbon. Although peatlands only occupy approximately 4% of the global land surface, they store about 30% of the global soil C (Lavoie et al., 2005). While typical agricultural soils may contain between 2 and 10 kg C m^{-2}, reported values for Histosols typically are an order of magnitude greater, with some values >200 kg C m^{-2} (Table 33.10). The quantity of C stored in some very deep Histosols is undoubtedly even higher.

Histosols are very dynamic and may be particularly significant in the overall terrestrial C budget. Many Histosols continue to sequester C at significant rates. This is particularly true for soils of coastal marshes, where rising sea level continues to power the engines of marsh accretion and C storage. Therefore, Histosols are generally viewed as a significant C sink although some studies indicate that climate change either through increased frequency and intensity of fire (Turetsky et al., 2002) or through elevated decomposition as a result of rising temperatures (Billett et al., 2004) is possibly switching some peatlands to sources of C to the atmosphere. In addition, if Histosols are drained or in some other way exposed to an aerobic environment, they may begin to oxidize and yield large quantities of C to the atmosphere (e.g., Nykanen et al., 1997), although other studies indicate that drainage can increase C sequestration as a result of greater aboveground and belowground productivity (e.g., Minkkinen and Laine, 1998).

Most of the discussion of possible global warming and greenhouse gas emission has focused on rising levels of CO$_2$ in the atmosphere. Methane (CH$_4$), however, is 32 times more efficient than CO$_2$ in trapping infrared radiation. Because many Histosols are strongly reducing (low Eh), they represent an ideal environment for the formation of CH$_4$. In systems where SO$_4^{2-}$ is more abundant in the soil solution, such as in coastal or estuarine environments, sulfate reduction is favored over methanogenesis and methane production may be more limited (Bartlett et al., 1987; Widdell, 1988; Dise and Verry, 2001). However, in many freshwater or inland areas, Histosols may be the locus of significant methane emission to the atmosphere (Table 33.11). Minerotrophic fens have higher methane

TABLE 33.9 Chemical Properties of Organic Soil Material as Related to Degree of Decomposition: Mean (Standard Deviation, N)[a]

Property	Sapric	Hemic	Fibric	AOV Probability[b]
Organic carbon, g kg^{-1}	313 (128, 129)	347 (135, 61)	372 (130, 26)	0.055
Total nitrogen, g kg^{-1}	18 (9, 131)	16 (6, 54)	14 (5, 23)	0.058
C:N ratio	21 (10, 113)	25 (11, 48)	27.5 (10, 20)	0.007
CEC, cmol kg^{-1}	101 (44, 129)	88 (41, 61)	83 (33, 28)	0.046
CEC, cmol L^{-1}	76 (42, 28)	44 (25, 25)	21 (2, 5)	0.000
Ash, g kg^{-1}%	250 (nd)	178 (110, nd)	100 (50, nd)	nd
pH	5.1 (1.2, 143)	4.9 (1.2, 59)	4.5 (1.2, 27)	0.024
Al, mol Al mol^{-1} TEA[c]	0.074 (0.076, 49)	0.022 (0.024, 23)	0.038 (0.050, 16)	0.004

[a] Data (except for ash) is for all organic horizons of Histosols in the USDA Natural Resources Conservation Service, National Soil Survey Laboratory characterization database. Ash content is taken from Lee et al. (1988; data for 1300 samples of Wisconsin Histosols, nd—no data).

[b] F-test probability from one-way analysis of variance (AOV) between fibric, hemic, and sapric peats.

[c] KCl extractable Al divided by total NH$_4$ acetate extractable acidity.

TABLE 33.10 Carbon Storage Values for Organic Soils

Site Characteristics	Location	Carbon Accumulation Rate (kg m⁻² year⁻¹)	Quantity of Stored C (kg m⁻²)	References
10 year averages	West Siberia	0.021		Golovatskaya and
Pine peatland		0.11		Dyukarev (2009)
Stunted pine peatland		0.10		
Sedge fen				
Fen, last 50 years	Alberta, Canada	0.14–0.25		Vitt et al. (2009)
Various peatland types	Alberta, Canada		53–165, mean = 129	Beilman et al. (2008)
Bog, last 100 years	Western Canada	0.09		Turetsky et al. (2007)
Fen 1	Saskatchewan, Canada		29–210	Robinson (2006)
Fen 2			20–120	
Rich fen	Northwest territories, Canada	0.014		Robinson (2006)
Poor fen		0.018		
Bog		0.019		
Rich fen	Western Canada	0.025		Yu (2006)
Tropical peatlands	Micronesia	0.3		Chimner and Ewel (2005)
Sloping, open peatland	British Columbia, Canada	0.007–0.039		Asada and Warner (2004)
Bog	Southern Sweden	0.060–0.072		Belyea and Malmer (2004)
River valley fen	West Siberia	0.069		Borren et al. (2004)
Raised bog		0.033		
Forested fen		0.034		
Bog		0.027		
Flow-through fen		0.021		
Flow-through fen		0.020		
Bog		0.019		
Bog		0.040		
Variety peatland types	West Siberia		<30 to >300	Sheng et al. (2004)
Various peatland types	North Dakota, USA	0.038		Gorham et al. (2003)
	Minnesota, USA	0.021–0.038		
	Quebec, Canada	0.022–0.028		
	Maine, USA	0.022–0.030		
	New Brunswick, Canada	0.017–0.019		
	Nova Scotia, Canada	0.019–0.041		
	Newfoundland, Canada	0.008–0.016		
	Alaska, USA			
Raised bog	Southern Sweden			Malmer and Wallen (1999)
Hummock		0.05–0.18		
Hollow		0.03–0.15		
Last 100 years	Eastern USA			Wieder et al. (1994)
Minnesota bog		0.16		
Pennsylvania fen		0.14		
Maryland fen		0.15		
West Virginia bog 1		0.15		
West Virginia bog 2		0.18		
Coastal marsh	Los Angeles, USA			Derived from Nyman
Fresh		0.17–0.22		and DeLaune (1991)
Brackish		0.17–0.27		
Saline		0.21–0.22		
Coastal marsh	Chesapeake Bay, MD	0.12–0.42		Derived from Kearney
Brackish				and Stevenson (1991)
Coastal marsh	Chesapeake Bay, MD		59 (range 18–166)	Derived from Griffin
Brackish				and Rabenhorst (1989)
Barataria Basin, coastal marsh	Louisiana, USA	0.18–0.30		Smith et al. (1983)
Coastal marshes	Atlantic and Gulf Coasts, USA		64 (range 9–191)	Rabenhorst (1995)
Sphagnum peat	Slave Lake, Canada	0.014–0.035		Kuhry and Vitt (1996)
Based on 138 basal ¹⁴C dates	Subarctic and Boreal Canada	0.023–0.029		Gorham (1991)

TABLE 33.11 Reported Fluxes of CO_2 and CH_4 Emissions from Histosols

Location	Site Details	Notes	CO_2 Emission Rate (mmol m^{-2} day^{-1})		CH_4 Emission Rate (mmol m^{-2} day^{-1})		References
			Mean	Range	Mean	Range	
West Siberia	10 year averages	Growing season					Golovatskaya and Dyukarev (2009)
	Pine peatland		45.0		0.030		
	Stunted pine peatland		28.5		0.045		
	Sedge fen		29.9		3.30		
Quebec, Canada	Poor fen—control	Growing season		42–250		0.95–3.55	Strack and Waddington (2007)
	Poor fen—w/water table drawdown			46–242		0.63–6.40	
Western Canada	Bog	Transplant experiment	55		0.4		Turetsky et al. (2007)
Northern England	Acidic blanket peat	Growing season				0.15–4.2	Ward et al. (2007)
	Control			5.4–327		(range for	
	Burned			10.9–491		all	
	Grazed			5.4–436		treatments)	
Micronesia	Forested peatland	Annual	198				Chimner and Ewel (2004)
	Cultivated for taro		110				
New York, USA	Conifer/maple peatland	Annual	103	8.6–216	0.05	−0.17–0.69	Coles and Yavitt (2004)
Ontario, Canada	Mesocosms—controlled temperature, water table, and humidity	Net production reported		6.1–602		−0.25–1.1	Blodau and Moore (2003)
Quebec, Canada	Poor fen—control	Growing season			8.9		Strack et al. (2004)
	Poor fen—w/water table drawdown				4.0		
Minnesota, USA	Poor fen—control	Growing season			15.1	6.2–31.9	Dise and Verry (2001)
	Ammonium nitrate added				16.0	7.5–30.0	
	Ammonium sulfate added				10.2	5.0–15.6	
Minnesota, USA	Bog	Mesocosms	275–2500			1.3–41.3	Updegraff et al. (2001)
	Fen	Growing season					
Quebec, Canada	Gatineau Park			1.97–7.24		1.15–2.18	Buttler et al. (1994)
Wales	Peat monoliths	Laboratory study	14.7	9.6–21.0	14.4	4.7–34.4	Freeman et al. (1993)
Finland	Natural fen	Annual	11.3		3.49		Nykanen et al. (1995)
	Drained fen		30.8		0.03		
Finland	Ombrotrophic	12 C	88.2	42.5–141.3			Silvola et al. (1996)
Alaska, USA		Summer measurements			9.2		After Gorham (1991), after Crill et al. (1988)
Boreal Canada	Swamp ($n = 20$)	Annual averages			0.21		Derived from Moore and Roulet (1995)
	Fen ($n = 6$)				0.57		
	Bog ($n = 13$)				0.39		
Canada (lab study)	Bog flooded	19–23 C	0.005		0.012		Derived from Moore and Knowles (1989)
	Bog drained		0.19		0.006		
	Fen flooded		0.009		0.58		
	Fen drained		0.14		0.025		
Alaska	Moist tundra	August			0.3		Derived from Sebacher et al. (1986)
	Waterlogged tundra				7.4		
	Wet meadows				2.5		
	Alpine fen				18		
Northern Sweden	Ombrotrophic bog	Summer					Svensson and Rosswall (1984)
	Hummocks		10.12		0.05		
	Between hummocks		14.92		0.14		
	Shallow depressions		11.61		0.77		
	Deeper depressions		12.45		1.21		
	Ombrominerotrophic		12.49		2.73		
	Minerotrophic fen		11.48		16.89		

TABLE 33.11 (continued)   Reported Fluxes of CO$_2$ and CH$_4$ Emissions from Histosols

Location	Site Details	Notes	CO$_2$ Emission Rate (mmol m^{-2} day^{-1}) Mean	Range	CH$_4$ Emission Rate (mmol m^{-2} day^{-1}) Mean	Range	References
Minnesota	Bog	Sampled during			8.2	1.2–29.2	After Harriss et al. (1985)
	Fen	August			0.25	0.19–0.31	
Georgia, USA		During midsummer			6.6		After Gorham (1991), after Crill et al. (1988)
West Virginia, USA	Mountain bog	During midsummer			11.7		After Gorham (1991), after Crill et al. (1988)
Minnesota, USA	Forest bog	Summer measurements			3.6	0.5–33	After Crill et al. (1988), after Mitsch and Wu (1995)
	Forest fen				6.7	3.2–12	
	Open bog				14	0.9–41	
	Neutral fen				15	7.1–33	
	Acid fen				4.8		
Minnesota, USA	Open poor fen	Winter measurements			3.0		Dise (1992)
	Open bog				0.7		
	Forest bog hollow				0.8		
	Hummock				0.3		
Virginia	Coastal marsh	Summer			0.45	0.13–0.82	Derived from Bartlett et al. (1985)
	York river	Creek Bank			0.29	0.05–0.87	
	Chesapeake Bay estuary	Short Spartina High marsh			0.09	0–0.36	
Virginia	Coastal marsh	Summer					Derived from Bartlett et al. (1987)
	Tidal Creek in Chesapeake Bay estuary	Low salinity			11	5–16	
		Moderate salinity			7.5	4–11	
		High salinity			2	0.5–2.5	
West Virginia	Appalachian bog		127	75–250	17.0	0–53	Wieder et al. (1990)
Maryland	Appalachian bog		152	100–250	4.4	0–12	Wieder et al. (1990)
Louisiana, USA	Barataria basin and Coastal marsh	Annual averages		41–141			Smith et al. (1983)
Florida, USA		During midsummer			6.0		After Gorham (1991), after Crill et al. (1988)
Malaysia	Ombrotrophic bog		170				Wosten et al. (1997)
Malaysia	Drained and cultivated peatland			139–727			Murayama and Bakar (1996)

emissions than ombrotrophic bogs (Dise, 1993). Methane emission from Histosols seems to be directly related to the location of the water table, with greater generation when soils are saturated to or above the surface (Updegraff et al., 2001; Strack et al., 2004). Greater methane emissions occur when soils are saturated. If aerobic zone exists in the profile, soil microbes will utilize the methane as it passes through on its way to the soil surface. Carbon dioxide emissions, which are produced by oxidation of soil organic matter, have been found to be both greater (Nykanen et al., 1997) and unchanged (Updegraff et al., 2001; Strack and Waddington, 2007) following the lowering of water tables.

33.2.7.2 Sulfides

The biogeochemical environment in which Histosols form can also be conducive to the formation of sulfides. The occurrence of iron sulfides can, in some circumstances, lead to the generation of extreme acidity and acid sulfate soils (Van Breemen, 1982).

Sulfate reduction generally requires the presence of organic matter, which serves as an energy source, low redox potentials, sulfate that functions as an electron acceptor, and sulfate reducing bacteria (Rickard, 1973). If sulfide is formed in the presence of a reactive iron source, then such minerals as pyrite (FeS$_2$) can form. The C source and anaerobic conditions are almost always present in Histosols, but sulfate (SO$_4^{2-}$) levels may vary dramatically among environments. Many inland Histosols receive SO$_4^{2-}$ only in small amounts as atmospheric deposition, and under these circumstances, sulfidization (Rabenhorst and James, 1992) is insignificant, and most of the S in those soils is bound in organic S forms (Novak and Wieder, 1992). Some inland peats have developed acid sulfate conditions, although usually they are associated with deposits of coprogeneous earth (Lucas, 1982). Sulfate reduction is common in coastal Histosols, which contain an abundance of SO$_4^{2-}$ from sea water. Extensive areas of these Sulfihemist soils have been identified along the Atlantic coast of the United States. The distribution of pyrite within

coastal Histosols can be highly variable and is often related to microsite differences in the availability of either reactive iron or sulfide (Rabenhorst and Haering, 1989).

33.2.7.3 Acidity and Base Saturation

In bog Histosols, where inputs of bases are minimal because soil water is derived from rainfall that has never contacted mineral soil, base saturation is low (at least in the rooting zone) and the organic acids produced by partial decomposition of organic matter typically buffer soil water pH near 4; the soil pH in $CaCl_2$ is typically 3–4 (Figures 33.8 and 33.9; Gorham et al., 1985; Urban, 1987). Minerotrophic (fen) peats have higher base saturation,
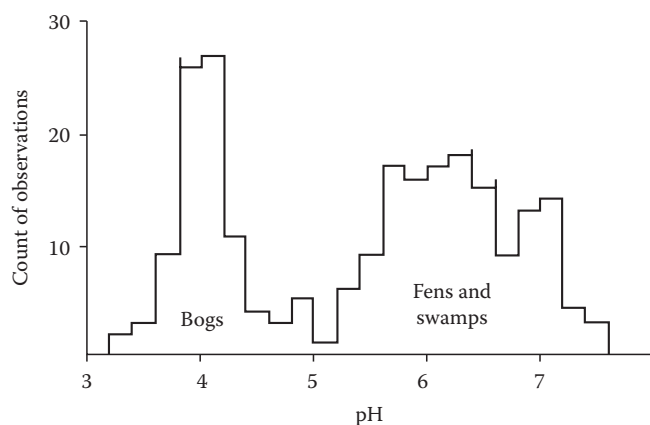
FIGURE 33.8 Frequency distribution of pH in peatland surface water for 232 sites in Minnesota. Samples with surface water pH near 4 are from bogs and those with pH above 5 are from fens (unforested) and swamps (forested). (From Swanson, D.K., and D.F. Grigal. 1989. Vegetation indicators of organic soil properties in Minnesota. Soil Sci. Soc. Am. J. 53:491–495.)
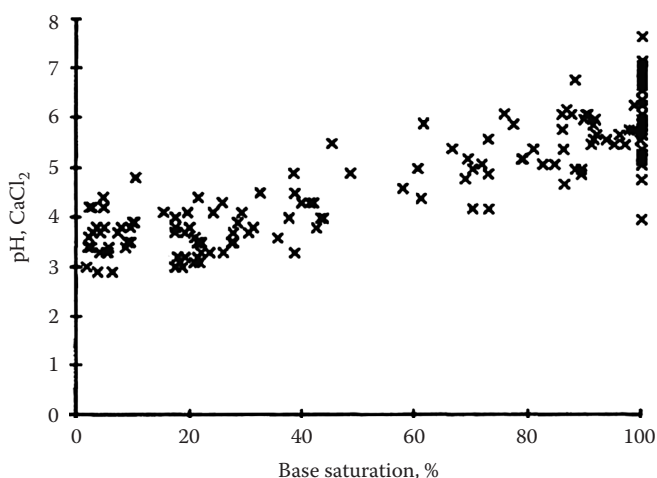
FIGURE 33.9 Relationship between base saturation and pH of organic soil horizons of Histosols. Peat samples with pH in $CaCl_2$ of less than 4.5 (presumably deposited in bogs) generally have less than 50% base saturation. Peat samples with pH greater than 4.5 (deposited in fens and swamps) typically have high base saturation. Data are for all organic horizons of Histosols in the USDA Natural Resources Conservation Service, National Soil Survey Laboratory characterization database.

and pH in water is generally above 5 ($CaCl_2$ pH above 4.5). The pH of minerotrophic peats is buffered by cation exchange within the soil (Bloom et al., 1983) or by carbonates if they are present. The aluminum ion, which is derived from silicate minerals, comprises a small proportion of the total acidity in most Histosols because of its low solubility at pH > 5.5 (Table 33.9). Less-decomposed peats generally have lower pH than more highly decomposed peats (Table 33.9; Lee et al., 1988).

The distinction between euic and dysic reaction classes at the family level in *Soil Taxonomy* separates the highly acid, bog peats (dysic) from less acid fen peats (euic) (Farnham and Finney, 1965). Because a pH (in $CaCl_2$) greater than 4.5 anywhere in the control section (i.e., anywhere within 130 or 160 cm of the surface) places the soil into the euic class, a bog with a highly acidic near-surface rooting zone may classify as euic rather than dysic due to the presence of higher pH horizons at depth in the soil.

Histosols which contain sulfide minerals such as pyrite have the potential to develop extreme acidity. Under saturated and anaerobic conditions, such sulfide bearing soils have circumneutral pH. If it is drained, dredged, or in some other way exposed to oxidizing conditions, the soil will undergo acid sulfate weathering, a process that is the result of sulfide mineral oxidation and the concomitant production of sulfuric acid (Van Breemen, 1982). The oxidation of pyrite can lead both to the extreme acidification of the soil (pH < 3.5) and also to the generation of acidity, which can be moved offsite through mobilization of acid generating soluble salts such as $FeSO_4$.

33.2.7.4 Cation Exchange Capacity

The CEC of Histosols is quite high as a result of the high cation exchange of organic matter (Table 33.9). The CEC is higher for more highly decomposed peats than fibric peats (Table 33.9; Lévesque et al., 1980). The difference between the CEC of sapric and fibric peats is even more dramatic if the CEC is expressed on a volume rather than mass basis, as a result of the very low bulk densities of fibric materials (Table 33.9). The CEC per unit soil volume is a more useful measure than CEC per unit soil mass in Histosols, because their bulk densities are very low and because roots occupy a volume rather than mass of soil. The CEC per unit volume of fibric peats, near 20 $cmol_c$ L^{-1}, is comparable to that of most mineral soils (5–20 $cmol_c$ L^{-1}) assuming mineral soil bulk density of 1.0–1.5 kg L^{-1} (Holmgren et al., 1993). Even when expressed per unit volume, the average CEC of sapric and hemic peats (Table 33.9) is much higher than the 5–20 $cmol_c$ L^{-1} of most mineral soils.

33.2.8 Physical Properties

33.2.8.1 Bulk Density

The physical properties of Histosols differ greatly from those of mineral soils. Bulk densities for organic soil materials generally are quite low, ranging from as little as 0.02 up to 0.8 g cm^{-3} (Table 33.12). Bulk density is related to the degree of decomposition

TABLE 33.12 Physical and Hydraulic Properties of Organic Soils

Location	Site Characteristics	Bulk Density (g cc⁻¹)	Hydraulic Conductivity (10⁻⁵ cm s⁻¹)	References
Alberta Canada	Rich fens	0.04–0.12		Vitt et al. (2009)
Western Canada	Bog	0.056		Turetsky et al. (2007)
Manitoba, Saskatchewan and Alberta, Canada	Various peatland types	0.07–0.26		Bauer et al. (2006)
Northwest Territories, Canada	Rich fen	0.03–0.14		Robinson (2006)
	Bog	0.08–0.20		
Western Canada	Rich fen	0.18		Yu (2006)
Micronesia	Tropical peatlands	0.11–0.13		Chimner and Ewel (2005)
West Siberia	Various peatland types	0.05–0.41		Sheng et al. (2004)
England	Blanket peats (fens)	0.15–0.27	0.01–1.04, mean = 0.24	Holden and Burt (2003)
New Zealand	Bog	0.06		Schipper and McLeod (2002)
Poland	Various peatland types	0.07–0.60		Bogacz (2000)
Finland	Drained and harvested fens		0.4–60	Klove (2000)
Southern Sweden	Bog			Malmer and Wallen (1999)
	Lichen hummock	0.18		
	Sphagnum hummock	0.27		
	Sphagnum lawn	0.27		
Finland	Various peatland types	0.12–0.16		Minkkinen and Laine (1998)
Southeastern Norway	Bog	0.2–0.8		Ohlson and Okland (1998)
Northern Minnesota	Undecomposed		3,810–15,000, mean = 8,650	Boelter (1965)
	Partially decomposed		13.9–132, mean = 63	
	Decomposed		0.9–15 mean = 5.1	
Wyoming	Mountain bog 46 cm	0.16–0.22	0.0277	Sturges (1968)
	91 cm		0.0185	
Ottawa and St. Lawrence River Valleys, Canada	Swamp and bog			Mathur and Levesque (1985)
	0–60 cm		624	
	0–100		366	
	0–125		269	
Northern Minnesota	Lost river peatland	0.06–0.14		Chason and Siegel (1986)
	Bog		25–560	
	Fen margin		150–2,600	
	Spring fen		67–1,600	
Quebec, Canada	Gatineau Park	0.03–0.10		Buttler et al. (1994)
Eastern New Brunswick, Canada	Sphagnum peat from raised bogs			Korpijaako and Radforth (1972)
	Van post scale 1–2	0.02–0.08	90–175	
	3–4	0.03–0.10	6.9–56	
	5–6	0.06–0.11	1.4–17	
	7–8	0.09–0.13	0.14–2.8	
Wisconsin, USA	Fibric	0.13		Lee et al. (1988)
	Femic	0.17		
	Sapric	0.20, 0.24		
Minnesota	Bog			Gafni and Brooks (1990)
	0–10 cm Von Post 1		23,495	
	10–20 cm Von Post 2–3		7,697	
	20–30 cm Von Post 4–5		5,498	
	30–40 cm Von Post 5–6		799	
	40–50 cm Von Post 5–7		995	
	Fen			
	0–10 cm		31,597	
	10–20 cm		5,000	
	20–30 cm		1,400	

of the organic materials with bulk density generally increasing as the materials become more highly decomposed (Figure 33.10). Because organic materials contain varying amounts of mineral matter, this also can affect the bulk density of organic soil horizons. For coastal marsh peat, the organic matter content is generally about twice the content of organic C, with the remainder representing the mineral fraction. Figure 33.11 illustrates the relationship between the mineral content (roughly the difference remaining from twice the OC content) and the bulk density for Oe and Oa horizons from Sulfihemists along the Atlantic Coast (Griffin and Rabenhorst, 1989).

33.2.8.2 Porosity, Hydraulic Conductivity, and Water Retention

Histosols have very high porosity levels, which can reach over 80% as shown by the water content at saturation (Figure 33.12; Boelter, 1969). The high porosity and low bulk density of organic
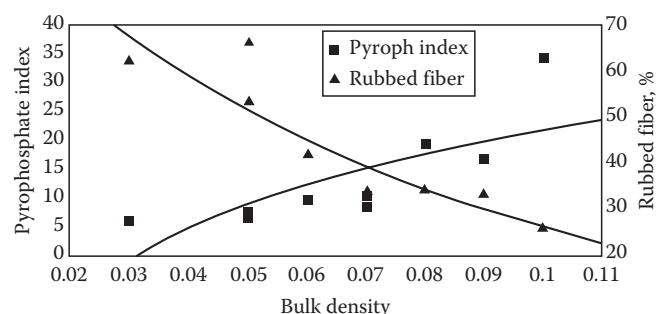
FIGURE 33.10 Relationship between bulk density of organic materials and indices of decomposition such as rubbed fiber content or pyrophosphate index. (Derived from Buttler, A., H. Dinel, and P.E.M. Levesque. 1994. Effects of physical, chemical and botanical characteristics of peat on carbon gas fluxes. Soil Sci. 158:365–374.)
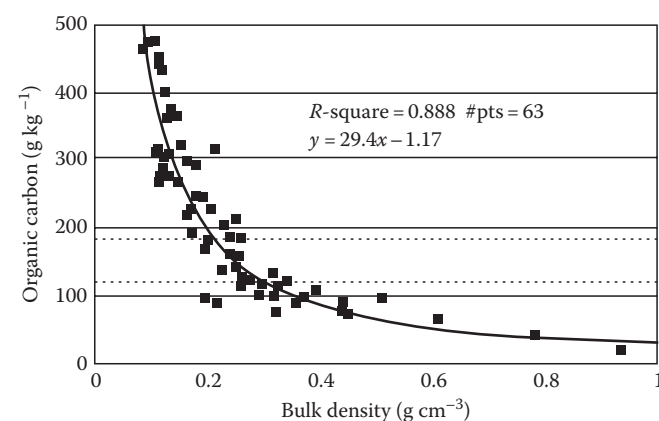
FIGURE 33.11 The effect of mineral content (roughly the difference remaining from 2× the OC content) on the bulk density of Oe and Oa horizons from Sulfihemists along the Atlantic Coast. Dashed lines represent 12% and 18% OC, which is necessary for soil materials to be considered organic. (Derived from Griffin, T.M., and M.C. Rabenhorst. 1989. Processes and rates of pedogenesis in some Maryland tidal marsh soils. Soil Sci. Soc. Am. J. 53:862–870.)
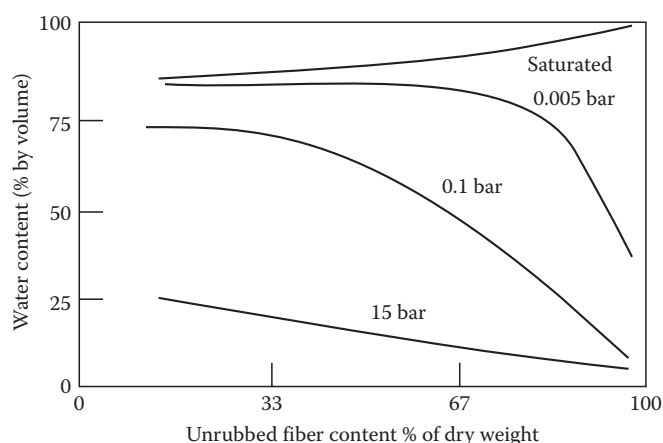
FIGURE 33.12 Volumetric soil water content as a function of unrubbed fiber content at various moisture tensions. (After Boelter, D.H. 1969. Physical properties of peats as related to degree of decomposition. Soil Sci. Soc. Am. J. 33:606–609.)

soils would lead one to expect high rates of water transmission through Histosols. Weakly or undecomposed peat often has a fairly high hydraulic conductivity; however, as the material becomes more decomposed, the hydraulic conductivity decreases (Table 33.12). The hydraulic conductivity of sapric and some hemic peats is quite low, comparable to that of clay (i.e., 10^{-5} m s^{-1} or less). Because the peat in the lower part of Histosol profiles is often sapric or hemic, little deep percolation occurs and water tends to evaporate or move laterally through the less decomposed, surface horizons (Ivanov, 1981).

Water retention in Histosols is also closely related to the degree of decomposition (Figure 33.12). The large pores of slightly decomposed peats drain readily at low suction. In contrast, the fine pores of more well-decomposed peats retain considerable water at low suctions (Boelter, 1974). The water retention difference (water content at 0.01 MPa minus that at 1.5 MPa suction; approximates plant-available soil water capacity) of most Histosols is very high, approaching 50% in some peats and exceeding 25% in all others except the least-decomposed peats (Figure 33.12).

33.2.9 Utilization and Management of Histosols

33.2.9.1 Interest in Preservation as Coastal and Nontidal Wetlands

Apart from any benefits which may be achieved from managed systems, such as commercial forestry, grazing, or intensive agriculture, Histosols perform a number of environmental and ecological functions. Because most Histosols occur in wetland environments, they typically provide wetland functions including wildlife habitat, floodwater control, groundwater recharge, nutrient and biogeochemical cycling, ion sorption, purification of surface and shallow groundwater, as well as functioning as an important sink for terrestrial C. The importance of peatlands as

paleoenvironmental and archaeological archives has also been documented (Godwin, 1981). Throughout the years, the value society has placed on these functions has been minimal, and the Histosols of peatlands and coastal marshes have been exploited and extensive areas have been destroyed. More recently, people have recognized that the functions which Histosols perform in a natural setting have significant benefit for society. Thus, legislation has been passed in the United States and elsewhere to preserve Histosols and other wetlands.

33.2.9.2 Histosols as Agricultural Resources

As was mentioned earlier, Histosols are utilized as important agricultural soils in many areas, so long as water tables can be effectively managed. For example, in Japan, there are over 70,000 ha of rice that is grown on peatlands. Peatlands in a number of tropical countries such as Cuba, Guyana, Malaya, and Indonesia have been reclaimed from mangroves and are used for the production of sugarcane (Moore and Bellamy, 1974). Within the United States, there has also been extensive growth of vegetables on peatlands in the Northern United States (such as Michigan), and there has also been extensive agricultural development in the Everglades region of Florida. The agricultural use of Histosols presents some special challenges regarding nutrient management and fertility, but the most significant problem is the high water tables requiring drainage.

33.2.9.3 Impacts of Drainage

Where Histosols have been converted for higher intensity land uses, such as in agriculture, horticulture, or silviculture, they are almost always drained to lower the water table. Such drainage results in a number of short-term effects, such as shrinkage and consolidation due to desiccation, the loss of the buoyant force of groundwater, and compaction. There is also ongoing consolidation and soil alteration caused by the enhanced decomposition of the organic materials following the shift from an anaerobic to an oxidizing regime (Stephens and Speir, 1969; Minkkinen and Laine, 1998). Reported rates of long-term peat subsidence range up to 10 cm year^{-1} but most reports are in the range of 2–5 cm year^{-1} (Table 33.13). This consolidation is accompanied by changes in the physical properties of the peat, including higher bulk densities and lower moisture contents.

33.2.9.4 Histosols as Energy Resources

Peat is mined mainly in northern Europe and used as fuel (Table 33.14). In addition, there is extensive mining and export of peat as a horticultural amendment. Important examples of the latter uses include ingredients for potting soils, and mixed fertilizers, components of mushroom beds, as a seed inoculant, as a material for packing of flowers and other plants as well as a general soil amendment to increase organic matter content in gardens, golf courses, etc. Approximately 635,000 metric tons of peat are utilized annually for these types of uses in the United States (Public communication, 2009c).

33.2.9.5 Engineering Properties

Histosols are notorious for their low strength and great compressibility, which make them poor foundation materials for

TABLE 33.13 Reported Rates of Subsidence after Drainage of Organic Soils

Location	Site Characteristics	Subsidence Rate (cm year^{-1})	Length of Record (year)	References
New Zealand	Bog	3.4	40	Shipper and McLeod (2002)
Finland	Various peatland types	0.4	60	Minkkinen and Laine (1998)
Florida, USA	Everglades, muck, and peat	3.2	41	Thomas (1965)
Hunts, England	Holme marsh	3.4	103	Nickolson (1951)
California, USA	Sacramento–San Joaquin delta, muck, and peat 1–3 m deep	6.4–9.8	26	Weir (1950)
Indiana, USA		3.0	30+	Ellis and Morris (1945)
Northern Indiana, USA	Muck	1.1–3.0 dependent on WT level	6	Jongedyk et al. (1950)
Michigan, USA		0–3.6	5	Davis and Engberg (1955) (after Thomas, 1965)
Southern Ontario, Canada	Holland marsh and deep loose muck	3.3	19	Mirza and Irwin (1964)
Minnesota, USA		5.1	3	Row (1940)
Florida, USA	Everglades	3	55	Stephens and Speir (1969)
Florida, USA	Everglades	2.3–1.8	20	Shih et al. (1981)
California USA	Sacramento–San Joaquin delta	2.3	78	Rojstaczer and Deverel (1995)
Quebec, Canada		2.1	38	Parent et al. (1982)
Malaysia	Ombrotrophic bog	2	21	Wosten et al. (1997)

TABLE 33.14 Peat Mining by Country, 2007 (in Thousands of Metric Tons)[a]

Country	Horticultural Use	Fuel Use	Total
Argentina	15	0	15
Australia	nd	nd	7
Belarus	100	2400	2500
Burundi	0	10	10
Canada	1250	0	1250
Denmark	300	0	300
Estonia	1300	600	1900
Finland	900	8200	9100
France	200	0	200
Germany	120	0	120
Ireland	500	3800	4300
Latvia	nd[b]	nd	1000
Lithuania	nd	nd	307
Moldova	0	475	475
New Zealand	27	0	27
Norway	30	0	30
Poland	500	0	500
Russia	nd	nd	1300
Spain	nd	nd	60
Sweden	380	900	1280
Ukraine	nd	nd	395
United States	635	0	635

[a] Data from Public communication (2009c).
[b] No data.

roads, buildings, and other structures. Compression and settlement of peats may continue for years after loading (MacFarlane, 1969; MacFarlane and Williams, 1974; Dhowian and Edil, 1980). Special engineering techniques, such as removal of the peat or precompression before construction, are thus required. The high water content and acidity of most Histosols also make corrosion of concrete and metal structures a potential problem (MacFarlane, 1969; MacFarlane and Williams, 1974). While Histosols are poor foundation materials, their high porosity and great adsorption capacity make them very useful for treatment of wastewater. Peats have potential for treatment of municipal effluent and removal of heavy metals and hydrocarbon pollutants from wastewater (Malterer et al., 1996).

References

Aaby, B. 1986. Paleoecological studies of mires, p. 145–164. *In* B.E. Berglund (ed.) Handbook of holocene palaeoecology and palaeohydrology. John Wiley & Sons, New York.

Almquist-Jacobson, H., and D.R. Foster. 1995. Toward an integrated model for raised-bog development: Theory and field evidence. Ecology 76:2503–2516.

Anderson, J.A.R. 1983. The tropical peat swamps of western Malesia, p. 181–199. *In* A.J.P. Gore (ed.) Mires—Swamp, bog fen, and moor. V. 4B Regional studies. Elsevier, Amsterdam, the Netherlands.

Andrus, R.E. 1986. Some aspects of *Sphagnum* ecology. Can. J. Bot. 64:416–426.

Asada, T., and B.G. Warner. 2004. Surface peat mass and carbon balance in a hypermaritime peatland. Soil Sci. Soc. Am. J. 69:549–562.

Bartlett, K.B., D.S. Bartlett, R.C. Harriss, and D.I. Sebacher. 1987. Methane emissions along a salt marsh salinity gradient. Biogeochemistry 4:183–202.

Bartlett, K.B., R.C. Harriss, and D. Sebacher. 1985. Methane flux from coastal salt marshes. J. Geophys. Res. 90:5710–5720.

Bauer, I.E., J.S. Bhatti, K.J. Cash, C. Tarnocai, and S.D. Robinson. 2006. Developing statistical models to estimate the carbon density of organic soils. Can. J. Soil Sci. 86:295–304.

Beilman, D.W., D.H. Vitt, J.S. Bhatti, and S. Forests. 2008. Peat carbon stocks in the southern Mackenzie River Basin: Uncertainties revealed in high-resolution case study. Global Change Biol. 14:1–12.

Belyea, L.R., and N. Malmer. 2004. Carbon sequestration in peatland: Patterns and mechanisms of response to climate change. Global Change Biol. 10:1043–1052.

Bennike, O. 1992. Paleoecology and paleoclimatology of a late Holocene peat deposit from Broendevinsskaer, Central West Greenland. Arct. Alp. Res. 24:249–252.

Billett, M.F., S.M. Palmer, D. Hope, C. Deacon, R. Storeton-West, K.J. Hargreaves, C. Flechard, and D. Fowler. 2004. Linking land-atmosphere-stream carbon fluxes in a lowland peatland system. Global Biogeochem. Cycles 18:GB1024.

Birks, H.J.B., and H.H. Birks. 1980. Quaternary paleoecology. Edward Arnold, London, U.K.

Blodau, C., and T.R. Moore. 2003. Experimental response of peatland carbon dynamics to a water table fluctuation. Aquat. Sci. 65:47–62.

Bloom, P.R., W.E. Elder, and J. Grava. 1983. Chemistry and mineralogy of mineral elements in Minnesota histosols, p. 29–43. *In* Proc. 26th Ann. Meet. Manitoba Society of Soil Sci. Winnipeg, Manitoba, Canada.

Bloom, A.L., and M. Stuvier. 1963. Submergence of the Connecticut coast. Science 139:333–334.

Boelter, D.H. 1965. Hydraulic conductivity of peats. Soil Sci. 100:227–231.

Boelter, D.H. 1969. Physical properties of peats as related to degree of decomposition. Soil Sci. Soc. Am. J. 33:606–609.

Boelter, D.H. 1974. The hydrologic characteristics of undrained organic soils in the Lake states, p. 33–46. *In* A.R. Aandahl (ed.) Histosols: Their characteristic, classification, and use. SSSA Special Publication No. 6. SSSA, Madison, WI.

Bogacz, A. 2000. Physical properties of organic soils in the Stolowe Mountains National Park (Poland). Suo 51:105–113.

Borren, W., W. Bleuten, and E.D. Lapshina. 2004. Holocene peat and carbon accumulation rates in the southern taiga of western Siberia. Quat. Res. 61:42–51.

Botch, M.S., and V.V. Masing. 1983. Mire ecosystems in the USSR, p. 95–152. *In* A.J.P. Gore (ed.) Ecosystems of the world, 4B, Mires: Swamp, bog, fen and moor, regional studies. Elsevier, Amsterdam, the Netherlands.

Buol, S.W., R.J. Southard, R.C. Graham, and P.A. McDaniel. 2003. Soil genesis and classification. 5th edn. Iowa State University Press, Ames, IA.

Buttler, A., H. Dinel, and P.E.M. Levesque. 1994. Effects of physical, chemical and botanical characteristics of peat on carbon gas fluxes. Soil Sci. 158:365–374.

Canada Committee on Ecological (Biophysical) Land Classification, National Wetlands Working Group. 1988. Wetlands of Canada. Polyscience, Montreal, Canada.

Chason, D.B., and D.I. Siegel. 1986. Hydraulic conductivity and related physical properties of peat, Lost River peatland, northern Minnesota. Soil Sci. 142:91–99.

Chimner, R.A., and K.C. Ewel. 2004. Differences in carbon fluxes between forested and cultivated Micronesian tropical peatlands. Wetl. Ecol. Manage. 12:419–427.

Chimner, R.A., and K.C. Ewel. 2005. A tropical freshwater wetland: II. Production, decomposition, and peat formation. Wetl. Ecol. Manage. 13:671–684.

Clymo, R.S. 1983. Peat, p. 159–224. *In* A.J.P. Gore (ed.) Mires—Swamp, bog, fen, and moor. Vol. 4A. General studies. Elsevier, Amsterdam, the Netherlands.

Coles, J.R.P., and J.B. Yavitt. 2004. Linking belowground carbon allocation to anaerobic CH_4 and CO_2 production in a forested peatland, New York state. Geomicrobiol. J. 21:445–455.

Collins, M.E., G.W. Schellentrager, J.A. Doolittle, and S.F. Shih. 1986. Using ground-penetrating radar to study changes in soil map unit composition in selected histosols. Soil Sci. Soc. Am. J. 50:408–412.

Cooper, D.J., and R.E. Andrus. 1994. Patterns of vegetation and water chemistry in peatlands of the west-central Wind River Range, Wyoming, U.S.A. Can. J. Bot. 72:1586–1597.

Crawford, R.M.M. 1983. Root survival in flooded soils, p. 257–283. *In* A.J.P. Gore (ed.) Mires—Swamp, bog fen, and moor. Vol. 4A. General studies. Elsevier, Amsterdam, the Netherlands.

Crill, P.M., K.B. Bartlet, R.C. Harriss, E. Gorham, E.S. Verry, D.I. Sebacher, L. Mazdar, and W. Sanner. 1988. Methane flux from Minnesota peatlands. Global Biogeochem. Cycles 2:317–384.

Darmody, R.G., and J.E. Foss. 1979. Soil-landscape relationships of the tidal marshes of Maryland. Soil. Sci. Soc. Am. J. 43:534–541.

Davis, J.F., and C.A. Engberg. 1955. A preliminary report of investigations of subsidence of organic soil in Michigan. MI Agric. Exp. Stn. Q. Bull. 37:498–505.

Dhowian, A.W., and T.B. Edil. 1980. Consolidation behaviour of peats. Geotech. Test. J. 2:105–114.

Dise, N.B. 1992. Winter fluxes of methane from Minnesota peatlands. Biogeochemistry 17:71–83.

Dise, N.B. 1993. Methane emissions from Minnesota peatlands: Spatial and seasonal variability. Global Biogeochem. Cycles 7:123–142.

Dise, N.B., and E.S. Verry. 2001. Suppression of peatland methane emission by cumulative sulfate deposition in simulated acid rain. Biogeochemistry 53:143–160.

Ellis, N.K., and R. Morris. 1945. Preliminary observations on the relation of yield of crops grown on organic soil with controlled water table and the area of aeration in the soil and subsidence of the soil. Soil Sci. Soc. Am. Proc. 10:282–283.

Everett, K.R. 1971. Composition and genesis of the organic soils of Amchitka Island, Aleutian Islands, Alaska. Arct. Alp. Res. 3:1–6.

Farnham, R.S., and H.R. Finney. 1965. Classification and properties of organic soils. Adv. Agron. 17:115–162.

Finney, H.R., E.R. Gross, and R.S. Farnham. 1974. Limnic materials in peatlands, p. 21–31. *In* A.R. Aandahl (ed.) Histosols: Their characteristic, classification, and use. SSSA Special Publication No. 6. SSSA, Madison, WI.

Foster, D.R., C.A. King, P.H. Glaser, and H.E. Wright, Jr. 1983. Origin of string patterns in boreal mires. Nature 306:256–258.

Fox, C.A. 1985. Micromorphological characterization of histosols, p. 85–104. *In* L.A. Douglas and M.L. Thompson (eds.) Soil micromorphology and soil classification. SSSA Special Publication No. 15. SSSA, Madison, WI.

Freeman, C.M., A. Lock, and B. Reynolds. 1993. Fluxes of CO_2, CH_4 and N_2O from a Welsh peatland following simulation of water table draw-down: Potential feedback to climatic change. Biogeochemistry 19:51–60.

Gafni, A., and K.N. Brooks. 1990. Hydraulic characteristics of four peatlands in Minnesota. Can. J. Soil Sci. 70:239–253.

Gesch, R.W., D.C. Reicosky, R.A. Gilbert, and D.R. Morris. 2007. Influence of tillage and plant residue management on respiration of a Florida Everglades histosol. Soil Tillage Res. 92:156–166.

Gill, C.J. 1970. The flooding tolerance of woody species. For. Abstr. 31:671–678.

Glaser, P.H. 1992. Vegetation and water chemistry, p. 15–26. *In* H.E. Wright, B.A. Coffin, and N.E. Aaseng (eds.) The patterned peatlands of Minnesota. University of Minnesota, Minneapolis, MN.

Glebov, F.Z., L.V. Karpenko, and I.S. Dashkovskaya. 2002. Climate changes, successions of peatlands and zonal vegetation, and peat accumulation dynamics in the Holocene (the West-Siberia peat profile 'Vodorasdel'). Clim. Change 55:175–181.

Godwin, H. 1981. The archives of the peat bogs. Cambridge University Press, New York.

Golovatskaya, E.A., and E.A. Dyukarev. 2009. Carbon budget of ombrotrophic mire sites in the Southern Taiga of Western Siberia. Plant Soil 315:19–34.

Gore, A.J.P. 1983. Mires—Swamp, bog, fen, and moor. Vol. 4A. General studies. Elsevier, Amsterdam, the Netherlands.

Gorham, E. 1987. The ecology and biogeochemistry of *Sphagnum* bogs in central and eastern North America, p. 3–15. *In* A.D. Laderman (ed.) Atlantic white cedar wetlands. Westview Press, Boulder, CO.

Gorham, E. 1991. Northern peatlands: Role in the carbon cycle and probable responses to climatic warming. Ecol. Appl. 1:182–195.

Gorham, E., S.J. Eisenreich, J. Ford, and M.V. Santelmann. 1985. The chemistry of bog waters, p. 339–363. *In* W. Stumm (ed.) Chemical processes in lakes. John Wiley & Sons, New York.

Gorham, E., J.A. Janssens, and P.H. Glaser. 2003. Rates of peat accumulation during the postglacial period in 32 sites from Alaska to Newfoundland, with special emphasis on northern Minnesota. Can. J. Bot. 81:429–438.

Griffin, T.M., and M.C. Rabenhorst. 1989. Processes and rates of pedogenesis in some Maryland tidal marsh soils. Soil Sci. Soc. Am. J. 53:862–870.

Hammond, R.F., and J.F. Collins. 1983. Microfabrics of a sphagno-fibrist and related changes resulting from soil amelioration, p. 689–697. *In* P. Bullock and C.P. Murphy (eds.) Soil micromorphology. Vol. 2. Soil genesis. A B Academic Publishers, Berkhamsted, U.K.

Harriss, R.C., E. Gorham, D.I. Sebacher, K.B. Bartlett, and P.A. Flebbe. 1985. Methane flux from northern peatlands. Nature 315:652–654.

Hatton, R.S., R.D. DeLaune, and W.H. Patrick, Jr. 1983. Sedimentation, accretion, and subsidence in marshes of Barataria Basin, Louisiana. Limnol. Oceanogr. 28:494–502.

Heinselman, M.L. 1963. Forest sites, bog processes, and peatland types in the glacial Lake Agassiz region, Minnesota. Ecol. Monogr. 33:327–374.

Heinselman, M.L. 1970. Landscape evolution, peatland types and the environment in the Lake Agassiz Peatlands Natural Area, Minnesota. Ecol. Monogr. 40:235–261.

Hofstetter, R.H. 1983. Wetlands in the United States, p. 201–244. *In* A.J.P. Gore (ed.) Mires—Swamp, bog, fen, and moor. Vol. 4B. Regional studies. Elsevier, Amsterdam, the Netherlands.

Holden, J., and T.P. Burt. 2003. Hydraulic conductivity in upland blanket peat: Measurement and variability. Hydrol. Processes 17:1227–1237.

Holmgren, G.G.S., M.W. Meyer, R.L. Chaney, and R.B. Daniels. 1993. Cadmium, lead, zinc, copper, and nickel in agricultural soils in the United States of America. J. Environ. Qual. 22:335–348.

Hussein, A.H. 1996. Soil chronofunctions in submerging coastal areas of Chesapeake Bay. Ph.D. Thesis. University of Maryland. College Park, MD.

IPCC. 2007. Climate change 2007: Synthesis report. Contribution of Working Groups I, II, and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [core writing team, R.K. Pachauri and A. Reisinger (eds.)]. IPCC, Geneva, Switzerland.

Ivanov, K.E. 1981. Water movement in mirelands. Academic Press, London, U.K.

Janssens, J.A. 1983. A quantitative method for stratigraphic analysis of bryophytes in Holocene peat. J. Ecol. 71:189–196.

Janssens, J.A. 1992. Bryophytes, p. 43–57. *In* H.E. Wright, B.A. Coffin, and N.E. Aaseng (eds.) The patterned peatlands of Minnesota. University of Minnesota, Minneapolis, MN.

Janssens, J.A., and P.H. Glaser. 1986. The bryophyte flora and major peat-forming mosses at Red Lake peatland, Minnesota. Can. J. Bot. 64:427–442.

Jeglum, J.K. 1971. Plant indicators of pH and water level in peatlands at Candle Lake, Saskatchewan. Can. J. Bot. 49:1661–1672.

Jongedyk, H.A., R.B. Hickok, I.D. Mayer, and N.K. Ellis. 1950. Subsidence of muck soil in the Northern Indiana. Purdue Univ. Agric. Exp. Stn. Spec. Circ. 366:1–10.

Kearney, M.S., A.S. Rogers, J.R.G. Townshend, J.C. Stevenson, J. Stevens, E. Rizzo, and D. Stutzer. 2002. Landsat imagery shows decline of coastal marshes in Chesapeake Bay and Delaware Bays. Eos Trans. Am. Geophys. Union 83:173–178.

Kearney, M.S., and J.C. Stevenson. 1991. Island land loss and marsh vertical accretion rate evidence for historical sea-level changes in Chesapeake Bay. J. Coast. Res. 7:403–415.

Kettridge, N., X. Comas, A. Baird, L. Slater, M. Strack, D. Thompson, H. Jol, and A. Binley. 2008. Ecohydrologically important subsurface structures in peatlands are revealed by ground-penetrating radar and complex conductivity surveys. J. Geophys. Res. 113:G04030.

Kirkinen, J., K. Minkkinen, R. Sievanen, T. Penttila, J. Alm, S. Saarnio, N. Silvan, J. Laine, and I. Savolainen. 2007. Greenhouse impact due to different peat fuel utilisation chains—A lifecycle approach. Boreal Environ. Res. 12:211–223.

Kivinen, E. 1977. Survey, classification, ecology, and conservation of peatlands. Bull. Int. Peat Soc. 8:24–25.

Klinger, L.F. 1996. Coupling of soils and vegetation in peatland succession. Arct. Alp. Res. 28(3):380–387.

Klove, B. 2000. Effect of peat harvesting on peat hydraulic properties and runoff generation. Suo 51:121–129.

Korpijaako, M., and N.W. Radforth. 1972. Studies of the hydraulic conductivity of peat, p. 323–334. *In* Proc. 4th Int. Peat Congr. Vol. 3. Otaniemi, Finland.

Kozlowski, T.T. 1984. Responses of woody plants to flooding, p. 129–163. *In* T.T. Kozlowski (ed.) Flooding and plant growth. Academic Press, London, U.K.

Kuhry, P., and D.H. Vitt. 1996. Fossil carbon/nitrogen ratios as a measure of peat decomposition. Ecology 77:271–275.

Lal, R., J. Kimble, E. Levine, and C. Whitman. 1995. World soils and greenhouse effect: An overview, p. 1–8. *In* R. Lal, J. Kimble, E. Levine, and B.A. Stewart (eds.) Soils and global change. Lewis Publishers, Boca Raton, FL.

Lavoie, M., D. Pare, and Y. Bergeron. 2005. Impact of global change and forest management on carbon sequestration in northern forested peatlands. Environ. Rev. 13:199–240.

Lee, G.B. 1983. The micromorphology of peat, p. 485–501. *In* P. Bullock and C.P. Murphy (eds.) Soil micromorphology. Vol. 2. Soil genesis. A B Academic Publishers, Berkhamsted, U.K.

Lee, G.B., S.W. Bullington, and F.W. Madison. 1988. Characteristics of histic materials in Wisconsin as arrayed in four classes. Soil Sci. Soc. Am. J. 52:1753–1758.

Lee, G.B., and B. Manoch. 1974. Macromorphology and micromorphology of a Wisconsin saprist, p. 47–62. *In* A.R. Aandahl (ed.) Histosols: Their characteristics, classification, and use. SSSA Special Publication No. 6. SSSA, Madison, WI.

Levesque, M.P., and H. Dinel. 1982. Some morphological and chemical aspects of peats applied to the characterization of histosols. Soil Sci. 133:324–333.

Levesque, P.E.M., H. Dinel, and A. Larouche. 1988. Guide to the identification of plant macrofossils in Canada peatlands. Land Resource Centre, Research Branch, Agriculture Canada, Ottawa, Canada.

Lévesque, M., H. Dinel, and R. Marcoux. 1980. Evaluation des critères de différentiation pour la classification de 92 matériaux tourbeaux du Québec et de l'Ontario. Can. J. Soil Sci. 60:479–486.

Lewis, T., and L.M. Lavkulich. 1972. Some folisols in the Vancouver area, British Columbia. Can. J. Soil Sci. 52:91–98.

Lucas, R.E. 1982. Organic soils (histosols). Formation, distribution, physical and chemical properties, and management for crop production. Michigan State University Agricultural Experimental Station Research Report 435. Farm Science. Michigan State University Agricultural Experiment Station, East Lansing, MI.

MacFarlane, I.C. (ed.). 1969. Muskeg engineering handbook. National Research Council of Canada. University of Toronto Press, Toronto, Canada.

MacFarlane, I.C., and G.P. Williams. 1974. Some engineering aspects of peat, p. 79–93. *In* A.R. Aandahl (ed.) Histosols: Their characteristics, classification, and use. SSSA Special Publication No. 6. SSSA, Madison, WI.

Malmer, N., and B. Wallen. 1999. The dynamics of peat accumulation on bogs: Mass balance of hummocks and hollows and its variation through the millennium. Ecography 22:736–750.

Malterer, T., B. McCarthy, and R. Adams. 1996. Use of peat in waste treatment. Min. Eng. 48:53–56.

Mathur, S.P., and M. Levesque. 1985. Negative effect of depth on saturated hydraulic conductivity of histosols. Soil Sci. 140:462–466.

Mausbach, M.J., and J.L. Richardson. 1994. Biogeochemical processes in hydric soil formation. Curr. Top. Wetland Biogeochem. 1:68–127.

Minkkinen, K., and J. Laine. 1998. Long-term effect of forest drainage on the peat carbon stores of pine mires in Finland. Can. J. For. Res. 28:1267–1275.

Mirza, C., and R.W. Irwin. 1964. Determination of subsidence of an organic soil in Southern Ontario. Can. J. Soil Sci. 44:248–253.

Mitsch, W.J., and X. Wu. 1995. Wetlands and global change, p. 205–230. *In* R. Lal, J. Kimble, E. Levine, and B.A. Stewart (eds.) Soil management and greenhouse effect. Lewis Publishers, Boca Raton, FL.

Moore, P.D., and D.J. Bellamy. 1974. Peatlands. Springer Verlag, New York.

Moore, T.R., and R. Knowles. 1989. The influence of water table levels on methane and carbon dioxide emissions from peatland soils. Can. J. Soil Sci. 69:33–38.

Moore, T.R., and N.T. Roulet. 1995. Methane emissions from Canadian peatlands, p.153–164. *In* R. Lal, J. Kimble, E. Levine, and B.A. Stewart (eds.) Soils and global change. CRC Lewis Publication, Boca Raton, FL.

Murayama, S., and Z.A. Bakar. 1996. Decomposition of tropical peat soils. 2. Estimation of in situ decomposition by measurement of CO_2 flux. Jpn. Agric. Res. Q. 30:153–158.

Nickolson, H.H. 1951. Groundwater control in reclaimed marshlands. World Crop. 3:251–254.

Novak, M., and R.K. Wieder. 1992. Inorganic and organic sulfur profiles in nine *Sphagnum* peat bogs in the United States and Czechoslovakia. Water Air Soil Pollut. 65:353–369.

Nykanen, H., J. Alm, K. Lang, J. Silvola, and P.J. Martikainen. 1995. Emissions of CH_4, N_2O and CO_2 from a virgin fen and a fen drained for grassland in Finland. J. Biogeogr. 22:351–357.

Nykanen, H., J. Silvola, J. Alm, and P.J. Martikainen. 1997. The effect of peatland forestry on fluxes of carbon dioxide, methane and nitrous oxide, p. 325–339. *In* C.C. Trettin, M.F. Jurgensen, D.F. Grigal, M.R. Gale, and J.K. Jeglum (eds.) Northern forested wetlands: Ecology and management. CRC Press, Boca Raton, FL.

Nyman, J.A., and R.D. DeLaune. 1991. CO_2 emission and soil Eh responses to different hydrological conditions in fresh, brackish, and saline marsh soils. Limnol. Oceanogr. 36:1406–1414.

Ohlson, M., and R.H. Okland. 1998. Spatial variation in rates of carbon and nitrogen accumulation in a boreal bog. Ecology 79:2745–2758.

Parent, L.E., J.A. Millette, and G.R. Mehuys. 1982. Subsidence and erosion of a histosol. Soil Sci. Am. J. 46:404–408.

Peregon, A., M. Uchida, and Y. Shibata. 2007. *Sphagnum* peatland development at their southern climate range in West Siberia: Trends and peat accumulation patterns. Environ. Res. Lett. 2:045014 (5pp.).

Pons, L.J. 1960. Soil genesis and classification of reclaimed peat soils in connection with initial soil formation. Proc. 7th Int. Congr. Soil Sci. IV:205–211.

Public communication. 2009a. Peatlands: History and uses of peat. Accessed on June 8, 2009. www.peatlandsni.gov.uk/history/fuel.htm

Public communication. 2009b. Peat resources limited. Accessed on June 8, 2009. www.peatresources.com

Public communication. 2009c. Peat: World production, by Country. United States Geological Survey Mineral Resources Program. Accessed on June 8, 2009. www.indexmundi.com/en/commodities/minerals/peat/peat_t9.html

Rabenhorst, M.C. 1995. Carbon storage in tidal marsh soils, p. 93–103. *In* R. Lal, J. Kimble, E. Levine, and B.A. Stewart (eds.) Soils and global change. Lewis Publishers, Boca Raton, FL.

Rabenhorst, M.C. 1997. The chrono-continuum: An approach to modeling pedogenesis in marsh soils along transgressive coastlines. Soil Sci. 167:2–9.

Rabenhorst, M.C., and K.C. Haering. 1989. Soil micromorphology of a Chesapeake Bay tidal marsh: Implications for sulfur accumulation. Soil Sci. 147:339–347.

Rabenhorst, M.C., and B.R. James. 1992. Iron sulfidization in tidal marsh soils. Catena Suppl. 21:203–217.

Redfield, A.C. 1972. Development of a New England salt marsh. Ecol. Monogr. 42:201–237.

Redfield, A.C., and M. Rubin. 1962. The age of salt marsh peat and its relation to recent changes in sea level at Barnstable, Massachusetts. Proc. Natl Acad. Sci. U. S. A. 48:1728–1735.

Reiger, S. 1983. The genesis and classification of cold soils. Academic Press, New York.

Rickard, D.T. 1973. Sedimentary iron sulphide formation, p. 28–65. *In* H. Dost (ed.) Acid sulfate soils, Vol. I. ILRI Publication No. 18, Wageningen, the Netherlands.

Robinson, S.D. 2006. Carbon accumulation in peatlands, southwestern Northwest Territories, Canada. Can. J. Soil Sci. 86:305–319.

Rojstaczer, S., and S.J. Deverel. 1995. Land subsidence in drained histosols and highly organic mineral soils of California. Soil Sci. Am. J. 59:1162–1167.

Row, H.B. 1940. Some soil changes resulting from drainage. Soil Sci. Soc. Am. Proc. 4:402–409.

Rycroft, D.W., D.J. Williams, and H.A. Ingram. 1975. The transmission of water through peat: 1. Review. J. Ecol. 63:535–556.

Schipper, L.A., and M. McLeod. 2002. Subsidence rates and carbon loss in peat soils following conversion of pasture in the Waikato Region, New Zealand. Soil Use Manage. 18:91–93.

Sebacher, D.I., R.C. Harriss, K.B. Bartlet, S.M. Sebacher, and S.S. Grice. 1986. Atmospheric methane sources: Alaskan tundra bogs, alpine fens, and a subarctic boreal marsh. Tellus 38B:1–10.

Seppälä, M., and L. Koutaniemi. 1985. Formation of a string and pool topography as expressed by morphology, stratigraphy, and current processes on a mire in Kuusamo, Finland. Boreas 14:287–309.

Sheng, Y., L.C. Smith, G.M. MacDonald, K.V. Kremenetski, K.E. Frey, A.A. Velichko, M. Lee, D.W. Beilman, and P. Dubinin. 2004. A high-resolution GIS-based inventory of the west Siberian peat carbon pool. Global Biogeochem. Cycles 18:GB3004.

Shih, S.F., and J.A. Doolittle. 1984. Using radar to investigate organic soil thickness in the Florida Everglades. Soil Sci. Soc. Am. J. 48:651–656.

Shih, S.F., D.E. Vandergrift, D.L. Myhre, G.S. Rahi, and D.S. Harrison. 1981. The effect of land forming on subsidence in the Florida Everglades' organic soil. Soil Sci. Soc. Am. J. 45:1206–1209.

Silvola, J., J. Alm, U. Ahlholm, H. Nykanen, and P.J. Martikainen. 1996. CO_2 fluxes from peat in boreal mires under varying temperature and moisture conditions. J. Ecol. 84:219–228.

Sjors, H. 1963. Bogs and fens on the Attawapiskat River, northern Ontario. Mus. Can. Bull. Contrib. Bot. 186:43–133.

Sjors, H. 1985. A comparison between mires of southern Alaska and Fennoscandia. Aquilo Ser. Bot. 21:89–94.

Smith, C.J., R.D. DeLaune, and W.H. Patrick, Jr. 1983. Carbon dioxide emission and carbon accumulation in coastal wetlands. Estuar. Coast. Shelf Sci. 17:21–29.

Soil Survey Staff. 1996. Keys to soil taxonomy. 7th edn. USDA-NRCS, U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1998. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. USDA agricultural handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2010 Keys to soil taxonomy. 11th edn. USDA-NRCS, U.S. Government Printing Office, Washington, DC.

Stephens, J.C., and W.H. Speir. 1969. Subsidence of organic soils in the USA, p. 523–534. International Association of Scientific Hydrology. Special Publication No. 89.

Sturges, D.L. 1968. Hydrologic properties of peat from a Wyoming mountain bog. Soil Sci. 106:262–264.

Strack, M., and J.M. Waddington. 2007. Response of peatland carbon dioxide and methane fluxes to a water table drawdown experiment. Global Biogeochem. Cycles 21:GB1007.

Strack, M, J.M. Waddington, and E.S. Tuittila. 2004. Effects of water table drawdown on northern peatland methane dynamics: Implications for climate change. Global Biogeochem. Cycles 18:GB4003.

Svensson, B.H., and T. Rosswall. 1984. In situ methane production from acid peat in plant communities with different moisture regimes in a subarctic mire. Oikos 43:341–350.

Swanson, D.K., and D.F. Grigal. 1988. A simulation model of mire patterning. Oikos 53:309–314.

Swanson, D.K., and D.F. Grigal. 1989. Vegetation indicators of organic soil properties in Minnesota. Soil Sci. Soc. Am. J. 53:491–495.

Swanson, D.K., and D.F. Grigal. 1991. Biomass, structure, and trophic environment of peatland vegetation in Minnesota. Wetlands 11:279–302.

Thomas, F.H. 1965. Subsidence of peat and muck soils in Florida and other parts of the United States—A review. Soil Crop Sci. Soc. Fla. Proc. 25:153–160.

Tolonen, K. 1979. Peat as a renewable resource: Long-term accumulation rates in North European mires, p. 282–296. *In* Proceedings of the international symposium on classification of peat and peatlands, Hyytiala, Finland. International Peat Society, Helsinki, Finland.

Tolonen, K., R.B. Davis, and L.S. Widoff. 1988. Peat accumulation rates in selected Maine peat deposits. Maine Geological Survey Bulletin No. 33, Marine department of conservation.

Trewartha, G.T. 1968. An introduction to climate. McGraw-Hill, New York.

Turetsky, M., K. Wieder, L. Halsey, and D. Vitt. 2002. Current disturbance and the diminishing peatland carbon sink. Geophys. Res. Lett. 29:1526.

Turetsky, M.R., R.K. Wieder, D.H. Vitt, R.J. Evans, and K.D. Scott. 2007. The disappearance of relict permafrost in boreal North America: Effects on peatland carbon storage and fluxes. Global Change Biol. 13:1922–1934.

Ugolini, F.C., and D.H. Mann. 1979. Biopedological origin of peatlands in south east Alaska. Nature 281:366–368.

Updegraff, K., S.D. Bridgham, J. Pastor, P. Weishampel, and C. Harth. 2001. Response of CO_2 and CH_4 emissions from peatlands to warming and water table manipulation. Ecol. Appl. 11:311–326.

Urban, N.R. 1987. The nature and origins of acidity in bogs. Ph.D. Thesis. University of Minnesota. Minneapolis, MN.

USDA-NRCS. U.S. Department of Agriculture—Natural Resources Reservation Service. 1996. Field indicators of hydric soils in the United States. USDA-NRCS, Fort Worth, TX.

Van Breemen, N. 1982. Genesis, morphology and classification of acid sulfate soils in coastal plains, p. 95–108. *In* J.A. Kittrick, D.S. Fanning, and L.R. Hossner (eds.) Acid sulfate weathering. SSSA Special Publication No. 10. SSSA, Madison, WI.

Van Heuveln, B., A. Jongerius, and L.J. Pons. 1960. Soil formation in organic soils. Proc. 7th Int. Congr. Soil Sci. IV:195–204.

Vitt, D.H., and S. Bayley. 1984. The vegetation and water chemistry of four oligotrophic basin mires in northwestern Canada. Can. J. Bot. 62:1485–1500.

Vitt, D.H., and N.G. Slack. 1975. An analysis of the vegetation of *Sphagnum*-dominated kettle-hole bogs in relation to environmental gradients. Can. J. Bot. 53:332–359.

Vitt, D.H., R.K. Wieder, K.D. Scott, and S. Faller. 2009. Decomposition and peat accumulation in rich fens of boreal Alberta, Canada. Ecosystems 12:360–373.

Von Post, L., and E. Granlund. 1926. Sodra Sveriges torvtillgångar I. Sven. Geol. Unders., C. 335:307–411.

Wakeley, J.S., S.W. Sprecher, and R.W. Lichvar. 1996. Relationships among wetland indicators in Hawaiian rain forest. Wetlands 16:173–184.

Walter, H. 1977. The oligotrophic peatlands of Western Siberia—The largest peinohelobiome in the world. Vegetation 34:167–178.

Ward, S.E., R.D. Bardgett, N.P. McNamara, J.K. Adamson, and N.J. Ostle. 2007. Long-term consequences of grazing and burning on northern peatland carbon dynamics. Ecosystems 10:1069–1083.

Washburn, A.L. 1980. Geocryology. John Wiley, New York.

Weir, W.W. 1950. Subsidence of peat lands in the Sacramento–San Joaquin Delta of California. Hilgardia 20:37–55.

Widdell, F. 1988. Microbiology and ecology of sulfate- and sulfur-reducing bacteria, p. 469–585. *In* A.J.B. Zehnder (ed.) Biology of anaerobic microorganisms. John Wiley & Sons, New York.

Wieder, R.K., M. Novák, W.R. Schell, and T. Rhodes. 1994. Rates of peat accumulation over the past 200 years in five *Sphagnum*-dominated peatlands in the United States. J. Paleolimnol. 12:35–47.

Wieder, R.K., J.B. Yavitt, and G.E. Lang. 1990. Methane production and sulfate reduction in two Appalachian peatlands. Biogeochemistry 10:81–104.

Witty, J.E., and R.W. Arnold. 1970. Some folists on Whiteface mountain, New York. Soil Sci. Soc. Am. Proc. 34:653–657.

Wosten, J.H.M, A.B. Ismail, and A.L.M. van Wijk. 1997. Peat subsidence and its practical implications: A case study in Malaysia. Geoderma 78:25–36.

Wright, H.E., B.A. Coffin, and N.E. Aaseng (eds.) 1992. The patterned peatlands of Minnesota. University of Minnesota, Minneapolis, MN.

Yefimov, V.N. 1986. Peat soils and their fertility. Agropromizdat, Leningrad, Russia.

Yu, Z. 2006. Holocene carbon accumulation of fen peatlands in boreal western Canada: A complex ecosystem response to climate variation and disturbance. Ecosystems 9:1278–1288.

Zurek, S. 1976. The problem of growth of the Eurasia peatland in the Holocene, p. 99–122. Proc. 5th Int. Peat Congr., Poznan, Poland, Vol. 2.

33.3 Andisols

Paul A. McDaniel

David J. Lowe

Olafur Arnalds

Chien-Lu Ping

33.3.1 Introduction

Andisols are soils that typically form in loose volcanic ejecta (tephra) such as volcanic ash, cinders, or pumice. They are characterized by andic properties that include physical, chemical, and mineralogical properties that are fundamentally different from those of soils of other orders. These differences resulted in a proposal to recognize these soils at the highest level in the USDA soil classification system (Smith, 1978). In 1990, Andisols were added to *Soil Taxonomy* as the 11th soil order (Soil Survey Staff, 1990; Parfitt and Clayden, 1991). A very similar taxonomic grouping, Andosols, is 1 of the 32 soil reference groups recognized in the World Reference Base for Soil Resources (IUSS Working Group WRB, 2006). Andisols (and Andosols) are classified on the basis of selected chemical, physical, and mineralogical properties acquired through weathering and not on parent material alone. Both soil names relate to two Japanese words, *anshokudo* meaning "dark colored soil" (*an*, dark; *shoku*, color or tint; *do*, soil) and *ando* meaning "dark soil." *Ando* was adopted into western soil science literature in 1947 (Simonson, 1979).

The central concept of Andisols is one of deep soils commonly with depositional stratification developing mainly from ash, pumice, cinders (scoria), or other explosively erupted, fragmental volcanic material (referred to collectively as tephra) and volcaniclastic or reworked materials. Andisols occur much less commonly on lavas. Unlike many other soils, Andisol profiles commonly undergo "upbuilding pedogenesis" as younger tephra materials are deposited on top of older ones. The resulting profile character is determined by the interplay between the rate at which tephras are added to the land surface and classical "top-down" processes that form soil horizons. Understanding Andisol

genesis in many instances thus requires a stratigraphic approach combined with an appreciation of buried soil horizons and polygenesis.

The coarser fractions of Andisols are often dominated by volcanic glass. This glass weathers relatively quickly to yield a fine colloidal or nanoscale fraction (1–100 nm) dominated by short-range order materials composed of "active" Al, Si, Fe, and organic matter, especially humus. Previously described erroneously as "amorphous," short-range order materials comprise extremely tiny but structured nanominerals, the main ones being allophane and ferrihydrite (Hochella, 2008). A useful collective descriptor for them is "nanocrystalline" (Michel et al., 2007). Another colloidal constituent, imogolite, comprises long filamental tubes and therefore has both short- and long-range order (Churchman, 2000). The nanominerals, chiefly allophane, ferrihydrite, and metal–humus complexes, are responsible for many of the unique properties exhibited by Andisols.

Despite covering less of the global ice-free land area than any other soil order (~1%), Andisols generally support high population densities, about 10% of the world's population (Ping, 2000). This is because they typically have exceptional physical properties for plant growth and, in many localities, high native fertility because relatively frequent additions of tephra can renew potential nutrient sources (Ugolini and Dahlgren, 2002; Dahlgren et al., 2004). The majority of Andisols occur in humid regions where there is adequate rainfall. Andisols often have high organic carbon contents. These and other factors make Andisols generally well suited for agriculture production and historically allowed establishment of nonshifting agricultural practices. Despite their generally favorable properties for plant growth, Andisols do pose some engineering and fertility challenges. These soils have low bulk densities, resulting in low weight-bearing capacity. Andisols also exhibit thixotropy and sensitivity, properties that cause them to behave in a fluid-like manner when loading pressures are applied (Neall, 2006; Arnalds, 2008). Andisols may exhibit substantial fertility limitations, including

P fixation, low contents of exchangeable bases (especially K) and other nutrients, and strong acidity and Al toxicity (Shoji and Takahashi, 2002; Dahlgren et al., 2004; Lowe and Palmer, 2005).

33.3.2 Geographic Distribution

Andisols cover approximately 124 million ha or about 0.84% of the Earth's ice-free surface (Soil Survey Staff, 1999). They are closely associated with areas of active and recently active volcanism, and their global distribution is depicted in Figure 33.13. The greatest concentration of Andisols is found along the Pacific Ring of Fire, a zone of tectonic activity and volcanoes stretching from South through Central and North America via the Aleutian Islands to the Kamchatka Peninsula of Russia through Japan, Taiwan, the Philippines, and Indonesia to Papua New Guinea and New Zealand. Other areas include the Caribbean, central Atlantic ridge, northern Atlantic rift, the Mediterranean, parts of China, Cameroon, the Rift Valley of east Africa, and southern Australia (Soil Survey Staff, 1999). There are numerous volcanic islands where Andisols are common, including Iceland, the Canary Islands, Azores, Galapagos Islands, Hawaiian Islands, the West Indies, and various small islands in the Pacific.

The global distribution of Andisols encompasses a wide variety of climatic conditions—cold-to-hot and wet-to-dry. This suggests that climate is less important to the formation of Andisols than is proximity to volcanic or pyroclastic parent materials. Nevertheless, the majority of Andisols are found in higher-rainfall regions of the world. Almost two-thirds of Andisols occur in humid regions (udic soil moisture regimes) while fewer than 5% occur in aridic moisture regimes (Mizota and van Reeuwijk, 1989; Wilding, 2000). Approximately half of the world's Andisols occur in the tropics, with the remaining half being split between boreal and temperate regions (Wilding, 2000; IUSS Working Group WRB, 2006).

There are almost 15.6 million ha of Andisols in the United States (Soil Survey Staff, 1999). The largest areas occur in Alaska
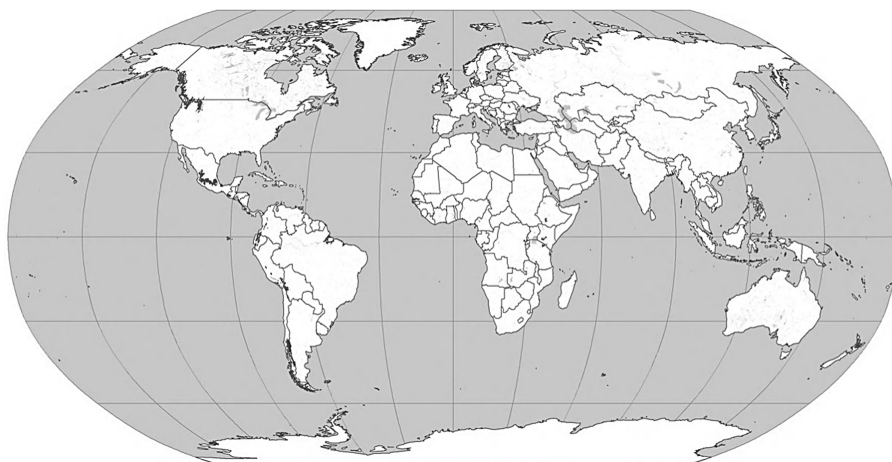
FIGURE 33.13 Global distribution of Andisols. (Courtesy of USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC, 2010.)

(~10 million ha) and in Washington, Oregon, Idaho, northern California, and western Montana (Pewe, 1975; Rieger et al., 1979; Ping et al., 1989; Southard and Southard, 1991; Ugolini and Dahlgren, 1991; Goldin et al., 1992; Takahashi et al., 1993; McDaniel and Hipple, 2010). In the Pacific Northwest region of Washington, Idaho, and Oregon, most Andisols are forested and occur at mid-to-high elevations in cooler temperature regimes (McDaniel et al., 2005). Few Andisols are found in warmer temperature regimes because the summers are normally too hot and dry to allow sufficient weathering or leaching to produce the required andic properties.

Iceland contains ~7 million ha of Andisols. These represent the largest area of Andisols in Europe (Arnalds, 2004). Andisols also occur in France, Germany, Spain, Italy, and Romania (Buol et al., 2003; Kleber et al., 2004; Quantin, 2004; IUSS Working Group WRB, 2006; Arnalds et al., 2007). In New Zealand, ~3.2 million ha of Andisols occur on the North Island, the majority now supporting agriculture or forestry (Parfitt, 1990; Lowe and Palmer, 2005). Japan has ~6.9 million ha of Andisols (Wada, 1986; Takahashi and Shoji, 2002).

Some soils classified as Andisols are also found in humid areas not associated with volcanic activity such as in the southern Appalachian Mountains, parts of Kyushu (Japan), Scotland, Spain, and the Alps. These soils have large quantities of Al or Fe associated with humus (see Section 33.3.3.2) and similar management constraints as those of soils formed from volcanic ejecta and also key out as Andisols. These attributes further highlight the importance of realizing that Andisols are not classified on parent material but on the properties acquired during weathering and leaching. By the same token, soils other than Andisols, such as Entisols, Inceptisols, Spodosols, Mollisols, Oxisols, Vertisols, Alfisols, or Ultisols, may form in association with volcanic or pyroclastic materials (e.g., Shoji et al., 2006).

33.3.3 Andisol Properties

33.3.3.1 Morphological Features

Most Andisols have distinct morphological features. They usually have multiple sequences of horizons (Figure 33.14) resulting from the intermittent deposition of tephras and ongoing top-down soil formation referred to as upbuilding pedogenesis (see Section 33.3.5.1). A horizons are typically dark, often overlying reddish brown or dark yellowish brown Bw cambic horizons. Buried A–Bw sequences are common (Figure 33.14). Layers representing distinct tephra-fall events are common, often manifested as separate Bw horizons or as BC or C horizons if the tephra shows limited weathering or is relatively thick. Horizon boundaries are typically distinct or abrupt where these thicker layers occur.

Andisols are usually light and easily excavated because of their low bulk density and weakly cohesive clay minerals. The high porosity allows roots to penetrate to great depths. Andisols
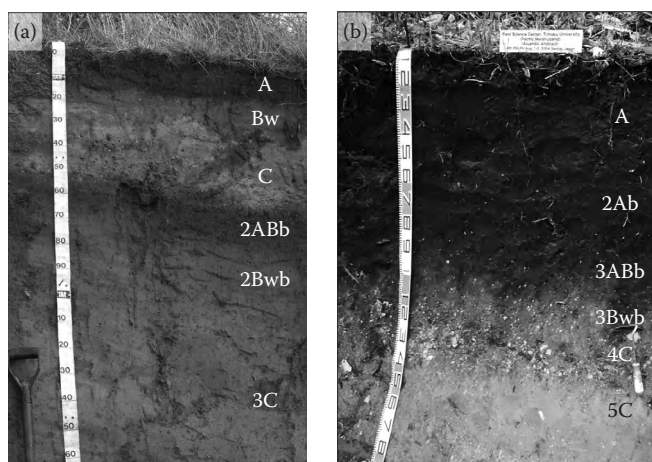
FIGURE 33.14 Andisol profiles (a) allophanic Taupo soil (Udivitrand) from New Zealand (scale divisions = 10 cm) and (b) nonallophanic Tohoku Farm soil (Melanudand) from Japan with thick (~1 m), dark, strongly humified horizons (melanic epipedon).

generally have granular structures in A horizons, but the structure in Bw horizons is generally weak subangular blocky, often crushing readily to crumb structure. Some Andisols (Udands) formed in areas of high rainfall have higher clay contents while soils that are subjected to wet and dry cycles form prismatic structures. At higher water contents, soils containing as little as 2% allophane have a characteristic greasy feel (Parfitt, 2009), an indication of sensitivity.

33.3.3.2 Mineralogical Properties

Tephra parent materials weather rapidly to form nanominerals that are responsible for many of the unique physical and chemical properties associated with Andisols. Although a wide range of clay minerals can be found in Andisols (such as gibbsite, kaolinite, vermiculite, smectite, crystalline Fe oxides such as hematite and goethite, and cristobalite), those of greatest interest are allophane, imogolite, ferrihydrite, and the Al– and Fe–humus complexes because they confer the characteristic andic properties (Dahlgren et al., 2004; Parfitt, 2009).

Allophane is nearly X-ray amorphous, but under an electron microscope it is structured over short distances, appearing as nanoparticles of hollow spheres 3.5–5 nm in diameter that have the chemical composition $(1–2)SiO_2 \cdot Al_2O_3 \cdot (2–3)H_2O$ (Figure 33.15a) (Wada, 1989; Churchman, 2000; Brigatti et al., 2006; Theng and Yuan, 2008). The most common type of allophane is the so-called Al-rich allophane with an Al:Si molar ratio of ~2 (it is sometimes called proto-imogolite allophane). There is also Si-rich allophane with an Al:Si ratio ~1 (also referred to as halloysite-like allophane).

Imogolite has the composition $(OH)SiO_3 \cdot Al_2(OH)_3$ and has both long- and short-range order. Under an electron microscope, it appears as long smooth and curved hollow threads or tubules with inner and outer diameters of ~0.7 and 2 nm, respectively
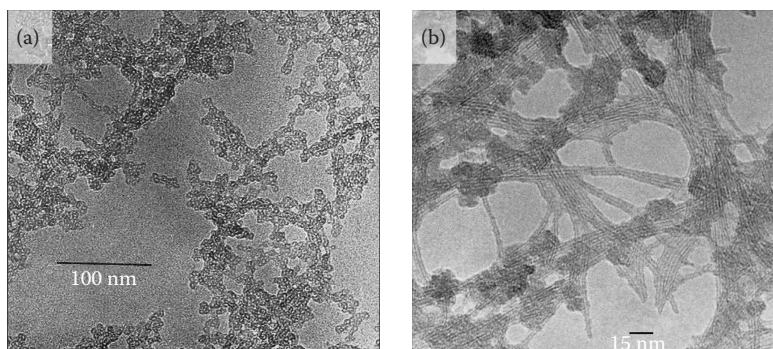
FIGURE 33.15 Micrographs of (a) allophane and (b) imogolite (external diameter of nanotubes is ~2 nm). (Reproduced with the kind permission of the Mineralogical Society of Great Britain and Ireland from a paper by Parfitt, R.L. 2009. Allophane and imogolite: Role in soil biogeochemical processes. Clay Miner. 44:135–155. With permission.)

(Figure 33.15b). These nanotubes typically appear as bundles of two or more threads 10–30 nm thick and several micrometers long (Theng and Yuan, 2008). Imogolite in Japan can be seen with the naked eye as a whitish gel film infilling pores in coarse pumice particles (Wada, 1989).

Allophane and imogolite both have high surface areas, ranging from 700 to 1500 m^2 g^{-1} (Parfitt, 2009), and this feature, coupled with their variable surface charge characteristics and exposure of $(OH)Al(OH_2)$ groups at wall perforations (defects), explains their strong affinity for water, metal cations, organic molecules, and other soil minerals (Harsh et al., 2002; Theng and Yuan, 2008). Even small amounts contribute huge reactive surface areas in soils (Lowe, 1995). Allophane and imogolite are soluble in ammonium (acid) oxalate solution, and the Si dissolved is used to estimate their contents in soils (Parfitt and Henmi, 1982; Parfitt, 2009). *Soil Taxonomy* uses oxalate-extractable Al (and Fe) to help define andic soil properties (see Figure 33.16 and Section 33.3.4). Allophane content of B horizons is quite

variable, ranging from about 2% in slightly weathered or metal–humus-dominated systems to >40% in well-developed Andisols. It typically increases with depth in upper subsoils, usually being highest in the Bw and buried horizons. But in many Andisol profiles in New Zealand, allophane decreases and halloysite concomitantly increases with depth in lower subsoils either because of the downward migration of Si into lower profiles or because of changes in climate during upbuilding, or both. Imogolite is more commonly found in B horizons under carbonic acid weathering regimes than in A horizons where organic acid weathering dominates (Dahlgren et al., 2004). Allophane may occur dispersed as groundmass, as coatings, bridges, or infillings (in vesicles or in root channels), or it may be disseminated through pseudomorphs of glass or feldspar grains (Jongmans et al., 1994, 1995; Bakker et al., 1996; Gérard et al., 2007).

Ferrihydrite is common in many Andisols, especially those associated with more basic parent materials, has a composition of $Fe_5HO_8 \cdot 4H_2O$ and imparts a reddish brown color (hues of 5YR–7.5YR; Bigham et al., 2002). Made up of spherical nanoparticles 2–5 nm in diameter (Schwertmann, 2008), ferrihydrite has large, reactive surface areas ranging from ~200 to 500 m^2 g^{-1} (Childs, 1992; Jambor and Dutrizac, 1998). Its abundance is commonly estimated from the amount of Fe extracted by ammonium oxalate solution multiplied by 1.7 (Parfitt and Childs, 1988). It is a widespread and characteristic component of young Fe-oxide accumulations precipitated from Fe-rich solutions in the presence of organic matter, such as in Iceland (Arnalds, 2004), and elsewhere, including New Zealand, Japan, and Australia where its precipitation may be inorganic or bacteria-driven (Childs et al., 1991; Lowe and Palmer, 2005). Ferrihydrite can transform to hematite via solid-state transformation or goethite through dissolution and reprecipitation (Schwertmann, 2008).

Metal–humus complexes are significant components of some Andisol colloidal fractions. These Al- and Fe-organic complexes are immobile and accumulate in dark or black surface horizons where organic materials are abundant, and dark (melanic) horizons may extend to depths as much as 2 m (see Figure 33.14b). Metal–humus complexes represent the active forms of Al and Fe in nonallophanic Andisols as described below (Dahlgren et al., 2004).
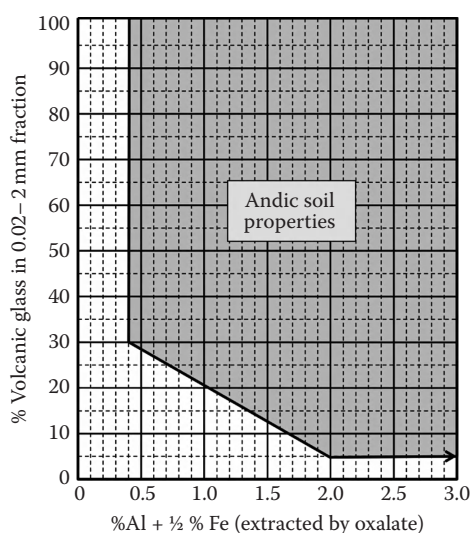
FIGURE 33.16 Andic soil properties as defined by *Soil Taxonomy*. (From Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS, Washington, DC.)

Halloysite is a relatively fast-forming 1:1 layer silicate that often exhibits tubular or spheroidal morphology (White and Dixon, 2002; Joussein et al., 2005). Its formation is favored in seasonally dry environments where higher Si concentrations are maintained (Shoji et al., 1993). These include areas of lower rainfall, restricted drainage, and Si-rich parent materials (Lowe, 1986; Churchman, 2000; Churchman and Lowe, 2011, Section 20.1). Halloysite surfaces are characterized by some permanent negative charge, allowing retention of cations across a wide range of pH values.

The soil solution in Andisols in a range of locations may contain large amounts of dissolved Si, which leads to the formation by nucleation of secondary silica minerals from the saturated solution (Ping et al., 1988; Shoji et al., 1993; Ping, 2000; Nanzyo, 2002, 2007; Waychunas and Zhang, 2008). Termed laminar opaline silica, this material is circular or elliptical in shape (0.2–0.5 μm diameter) and extremely thin. Precipitation of the silica may be aided by evaporation or freezing of soil water, or via plant-related processes related to Si uptake and recycling (Lowe, 1986; Drees et al., 1989; Churchman, 2000; Henriet et al., 2008). Such silica polymorphs can be distinguished from biogenic forms of silica (phytoliths) because the latter have more complex shapes inherited from biological cells (Kondo et al., 1994; Nanzyo, 2007).

Andisols dominated by allophane with subordinate imogolite and ferrihydrite in upper horizons are referred to as *allophanic Andisols*. These contrast with a second, strongly acid group known as *nonallophanic Andisols* in which metal–humus complexes dominate the colloidal mineralogy. Nonallophanic Andisols are common in Japan especially where they account for about 30% of soils formed on tephras (Takahashi and Shoji, 2002) and are known in around 20 other countries (Saigusa and Matsuyama, 2004). Examples of soils from each group are shown in Figure 33.14. In Table 33.15, the Thingvallasveit and Tirau soils are examples of allophanic Andisols; the Tohoku Farm soil is an example of a nonallophanic Andisol. The mineralogical differences between these two groups of Andisols lead to several important different physical and chemical properties (especially the strong acidity of nonallophanic Andisols) and significant management implications (Dahlgren et al., 2004).

In the silt and sand fractions of Andisols, the dominant components are volcanic glass (a mineraloid) and various primary minerals. The glass particles (shards) which, like shattered glass, have sharp angles and edges, are very abrasive. However, these glass particles are usually coated with colloidal materials including allophane, ferrihydrite, and other Fe oxides and their humus complexes, which all contribute to aggregate formation. It is noteworthy that volcanic glass is often quite vesicular and porous in nature (as is pumice), and thus can retain water and has more chemical activity than other common sandy materials (Ping, 2000; Lowe and Palmer, 2005).

33.3.3.3 Chemical Properties

One of the common characteristics of Andisols is accumulation of relatively large quantities of organic matter, both in the allophanic (moderate pH) and nonallophanic Andisols (low pH; Table 33.15). Allophanic Andisols typically contain up to ~8%–12% C, whereas nonallophanic soils may contain up to ~25%–30% C (Mizota and van Reeuwijk, 1989). The residence time of C in Andisols, as measured by ^{14}C, is much greater than that of other soil orders (Parfitt, 2009). In addition, upbuilding pedogenesis leads to the storage of C in lower parts of profiles, and especially in buried A horizons that are sealed off and isolated from most surface processes.

Andisols are almost always acid, with most pH (H_2O) values ranging from 4.8 to 6.0 (Shoji et al., 1993; Dahlgren et al., 2004). Uncultivated, nonallophanic Andisols with high organic matter contents typically have a pH (H_2O) < 4.5 (IUSS Working Group WRB, 2006).

One of the key factors affecting the chemistry of Andisols is the variable surface charge associated with the colloidal fraction (Qafoku et al., 2004). Electrical charges on colloid or nanoparticle surfaces can be either positive or negative, and change as a function of pH. Surfaces have positive charge at lower pH and retain anions, while cations are retained by the negative charge at higher pH. This variable charge greatly affects the behavior of ions that are retained in soils. Cation retention capacity of Andisols makes the soils susceptible to metal and radioactive fallout ^{137}Cs pollution (Adamo et al., 2003; Sigurgeirsson et al., 2005). Andisols also exhibit anion exchange properties, which can be important for nutrient retention (e.g., NO_3^- and SO_4^{2-}) (Shoji et al., 1993).

Measured cation exchange capacity of Andisols is relatively high, often 20 to >50 cmol$_c$ kg^{-1}. However, because the dominant colloids have variable charge, much of this CEC is pH dependent. This is especially true in allophanic Andisols (Dahlgren et al., 2004) and means that CEC decreases with decreasing pH. And because most Andisols are acid, CEC measurements made at pH 7 or 8.2 will be artificially high. In andic soils of the Pacific Northwest of United States, average CEC values determined using unbuffered extractants (effective CEC or ECEC) are approximately one-fourth of those determined at pH 8.2, 6.5 cmol$_c$ kg^{-1} vs. 26.4 cmol$_c$ kg^{-1} (McDaniel et al., 2005). This phenomenon needs to be considered when measuring CEC and base saturation or interpreting these data. The relatively low ECEC of Andisols can limit their ability to retain and exchange Ca, Mg, and K. Some representative cation exchange characteristics of Andisols are presented in Table 33.15.

The active Al and Fe compounds in Andisols (allophane/imogolite, metal–humus complexes, and ferrihydrite) also have the ability to sorb and strongly bind anions such as phosphate and fluoride (Shoji et al., 1993; Dahlgren et al., 2004; Parfitt, 2009). Much of this sorption is not reversible, leading to large quantities of phosphate being rendered unavailable for plant uptake. As described in Section 33.3.4, the amount of P retention in soils is used to define andic soil properties (Soil Survey Staff, 2010). Similarly, quantities of active Al and Fe compounds can also be estimated by reacting soil with NaF solution. Sorption of F$^-$ releases OH$^-$ into solution, thereby raising the pH. A resultant pH greater than ~9.5 indicates the presence of allophane/imogolite

TABLE 33.15 Selected Chemical and Mineralogical Properties of Representative Andisols

Horizon	Depth (cm)	pH H$_2$O	pH NaF	Clay (%)	Organic C %	ECEC[a]	CEC pH 7 (cmol$_c$ kg^{-1})	Sum of Exch. Bases	[b]Exch. Al^{3+}	P Retention	Al in Al–Humus Complexes[c] (%)	Allophane[d]	Ferrihydrite[e]
Thingvallasveit Series[f] (Haplocryand)—Iceland													
A1	0–12	5.4	11.0	—	7.9	—	31.9	9.9	—	99	—	13	8.5
A2	12–28	5.7	10.7	1	7.6	—	44.6	13.1	—	99	0.1	16	11.1
Bw1	28–61	5.6	10.8	—	7.4	—	43.8	10.1	—	99	—	15	12.7
Bw2	61–68	5.7	11.1	3	4.2	—	32.3	6.5	—	99	—	16	6.0
2Bw3	68–87	5.6	10.5	—	2.1	—	24.8	4.4	—	98	—	17	4.4
2C	87–142	5.5	10.3	1	0.8	—	11.8	2.8	—	99	—	13	3.6
Tirau Series[g] (Hapludand)—New Zealand													
Ap	0–18	5.6	9.6	19	7.9	14.6	29.7	14.4	0.2	88	0.7	9	1.0
Bw1	18–32	6.2	9.9	11	2.0	5.6	11.3	5.6	0	98	0.2	14	1.2
Bw2	32–48	6.2	9.8	13	1.0	5.2	10.5	5.2	0	98	0.2	16	1.2
Bw3	48–65	6.4	9.5	20	0.5	6.1	12.5	6.1	0	91	0.1	10	0.7
2BCb	65–90	6.4	9.4	18	0.5	7.5	13.2	7.5	0	86	0.1	8	0.8
Tohoku Farm[h] (Melanudand)—Japan													
A1	0–14	4.2	10.5	18	11.5	13.3	52.2	2.3	11.0	83	1.5	0.3	—
2A2	14–30	4.6	11.6	5	8.0	8.4	49.1	0.4	8.0	93	2.5	1.1	—
3A3	30–57	4.8	11.6	12	13.2	8.4	56.1	0.4	8.0	93	2.7	1.4	—
4Bw1	57–80	5.2	11.3	8	2.8	2.0	18.4	0.1	1.9	93	0.9	6.7	—
4Bw2	80–126	5.2	11.0	16	0.9	0.9	13.7	0.2	0.7	94	0.6	7.4	—
4Bw3	126–160	5.3	10.9	20	0.6	0.8	10.9	0.2	0.6	89	0.5	7.0	—
4C	160–200	5.3	10.9	22	0.5	1.8	11.5	0.4	1.4	84	0.5	4.9	—

[a] Effective cation exchange capacity.

[b] Exchangeable Al^{3+} extracted with KCl.

[c] Estimated using pyrophosphate-extractable Al.

[d] Estimated using oxalate-extractable Si (Parfitt, 1990).

[e] Estimated using oxalate-extractable Fe (Parfitt and Childs, 1988).

[f] Data are from Arnalds et al. (1995).

[g] Data are from Bakker et al. (1996).

[h] Data (Pedon #86P0091) are from Soil Survey Staff 2011.

and/or Al–humus complexes, and because of this, NaF field test kits can be used for field identification of Andisols (Fieldes and Perrott, 1966; IUSS Working Group WRB, 2006).

33.3.3.4 Physical Properties

Unique physical attributes of Andisols are related to structural assemblages of hollow spheres and tubular threads as mineral entities into resilient, progressively larger (silt-sized) aggregated domains. This aggregation results in low density, high porosity, high surface area, and high soil water retention even at low water potentials. The structural arrangement accounts for the low thermal conductivity of andic materials, which is three to four times less than that of the phyllosilicates in other mineral soils. It also accounts for the thixotropic and sensitivity character of these soils and several irreversible changes in physical properties that occur upon drying (Ping, 2000; Neall, 2006).

Andisols have low bulk density, usually <0.9 g cm^{-3}, because of the high organic matter and nanomineral contents (Table 33.16) and well-developed aggregation. This results in good tilth and makes them excellent rooting media. On the other hand, low bulk densities result in low weight-bearing capacity and make Andisols highly susceptible to wind and water erosion when surface cover is removed or degraded (Ping, 2000; Nanzyo, 2002; Dahlgren et al., 2004; Neall, 2006). Because of the nature of volcanic ejecta and its distribution, many Andisols in proximal locations contain appreciable amounts of gravel and stones, and coarse-grained tephra layers may have adverse effects on hydrological properties by interrupting capillary movement of water. Some additional adverse physical properties include a high glass content that can reduce the quantity and biodiversity of soil organisms such as earthworms, and the presence occasionally of impenetrable horizons such as thin Fe pans (placic horizons) in higher rainfall areas (Neall, 2006).

Andisols typically exhibit high water retention because of the presence of allophane, ferrihydrite, and metal–humus complexes, which have high surface areas as noted previously. As a result, moisture contents of many Andisols can exceed 100% on a weight basis, even at soil moisture tensions of 1500 kPa—this feature is illustrated by the data for the Hilo soil (Table 33.16). Water retention is greatest in Andisols that have undergone significant weathering and hence have high clay contents.

Most field-moist Andisols have a greasy feel when rubbed between the fingers and exhibit smeariness—these can be both indicators of the properties known generally as sensitivity and thixotropy (Soil Survey Staff, 1975; Torrance, 1992). Thixotropy is a reversible gel–sol transformation that occurs when shear forces are applied to a moist soil. The applied shear force causes the soil to abruptly lose strength, sometimes to the point of behaving as a fluid. When the shear force is removed, the soil will recover some or all of its original strength. Sensitivity (the term used more commonly by engineers) is the ratio of undisturbed to disturbed (remolded) shear strength, that is, the maximum strength of

TABLE 33.16   Selected Physical Properties of Representative Andisols

| Depth (cm) | Horizon | Texture | | Sand (%) | Silt (%) | Clay (%) | Bulk Density (g cm^{-3}) | | 1500 kPa H$_2$O (%) | |
		Field	Lab				Oven Dry	Moist	Air-Dry	Moist
Bonner Series[a] (Vitrixerand)—Idaho										
0–4	A	Sil	Sil	30.6	63.8	5.6	0.75	0.68	20.3	—
4–20	Bw1	Sil	Sil	38.1	58.0	3.9	—	—	10.0	—
20–48	Bw2	Sil	Sil	39.3	59.0	1.7	0.96	0.94	9.9	—
48–69	Bw3	—	Cosl	52.6	44.1	3.3	—	—	6.8	—
69–89	Bw4	—	Cosl	52.7	44.8	2.5	0.80	0.80	6.8	—
89–152	2C	—	Cos	88.6	9.8	1.6	—	—	2.2	—
Tirau Series[b] (Hapludand)—New Zealand										
0–18	Ap	Sil	L	35	46	19	0.75	—	23.3	31.1
18–32	Bw1	Sl	Sil	33	56	11	0.71	—	15.7	33.7
32–48	Bw2	Sl	Sil	33	54	13	0.69	—	16.5	33.6
48–65	Bw3	Sl	L/Sil	30	50	20	0.79	—	22.0	34.7
65–90	2BCb	Sil	Sil	20	62	18	0.87	—	20.4	33.9
Hilo Series[c] (Hydrudand)—Hawaii										
0–18	Ap1	Sicl	Sl	66.2	29.7	4.1	—	—	28.4	54.1
18–36	Ap2	Sicl	Cos	86.8	12.1	1.1	0.82	0.51	27.7	107.2
36–60	Bw1	Sicl	Cos	91.8	7.9	0.3	1.66	0.41	25.0	112.6
60–92	Bw2	Sicl	Cos	93.1	6.9	—	1.61	0.25	25.1	132.5
92–108	Bw3	Sicl	Cos	94.9	4.7	0.4	1.41	0.30	26.0	122.4

Sil, Silt loam; Sicl, Silty clay loam; Cosl, Coarse Sandy loam; Cos, Coarse sand; Sl, Sandy loam; L, Loam.

[a] Data (Pedon #78P0553) are from Soil Survey Staff 2011.

[b] Data are from Bakker et al. (1996).

[c] Data (Pedon #89P0658) are from Soil Survey Staff 2011.

an undisturbed specimen compared with the residual strength remaining after force or strain is applied (Mitchell and Soga, 2005; Neall, 2006). In sensitive materials, the original strength is generally not recovered—unlike thixotropic materials—after removal of the force. Both sensitivity and thixotropic behavior are best expressed in Andisols having high water retention and lack of layer silicates to provide cohesion and can pose significant engineering problems.

Many Andisols exhibit irreversible changes upon drying. Allophane nanospheres collapse upon dehydration and form larger aggregates that do not break down upon rewetting. This phenomenon can cause crust formation at the soil surface during hot dry periods. It also results in well-known unreliable particle size analyses of air-dried Andisols. Normally, clay content is underestimated and sand and silt contents are overestimated (Ping et al., 1989; Dahlgren et al., 2004). However, reliable sand-, silt-, and crystalline clay-size fraction data were obtainable for andesitic Andisols in New Zealand by Alloway et al. (1992) who analyzed grain-size distributions of residual material following selective dissolution of nanominerals and organic constituents via ammonium oxalate. Other irreversible changes that occur upon desiccation include increases in bulk density, decreases in water retention, and increases in cohesive strength. Note the difference in water retention and bulk density values between dried and moist samples in Table 33.16.

Plasticity in Andisols is different from that of soils containing layer silicate clay minerals. Generally, field-moist Andisols have high liquid (60%–350%) and plastic limits (70%–180%) (Warkentin and Maeda, 1980; Neall, 2006). The low plasticity index (0–10) also clearly separates Andisols from other soils. Plasticity measurements can be used as an index of physical behavior in Andisols and as a substitute for particle size analysis, which usually is not reliable. Air-dry samples, on the other hand, often show low plasticity because of the irreversible changes on drying and behave like sandy soils.

33.3.4 Classification of Andisols

Andisols are classified on the basis of having andic soil properties, which are quantitatively defined in *Soil Taxonomy* (Soil Survey Staff, 2010). In general, andic soil properties consist of combinations of properties that develop as tephras and other volcanic materials weather. These include relatively low bulk density values ($\leq 0.90\,g\,cm^{-3}$), relatively high phosphate retention (>25%–85%), the presence of volcanic glass, and the presence of nanoscale weathering products containing Al, Fe, and Si (Soil Survey Staff, 2010). The criterion of percentage Al extracted by ammonium oxalate (Al_o) plus half the percentage Fe extracted by ammonium oxalate (Fe_o) from short-range order nanominerals is used to quantitatively define andic properties. The quantity $0.5\,Fe_o$ is used to normalize the criterion because Fe has an atomic weight (56), approximately twice that of Al (27). The majority of soils with a glass content and percentage of $Al_o + 0.5Fe_o$ that falls within the shaded area of Figure 33.16 have andic soil properties. Soils possessing at least 36 cm of material

with andic soil properties are classified as Andisols (Soil Survey Staff, 2010). It is emphasized that freshly deposited volcanic ash does not have andic soil properties and would not be classified as an Andisol. It is not until some weathering has occurred—sufficient to generate $Al_o + 0.5Fe_o$ totaling at least 0.4%—that andic properties are developed.

Andisols are separated into suborders primarily on the basis of soil moisture and/or temperature regimes (Table 33.17). Aquands are poorly drained and have a water table at or near the soil surface for much of the year. These soils are usually restricted to low-lying landscape positions and have dark surface horizons. Because of excessive wetness, Aquands typically require drainage in order to be used for crop or pasture production.

Gelands are very cold Andisols that have a mean annual soil temperature (MAST) $\leq 0°C$ (Soil Survey Staff, 2010). Relatively little is known about the distribution of these soils, but they are found at higher latitudes in areas of either current or recent volcanic activity.

Cryands are cold Andisols having a cryic temperature regime ($0°C < MAST \leq 8°C$ and summers are cool; Soil Survey Staff, 2010). They are found at higher latitudes and higher elevations. In the United States, Cryands are found mainly in Alaska and mountainous regions of the Pacific Northwest (Ping, 2000; McDaniel and Hipple, 2010). They are extensive in Iceland (Arnalds and Kimble, 2001) and also on the Kamchatka Peninsula and occur in mountainous regions elsewhere including in eastern Africa, the Andes, and (uncommonly) New Zealand (Ping, 2000; Lowe and Palmer, 2005). Globally, there are ~26 million ha of Cryands, and they are the third most common suborder, representing ~28% of Andisols (Soil Survey Staff, 1999; Wilding, 2000).

Torrands are Andisols of very dry environments where moisture for plant growth is very limited. This lack of soil moisture slows down weathering processes and leaching, thereby inhibiting development of andic soil properties. Torrands are found in Oregon and Hawaii in the United States. They are the least extensive of any of the Andisol suborders, with only ~100,000 ha occupying the global ice-free land area (<1% of Andisols).

Xerands occur in temperate regions with xeric soil moisture regimes, which are characterized by cool, moist winters and very warm, dry summers (Mediterranean climates). In the United States, they occur primarily in northern California, Oregon, Washington, and Idaho where they have formed under coniferous forest. Elsewhere, Xerands occur in scattered localities including Italy, Canary Islands, Argentina, and South Australia (Broquen et al., 2005; Lowe and Palmer, 2005; Inoue et al., 2010). Xerands are uncommon globally (~4% of Andisols).

Vitrands are the only suborder that is not defined by a climatic regime. These Andisols are relatively young and only slightly weathered. They tend to be coarse-textured and have a high content of volcanic glass that may be strongly vesicular or pumiceous. In the United States, Vitrands are found in Washington, Oregon, and Idaho. They also occur in Argentina (Broquen et al., 2005) and are common in the North Island of New Zealand (Lowe and Palmer, 2005). Globally, there are ~28 million ha of Vitrands (~31% of Andisols) making them and

TABLE 33.17   Listing of Andisol Suborders and Great Groups

Suborders	Great Groups
Aquands—poorly drained Andisols with a water table at or near the surface for much of the year	
	Gelaquands—Aquands of very cold climates (mean annual soil temperature <0°C)
	Cryaquands—Aquands of cold climates (0°C < mean annual soil temperature ≤8°C)
	Placaquands—Aquands with a thin pan cemented by Fe, Mn, and organic matter
	Duraquands—Aquands with a cemented horizon
	Vitraquands—Aquands with coarse textures dominated by glassy materials
	Melanaquands—Aquands with a thick, dark, organic matter-rich surface layer
	Epiaquands—Aquands with a perched water table
	Endoaquands—Aquands with a groundwater table
Gelands—Andisols of very cold climates (mean annual soil temperature ≤0°C)	
	Vitrigelands—Gelands with coarse textures dominated by glassy materials
Cryands—Andisols of cold climates (0°C < mean annual soil temperature ≤8°C)	
	Duricryands—Cryands with a cemented horizon
	Hydrocryands—Cryands with very high water-holding capacity
	Melanocryands—Cryands with a thick, very dark, organic matter-rich surface layer
	Fulvicryands—Cryands with a thick, organic matter-rich surface layer
	Vitricryands—Cryands with coarse textures dominated by glassy materials
	Haplocryands—other Cryands
Torrands—Andisols of very dry climates	
	Duritorrands—Torrands with a cemented horizon
	Vitritorrands—Torrands with coarse textures dominated by glassy materials
	Haplotorrands—other Torrands
Xerands—Andisols of temperature regions with warm, dry summers and cool, moist winters	
	Vitrixerands—Xerands with coarse textures dominated by glassy materials
	Melanoxerands—Xerands with a thick, very dark, organic matter-rich surface layer
	Haploxerands—other Xerands
Vitrands—slightly weathered Andisols that are coarse textured and dominated by glassy material	
	Ustivitrands—Vitrands of semiarid and subhumid climates
	Udivitrands—Vitrands of humid climates
Ustands—Andisols of semiarid and subhumid climates	
	Durustands—Ustands with a cemented horizon
	Haplustands—other Ustands
Udands—Andisols of humid climates	
	Placudands—Udands with a thin pan cemented by Fe, Mn, and organic matter
	Durudands—Udands with a cemented horizon
	Melanudands—Udands with a thick, very dark, organic matter-rich surface layer
	Hydrudands—Udands with very high water-holding capacity
	Fulvudands—Udands with a thick, organic matter-rich surface layer
	Hapludands—other Udands

Source: Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS, Washington, DC.

Udands (similar in extent) the two most extensive suborders (Soil Survey Staff, 1999; Wilding, 2000).

Ustands occur in tropical or temperature regions that have ustic soil moisture regimes. The ustic soil moisture regime is characterized by an extended dry period, but moisture is normally present at a time when conditions are suitable for plant growth (Soil Survey Staff, 2010). Ustands are of fairly limited extent in the United States and are found mainly in Hawaii. They also occur in Mexico, on Pacific islands, and in eastern Africa (Dubroeucq et al., 1998; Takahashi and Shoji, 2002). About 6.3 million ha of Ustands are known globally (~7% of Andisols).

Udands are Andisols with a udic moisture regime, which is common to humid climates. Udands are thus characterized by well-distributed precipitation throughout the year and limited periods of soil moisture stress. In the United States, Udands are found in western parts of Washington and Oregon and in Hawaii. Elsewhere, they occur commonly on other parts of the Pacific Rim including in Patagonia (Argentina), Mexico, Japan, the Philippines, Indonesia, and New Zealand (Takahashi and Shoji, 2002; Van Ranst et al., 2002; Broquen et al., 2005; Prado et al., 2007). In total, Udands occupy nearly ~28 million ha of ice-free land globally (~30% of Andisols; Soil Survey Staff, 1999; Wilding, 2000).

Each of the suborders is further separated into great groups. These are listed and described in Table 33.17. A variety of characteristics are used to define great groups, including the presence or absence of certain types of soil horizons, moisture and temperature regimes, glass content and texture, and water retention.

There are also soils of other orders that have been influenced to a lesser degree by andic materials (Parfitt, 2009). Andic soil properties have developed, but their distribution within the soil is not of sufficient thickness for the soils to be classified as Andisols. These soils are therefore classified as andic subgroups of other orders. In cases where andic materials have been extensively mixed with other parent materials, or andic properties are only weakly expressed, soils are classified as vitrandic subgroups of other soil orders (Soil Survey Staff, 2010). In the United States, such soils are extensive in the Pacific Northwest where they are transitional to higher-elevation, forested Andisols (McDaniel and Hipple, 2010).

33.3.5 Formation of Andisols

33.3.5.1 Parent Material and Stratigraphy

In most cases, the parent materials from which Andisols have formed are of explosive volcanic origin (rather than effusive) and thus fragmental and unconsolidated. Such materials range in size from ash (<2 mm) and lapilli (2–64 mm) through to large angular blocks or part-rounded bombs (>64 mm)—that is, from fine dust to boulders. Collectively, these materials are termed pyroclastic deposits or tephra (Gk *ashes*) (Alloway et al., 2007). Tephra deposits typically are loose and very coarse and thick close to source vents but become markedly finer and thinner with increasing distance away from source so that at more than ~100 km from vent most comprise mainly ash-sized material (equivalent to sand or finer particles). The accumulation at a particular site of numerous tephra deposits from sequential eruptions from one or more volcanoes leads usually to the formation of Andisols with distinctive layered profiles and buried soil horizons, forming multisequal profiles (Figure 33.17). Such layered profiles, together with their andic soil properties, are special features of Andisols. Study of the layers and attaining ages for them (tephrostratigraphy) is an important aspect of understanding Andisol formation (Lowe and Palmer, 2005; Lowe and Tonkin, 2010).

During periods of quiescence between major eruptions, soil formation takes place, transforming the characteristics of the unmodified tephra via normal top-down pedogenesis whereby the materials are altered in a downward-moving front to form subsoil horizons. However, when new tephras are added to the land surface, upbuilding pedogenesis takes place. The frequency and thickness of tephra accumulation (and other factors) determine how much impact the top-down processes have on the ensuing profile character, and if "developmental" or "retardant" upbuilding, or both, will take place. Two contrasting scenarios can be considered.
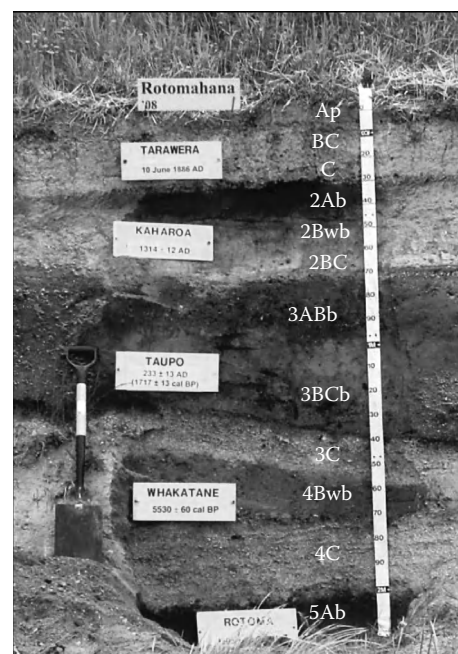
FIGURE 33.17 Udivitrand (Rotomahana soil) from New Zealand composed of multiple tephra deposits and associated buried soils and horizons; scale divisions = 10 cm; the oldest tephra layer (Rotoma) at bottom of exposure is 9505 ± 25 calibrated years BP.

In the first scenario, successive thin tephra deposits (ranging from millimeters to centimeters in thickness) accumulate incrementally and relatively infrequently, so that developmental upbuilding ensues. Such a situation occurs typically at distal sites. The thin materials deposited from each eruption become incorporated into the existing profile over time. Top-down pedogenesis continues as the tephras accumulate but its effects are lessened because any one position in the sequence is not exposed to pedogenesis for long before it becomes buried too deeply for these processes to be effective as the land surface gently rises (Figure 33.18). This history thus leaves the tephra materials with a soil fabric inherited from when the tephra was part of the surface A horizon or subsurface Bw horizon (Lowe and Palmer, 2005; Lowe and Tonkin, 2010). Each part of the profile has been an A horizon at one point, as illustrated in Figure 33.18.

In the second scenario, tephra accumulation is more rapid, as occurs in locations close to volcanoes, or when a much thicker layer (more than a few tens of centimeters) is deposited from a powerful eruption. In the latter case, the antecedent soil is suddenly buried and isolated beyond the range of most soil-forming processes (i.e., it becomes a buried soil horizon or paleosol). A new soil will thus begin forming at the land surface in the freshly deposited material. This scenario typifies retardant upbuilding, which recognizes that the development of the now-buried soil has been retarded or stopped, and the pedogenic "clock" reset to time zero for weathering and soil formation to start afresh. An example of a multisequal Andisol profile formed via retardant upbuilding pedogenesis since ~9500 calendar
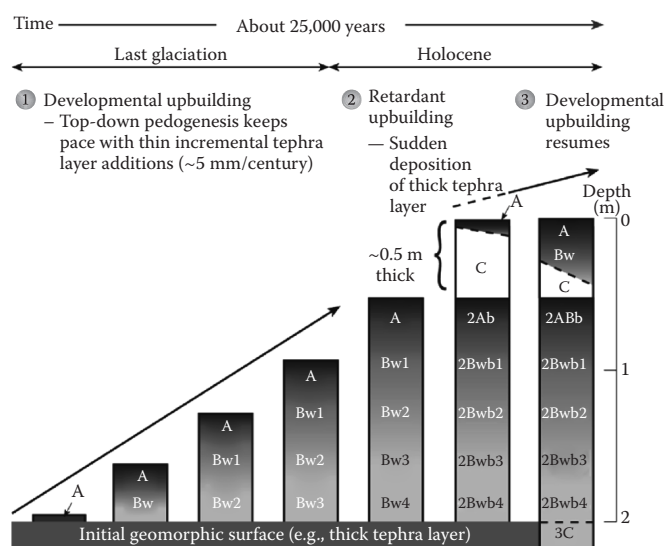
FIGURE 33.18 Model of upbuilding pedogenesis profiles over ~25,000 years in the Waikato region, New Zealand. (From Lowe, D.J., and P.J. Tonkin. 2010. Unravelling upbuilding pedogenesis in tephra and loess sequences in New Zealand using tephrochronology, p. 34–37. *In* Proc. IUSS 19th World Soil Congress, Symposium 1.3.2 Geochronological techniques and soil formation, Brisbane, Australia August 1–6, 2010. Published on DVD and online at: http://www.iuss.org. With permission.)

FIGURE 33.19 Volcanic glass compositions and dissolution of Al and Si to form allophane. (After Hiradate, S., and S.-I. Wada. 2005. Weathering processes of volcanic glass to allophane determined by ^{27}Al and ^{29}Si solid-state NMR. Clays Clay Miner. 53:401–408.)

(cal) years ago is shown in Figure 33.17. Each of five successive tephra deposits shows the imprint of top–down pedogenesis, as depicted by their horizonation. But the sudden arrival of a new deposit every few thousand years or so on average has buried and effectively isolated each of the weakly developed "mini" soil profiles as the land surface rises.

In addition to stratigraphic factors, Andisols are markedly affected by the mineralogical and physicochemical compositions of the parent tephras, or associated deposits derived from remobilization of volcanic and other material (collectively termed volcaniclastic deposits). For example, the marked influence of windblown dust (mainly basaltic glass) on soil properties in Iceland was described by Arnalds (2010). Tephras differ widely according to the chemical makeup of magmas of the volcanoes that generated them. The chemistry of magmas, especially Si content, governs the way a volcano erupts. Three main magma types and resulting eruptives can be identified according to their chemical composition—rhyolitic ($\gtrsim$70% SiO_2), andesitic/dacitic (~50%–70% SiO_2), and basaltic ($\lesssim$~50% SiO_2). All magmas generate volcanic glass, a noncrystalline, easily weatherable mineraloid, and various other primary silicate minerals (Shoji et al., 1993; Nanzyo, 2002; Smith et al., 2006; Alloway et al., 2007; De Paepe and Stoops, 2007). Glass especially provides much of the Si and Al required to dissolve and re-form as allophane or other aluminosilicate clay minerals, and the amounts differ according to the magma composition as shown in Figure 33.19. Feldspars also release Si and Al via weathering for clay formation.

Generally, the basaltic and intermediate tephras tend to weather more readily than rhyolitic tephras, and in all cases glasses weather very quickly (Neall, 1977; Kirkman and McHardy, 1980; Colman and Dethier, 1986; Hodder et al., 1996; Shikazono et al., 2005). Compared with hard rock, the fragmental tephra components, especially vesicular glass and pumice fragments, have a much greater surface area and higher porosity and permeability, and so break down to constituent compounds very readily. As this weathering occurs, various elements including Si and Al are released into chemical solution either for subsequent leaching, complexing with humic materials, plant uptake, or synthesis (neoformation) into relatively stable nanominerals and other clay minerals (Lowe, 1986; Vacca et al., 2003; Dahlgren et al., 2004). Andisols developed on (base-rich) basaltic eruptives are inherently more fertile than those on more siliceous eruptives (Wolff-Boenisch et al., 2004). Weakly weathered tephras give rise to the glass-rich Andisols that belong to the Vitrand suborder.

33.3.5.2 Climate

Climate plays an important role in Andisol formation, and soil moisture and temperature regimes are used to differentiate all the suborders except Vitrands (Ping, 2000). The majority of Andisols, as reported previously, are found under udic soil moisture regimes, and around half are found in the tropics, with the rest occurring in temperate or boreal regions. Climatic conditions help govern the combinations of processes, collectively referred to as andisolization, that occur in soils developing on tephras. Andisolization is the in situ formation of andic soil materials comprising nanominerals composed of "active" Al, Si, Fe, and humus (Dahlgren et al., 2004). The process is discussed in more detail in Section 33.3.6.

The essential conditions for the formation of allophane are the activity of silicic acid in the soil solution, the availability of Al species, and the opportunity for coprecipitation (Figure 33.19). These conditions are controlled largely by the leaching regime, the organic cycle, and pH, which, in turn, are potentially influenced by numerous environmental factors including rainfall, drainage, depth of burial, parent tephra composition and accumulation rate, dust accession, type of vegetation and supply of humic substances, and human activities (such as burning vegetative cover), together with thermodynamic and kinetic factors (Parfitt and Saigusa, 1985; Lowe, 1986, 1995; Hodder et al., 1990; Chadwick et al., 2003; Dahlgren et al., 2004; Parfitt, 2009). Availability of Al, derived mainly from the dissolution of glass or feldspars, is assumed to be unlimited in this model, though potentially more is available from andesitic and especially basaltic tephras than rhyolitic tephras (Lowe, 1986; Shoji et al., 1993; Cronin et al., 1996). In contrast, in pedogenic environments rich in organic matter and with pH ≤ 5, humus effectively competes for dissolved Al, leaving little Al available for coprecipitation with Si to form allophane or halloysite (Shoji et al., 1993; Dahlgren et al., 2004; see Section 33.3.5.4).

In New Zealand, both mineralogical and soil solution studies on soils derived from tephras extending across a rainfall gradient showed that rainfall, coupled with through-profile drainage, helps govern Si concentration [Si] in soil solution and thus the likelihood of allophane being formed or not (Parfitt et al., 1983; Singleton et al., 1989; Parfitt, 2009). The Si leaching model shown in Figure 33.20 is summarized as follows: where [Si] is less than ~10 ppm (mg L^{-1}), allophane is formed; where [Si] is greater than ~10 ppm, halloysite is formed. If [Si] is close to ~10 ppm, then either allophane or halloysite may predominate. A profile throughflow threshold of approximately 250 mm year^{-1} of drainage water likely controls [Si]—less than ~250 mm year^{-1} means that the loss of Si is insufficient for Al-rich allophane to form and halloysite (or Si-rich allophane) forms instead (Parfitt et al., 1984; Lowe, 1995).

It is emphasized, however, that allophane and other nanominerals can form under lower rainfalls (~700 mm year^{-1}) such as in Mexico, Iceland, and South Australia (Arnalds et al., 1995;

FIGURE 33.20 Simplified allophane–halloysite rainfall leaching model.

Dubroeucq et al., 1998; Ugolini and Dahlgren, 2002; Lowe and Palmer, 2005). As well, other factors such as parent tephra composition and vegetation are also important.

33.3.5.3 Topography

Andisols are found on all types of topography and at a wide range of elevations (from sea-level to >3000 m). Because of their strong association with the products of volcanism, Andisols occur most commonly in volcanic landscapes, which can range from mountainous to hilly terrains on the flanks of stratovolcanoes, dome complexes, shield volcanoes, and scoria or tuff cones through to rolling or essentially flat-lying landscapes, some associated with plateaus of tephra-draped welded ignimbrite sheets derived from large caldera eruptions, such as occur for example in the western United States, Japan, and northern New Zealand (Lowe and Palmer, 2005). Because tephras are carried long distances by the wind, and can easily be reworked by wind and water, they can be deposited in a variety of sedimentary and other landscapes including dune plains, fluvial terraces or plains, and flat intermontane valleys (Ping, 2000; Lowe and Palmer, 2005). The thickest tephra-fall deposits tend to be downwind from the volcanic source areas on rolling or flattish land surfaces (Lowe and Palmer, 2005). On hilly landscapes, tephras are variable in thickness because of erosion (especially during the last glaciation), and profiles typically are thinner on slopes than on more stable geomorphic surfaces (Neall, 2006). The effects of specific topographic positions on Andisol formation have been evaluated, for example, by Navarette et al. (2008) in the Philippines.

33.3.5.4 Vegetation

Because Andisols occur in all moisture and temperature regimes on different landscapes at different elevations, vegetation is accordingly highly variable. In general, Andisols are usually moderately acid to very acid and so vegetation associated with acid soils will predominate. Because of their relatively high water-holding capacities, Andisols, even in xerophytic areas, are likely to support more luxuriant vegetation than other nonandic soils in the same environment. When the vegetation cover is grass, a more humus-rich profile usually develops. A thick, dark surface known as a melanic epipedon is common in many Andisols that have formed under grasses. Melanic epipedons contain ≥6% organic C as a weighted average, have black soil colors (melanic index ≤1.70), and attain a thickness of ≥30 cm (Soil Survey Staff, 2010). Andisols with melanic epipedons are usually classified as Melanudands, an example of which is shown in Figure 33.14b.

Although the vegetation on Andisols is quite variable, it clearly has a major influence on the type of Andisol formed (Ping, 2000). In the Pacific Northwest and parts of Alaska, and in New Zealand, Udands generally formed under forest. However, Udands are known to form under other vegetation types such as grasses in south central Alaska (Ping, 2000) and in Japan, where pampas grass (*Miscanthus sinensis*), known as "susuki,"
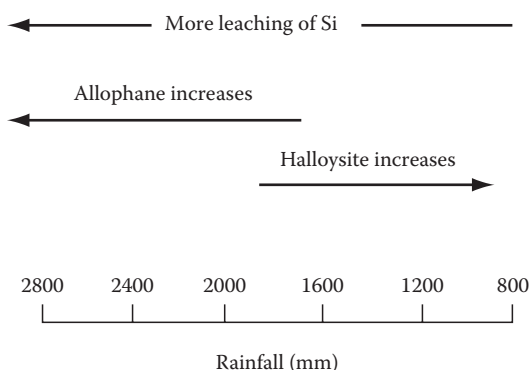
has formed Melanudands, Sitka spruce (*Picea sitchensis*) Hydrudands, and blue joint grass (*Calamagrostis canadensis*) Fulvudands. The organic matter is dominated by humic acids in the first cases and by fulvic acids in the latter (Ping et al., 1989; Shoji et al., 1993; Nanzyo, 2002; Dahlgren et al., 2004).

In Japan and elsewhere, the humic acids in many Andisols and associated soils are characterized by their stability and aromatic (humified) structure (Shoji et al., 1993; Dahlgren et al., 2004; Hiradate et al., 2004; Parfitt, 2009). These features arise from the presence of labile and active metals, chiefly Al and Fe, which are able to bind with humic substances to form nanoparticles of Al- and Fe-humic acid complexes that are very resistant to degradation or leaching (Hiradate et al., 2004; Basile-Doelsch et al., 2005). As well as strong bonding, the metal-humic acid complexes are also evidently protected by Al toxicity to microorganisms or enzymes, physical encasement within abundant micropeds, and a P deficiency of microorganisms arising from high P retention (Ugolini and Dahlgren, 2002; Parfitt, 2009). The resultant very dark or black melanic epipedons can contain as much as 25%–30% organic carbon (Drijber and Lowe, 1990; Nanzyo, 2002; Hiradate et al., 2004).

Most Xerands, such as those in central Washington, Oregon, and northern California, have formed under mixed vegetation of grass and forest whereas those in South Australia formed under dense to semi-open woodland and grassland/shrubland or fernland (Ping, 2000; Lowe and Palmer, 2005; Takesako et al., 2010). Torrands formed under shrubby or grass vegetation, and Vitrands under grass or forest in North America and Mexico and under broadleaf–podocarp forest in New Zealand. Cryands generally formed under forest, but in Iceland and other high-latitude polar regions including the Aleutians, these soils formed under tundra or moss, heath, grass, and "desert" covers (Arnalds et al., 1995; Ping, 2000; Arnalds and Kimble, 2001). In the tropics, the vegetation is often rain forest, particularly on the slopes of volcanoes where grassland and savanna types also occur.

Some black A horizons on Andisols in New Zealand, typically ≤20 cm thick, have been attributed largely to the effects of bracken fern (*Pteridium* spp.), which replaced much of the original forest after human-induced deforestation by Polynesian burning from about AD 1300 (Newnham et al., 1999; McGlone et al., 2005; Lowe, 2008). In northern Idaho, establishment of bracken fern following forest canopy removal is associated with decadal-scale changes in soil properties. These include increased soil C, darker soil colors, lower pH, and increased organic forms of active Al (Johnson-Maynard et al., 1997).

33.3.5.5 Time

Andisols generally are found on relatively young parent materials in humid climates because, on older and more stable landscapes, more weathering would normally have taken place resulting in the transformation to other soil orders such as Ultisols or Oxisols. The drier and cooler the environment, the longer it usually takes for an Andisol to form (Shoji et al., 2006). Tephras or volcanic materials, either very young or in extremely dry areas, and not sufficiently weathered to form the minimum requirements for andic soil materials, are placed in the Entisol order. An A horizon can form in newly deposited tephra in a few hundred years or so whereas the development of BC and then Bw horizons beneath, it may take several hundreds to several thousands of years according to environmental conditions and the nature of the tephra (Shoji et al., 1993; Ping, 2000; Lowe and Palmer, 2005; Lowe, 2008). However, in northeast Kodiak Island, Alaska, a Cryand formed on tephra erupted in the 1912 Mount Katmai event already has thin E and Bw horizons (Ping, 2000), and soils on young basaltic tephra surfaces in Iceland (<100 years) are classified as Andisols (Arnalds and Kimble, 2001).

In situations where just a single eruption event has taken place to produce a parent deposit, then Andisols developed in that material have the same age as the eruption. However, where Andisols comprise upbuilding sequences of tephras and soil horizons (as commonly occurs), the maximum age of the soil-profile constituents depends on the depth at which the profile "base" is drawn. In considering the uppermost 1–2 m of Andisols, some generalizations about their *composite* ages can be made. For example, in Japan, many Andisols are ~5,000–10,000 cal years old although some are considerably older (~25,000 cal years old; Takesako and Muranaka, 2006). In southeast Alaska, Cryands may be up to ~15,000 cal years old (Ping, 2000). In Iceland, most of the Andisols range in age from a few decades to ~10,000 cal years old (Arnalds et al., 1995). In New Zealand, there is a much wider range of ages on Andisols. Most of the Vitrands are ~700 years old (on rhyolitic Kaharoa Tephra) and ~1,800 years old (on rhyolitic Taupo Tephra); extensive Udands date back ~20,000–25,000 cal years and some are as old as ~60,000 years. In all these countries and others, tephrochronology—the identification and correlation of tephras and their application as a linking and dating tool (Alloway et al., 2007; Lowe, 2011)—has been used to provide age models for the Andisols. These age models rely mainly on radiocarbon and other techniques including dendrochronology and depositional age-depth modeling as well as wiggle-match dating (Lowe, 2011). Historical accounts provide ages for the youngest tephras (Alloway et al., 2007).

33.3.6 Pedogenic Processes in Andisols

The importance of upbuilding and polygenesis in the formation of Andisols has already been described. In this section, the collective process of andisolization in tephra materials with a significant content of volcanic glass is outlined. Andisolization leads to the formation of a fine colloidal fraction dominated by nanoscale minerals, notably allophane, ferrihydrite, and metal–humus complexes, with typically subordinate amounts of imogolite. Nanoscale minerals are defined as having at least one dimension in the nanorange, which is 1–100 nm (Hochella, 2008). These nanoscale minerals are composed of active forms of Al, Si, Fe, and humus and impart andic soil properties (Theng and Yuan, 2008). Although volcanic glass is a common

component in many Andisols, it is not a requirement of the Andisol order and some soils develop andic properties without the influence of glass (Soil Survey Staff, 1999). Soil solution and micromorphological studies show that, in contrast to podzolization, translocation of Al, Fe, organic matter, and clays is normally minimal in Andisols (Jongmans et al., 1994; Soil Survey Staff, 1999; Ugolini and Dahlgren, 2002; Stoops, 2007).

Allophane and imogolite (both aluminosilicates) and ferrihydrite (an iron hydroxide) are formed by the synthesis of soluble Al, Si, or Fe that are released from glass or other minerals (mainly feldspars or various mafic minerals) via rapid dissolution and hydrolysis by carbonic acid (H_2CO_3) normally under moderately acid and humid conditions. As soil solutions become rapidly oversaturated with respect to these nanominerals, they are precipitated preferentially because their nucleation is kinetically favored over that of less-soluble, long-range-order (crystalline) phases (Ugolini and Dahlgren, 2002; Theng and Yuan, 2008). Thermodynamic stability diagrams show that imogolite and very likely Al-rich allophane are more stable than halloysite over a wide range of Si activity, with the latter more stable than both imogolite and Al-rich allophane only at high Si activity (Farmer et al., 1991; Lowe and Percival, 1993; Hodder et al., 1996; Harsh, 2000). The central role of the Si leaching model in helping govern Si activity (i.e., concentration of $Si(OH)_4$ or monosilicic acid) was discussed earlier. Al-rich allophane forms at pH values between ~5 and 7, a value of at least ~4.8 being required for it to precipitate (Lowe, 1995; Dahlgren et al., 2004). Al undergoes hydrolysis and polymerizes in gibbsite-like octahedral sheets as $(OH)_3Al_2$ and

then combines with Si in tetrahedral coordination as O_3SiOH to form imogolite or allophane (Figure 33.19; Hiradate and Wada, 2005; Theng and Yuan, 2008). Allophane can form rapidly and has been observed to precipitate in a few months on the face of open soil pits containing rhyolitic tephra (Parfitt, 2009).

Note that a long-standing previous model in which allophane was suggested to weather or "crystallize" by solid-state transformation to form halloysite with time (because the latter often occurs in older, that is, deeper, tephras and associated soil horizons) has been largely discounted and replaced by this Si leaching model and its variants. For allophane to alter to halloysite requires a complete rearrangement of the atomic structures, and this could only occur by dissolution and reprecipitation processes because the allophane would need to "turn inside out" so that Si-tetrahedra are on the outside, not inside, the curved Al-octahedral sheets (Figure 33.21; Parfitt et al., 1983; Ildefonse et al., 1994; Lowe, 1995; Churchman, 2000; Hiradate and Wada, 2005; Parfitt, 2009). The effect of time is clearly subordinate because glass can weather directly via dissolution either to allophane or halloysite depending on glass composition and both macro- and microenvironmental conditions, not time (Lowe and Percival, 1993; Lowe, 1995; Parfitt, 2009). As well, allophane can persist for hundreds of thousands of years under favorable conditions (Stevens and Vucetich, 1985; Churchman and Lowe, 2011, Section 20.1).

The formation of ferrihydrite in Andisols and other soils has been described in a number of papers including Schwertmann and Taylor (1989), Childs (1992), Lowe and Percival (1993),
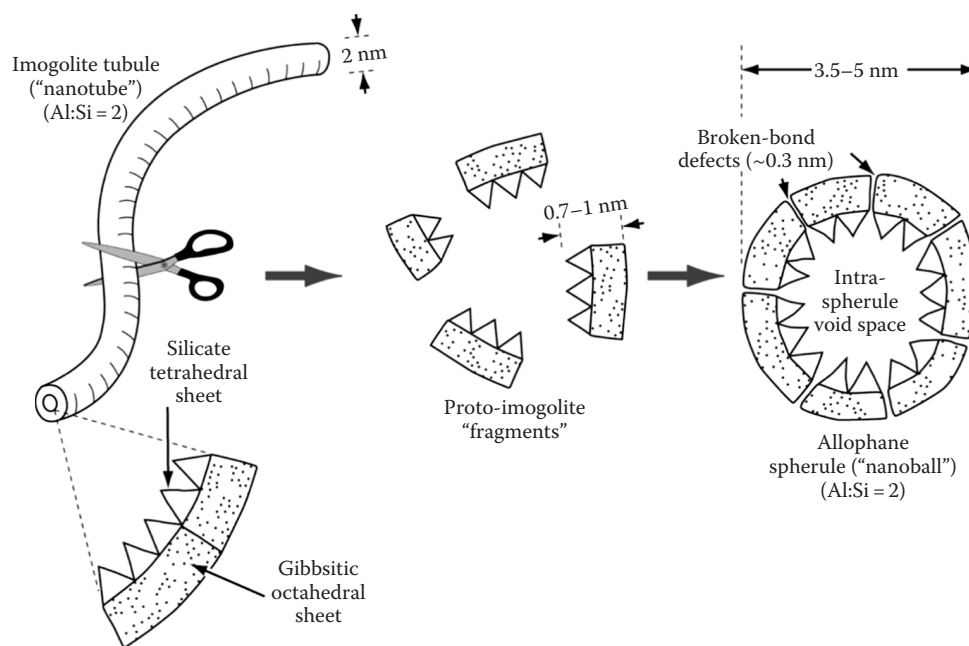
FIGURE 33.21 Diagram of imogolite tubes and allophane spherules. "Proto-imogolite" fragments have an imogolite atomic structure over a short range and link to form porous, hollow nanospheres with water molecules occupying the intra-spherule interior as well as being adsorbed to the outer AlOH surface (Parfitt, 2009). (After Lowe, D.J. 1995. Teaching clays: From ashes to allophane, p. 19–23. *In* G.J. Churchman, R.W. Fitzpatrick, and R.A. Eggleton (eds.) Clays: Controlling the environment. CSIRO Publishing, Melbourne, Australia.)

Churchman (2000), Bigham et al. (2002), Nanzyo (2002), and Dahlgren et al. (2004). After release of structural Fe from mafic primary minerals such as pyroxenes, amphiboles, biotite, and olivine (and presumably also from glass), Fe^{2+}-containing soil water oxidizes relatively quickly to Fe^{3+}, which immediately hydrolyzes to ferrihydrite. This occurs in the presence of organic matter, silicate, or phosphate, all of which inhibit the formation of more thermodynamically stable crystalline iron oxides. Ferrihydrite can redissolve readily by reduction (Schwertmann, 2008).

The availability of Al^{3+} is the critical factor regulating the formation of nonallophanic Andisols, which form preferentially in environments rich in organic matter and with pHs of ~5 or less, and which typically contain 2:1 layer silicate clays (Shoji et al., 1985, 1993; Nanzyo, 2002; Ugolini and Dahlgren, 2002; Dahlgren et al., 2004). At these pHs, organic acids are the dominant proton donors lowering pH and aqueous Al^{3+} activities through the formation of Al–humus complexes (also Fe–humus complexes, but less commonly). Under these conditions, carboxyl groups of humus and the 2:1 layer silicates (derived mainly from windblown dust) effectively compete for dissolved Al, leaving little Al available for coprecipitation with Si to form allophane or imogolite (Mizota and van Reeuwijk, 1989; Dahlgren et al., 2004). The preferential incorporation of Al into Al–humus complexes (and hydroxyl-Al interlayers of 2:1 layer silicates) has been termed the *antiallophanic* effect (Shoji et al., 1993). In Japan, allophanic Andisols tend to be formed where parent tephras are more basic (basaltic to andesitic) and rainfall is <1000 mm per year, so that higher pH values (>5) are favored (Dahlgren et al., 2004). Their distribution pattern is also attributed in part to proximity to volcanoes and the frequency therefore of tephra accretion vs. exotic dust accession (Saigusa and Matsuyama, 2004): sites closer to volcanoes effectively received a "top up" of Al through more regular deposition of weatherable tephras (resulting in allophanic Andisols) than at distal sites where the Al became quickly depleted (resulting in nonallophanic Andisols) through both humus complexing and "dilution" from the ongoing fallout of 2:1 mineral-bearing dust blown from China.

Changes in elemental composition during andisolization in Japan were studied by Nanzyo and Takahashi (2005).

33.3.7 Management of Andisols

Andisols support a much greater population than their limited extent would suggest. One reason for this is that crop productivity on Andisols is very high due to good physical properties and relatively high native fertility in some locations. Periodic additions of fresh tephra, especially when intermediate to basic in composition, can resupply potential nutrients and maintain favorable fertility levels in areas where other soils tend to be depleted in nutrients. For example, a "dusting" of andesitic ash fallout from the 1995–1996 Mt. Ruapehu eruptions in New Zealand added between 30 and 1500 kg ha^{-1} of sulfur as well as small quantities of Se, Mg, and K to substantial land areas (Lowe and Palmer, 2005). Consequently, in many areas, few inputs with the exception of P have been required for successful agricultural production on many Andisols. Andisols are also productive forest soils. In the Pacific Northwest region of United States, the presence of tephra (ash) mantles and their water-holding capacity is closely linked to forest productivity (Kimsey et al., 2008). Vitrands in New Zealand also support very productive plantation forestry (Ross et al., 2009).

Nevertheless, in other locations such as New Zealand where rainfall and leaching are generally high, and where siliceous tephras low in metals are common, many Andisols have low fertility because of a range of chemical limitations both acquired and inherited. As well as P fixation, exchangeable bases (especially K) tend to be low and other elemental deficiencies may include Mg, S, Co, and others (Lowe and Palmer, 2005). The high productivity of these soils is thus maintained by regular fertilizer input in combination with their generally excellent physical properties.

Most Andisols are acid and where acidity is particularly high, high levels of exchangeable Al may inhibit crop growth (Al toxicity). This is especially a problem in nonallophanic Andisols (Shoji et al., 1993; Shoji and Takahashi, 2002; Dahlgren et al., 2004). In such instances, addition of lime can reduce the amounts of organically complexed Al and exchangeable Al (Takahashi et al., 2006), or acid-tolerant crops must be selected (Michaelson and Ping, 1987; Nanzyo, 2002; Shoji and Takahashi, 2002). Where Andisols have been acidified by long-term N inputs, serious consequences for crop production can arise because as pH decreases, CEC also decreases. As a result, such soils are able to hold few base cations (Ca, Mg, and K) for plant uptake (Sumner and Hylton, 1993).

Requirements for N vary according to the ease with which organic matter is mineralized in these soils. Nitrogen mineralization is slower in allophanic Andisols than in nonallophanic Andisols because organic matter interacts with and is stabilized by allophane in the former (Parfitt, 2009). Liming often promotes the release of substantial amounts of N from nonavailable reserves, thereby improving N fertility.

As noted above, the high P-fixing capacities of most Andisols necessitate inputs of P for successful agricultural production (Michaelson and Ping, 1990; Dahlgren et al., 2004; Lowe and Palmer, 2005). Such inputs in the form of rock phosphate are effective on very acid Andisols but in any event, soluble P sources should be band-placed or otherwise protected from intimate mixing with the soil (Sumner and Hylton, 1993).

Other nutrients may sometimes be deficient in Andisols depending of the chemical composition of the parent material. One notable example occurs in the central North Island of New Zealand where Vitrands, formed in the silica-rich Taupo and Kaharoa tephras, are deficient in micronutrients including Co, Se, Cu, B, I, and Mo. Inherently low Co in these pumiceous parent tephras led to low Co levels in soils and herbage, and ultimately to a deficiency in ruminant animals (sheep and cows)

that developed a serious and commonly fatal wasting disease known as "bush sickness" (Lowe and Palmer, 2005; Neall, 2006). This term referred to a variety of symptoms exhibited by livestock stemming from their inability to produce vitamin B12, of which Co is an essential component (Cornforth, 1998). The association of this condition with the Vitrands was recognized in the early 1930s (Grange and Taylor, 1932), and the cause identified a few years later. Acquired Co deficiency also occurred in some Udands and non-Andisol orders where high rainfall and strong leaching were the cause. Subsequent use of cobaltized superphosphate, and other methods, corrected the problem (Neall, 2006).

Andisols, in general, are noted for their good physical properties including high water-holding capacity, free drainage, good tilth and friability, resistance to water erosion, stable aggregation, low bulk density and high porosity, and hence good aeration. Consequently, they are usually excellent media for seedling emergence and root growth and proliferation. It should be emphasized that these properties can and often are altered through management (e.g., Candan and Broquen, 2009). Where uncultivated, undisturbed Andisols may be resistant to erosion, but loss of vegetative cover, compaction, and subsequent reduction in infiltration can lead to severe runoff and erosion including deep gullying in Vitrands (Lowe and Palmer, 2005). Agricultural production on Andisols can result in erosion and compaction, and no-till or reduced tillage farming practices are recommended to minimize these problems in agricultural systems (Dahlgren et al., 2004). Heavy livestock grazing can also compact Andisols (Schlichting, 1988). In the Pacific Northwest of United States, compaction resulting from mechanized timber harvesting can reduce infiltration and water-holding capacity of Andisols (Cullen et al., 1991). Severe erosion from harvested sites can also significantly reduce water-holding capacity to the point where site productivity is drastically diminished (McDaniel et al., 2005). In Iceland, Cryands have been overgrazed, denuded, and subject to severe wind erosion (Arnalds et al., 1995; Arnalds and Kimble, 2001).

As noted earlier, properties of sensitivity and thixotropy can cause Andisols to behave in a fluid-like manner when loading pressures are applied. Thus, Andisols are susceptible to failure when disturbed on slopes, which can cause them to reach the liquid limit, generating landslides (Basile et al., 2003; Neall, 2006). Mantling of harder bedrock by andic soils and the layering associated with tephra deposits and buried soils can form planes of failures (Arnalds, 2008).

Because of their sorptive capacity for anions and metals, Andisols offer potential environmental benefits as active filters. High nitrate adsorption by allophanic Andisols was correlated with retarded nitrate transport rates under coffee plantations in Costa Rica (Reynolds-Vargas et al., 1994; Ryan et al., 2001). Reduced phosphorus leaching losses from surface-applied domestic effluent in New Zealand have been attributed the relatively high P sorption capacity of Andisols (Barton et al., 2005). Similarly, the sorption of arsenate by Andisols offers potential for purifying drinking water or for remediating As-contaminated soils (Arai et al., 2005; Theng and Yuan, 2008). McLeod et al. (2008) have also demonstrated that New Zealand Andisols leached very small amounts of surface-applied microbial tracers via bypass flow, attributable to their sorptive capacity and the fine, porous nature of the soils. These and other studies suggest that Andisols may play an important role in reducing groundwater contamination.

Acknowledgments

References

Adamo, P., L. Denaix, F. Terribile, and M. Zampella. 2003. Characterization of heavy metals in contaminated volcanic soils of the Solofrana river valley (southern Italy). Geoderma 117:347–366.

Alloway, B.V., G. Larsen, D.J. Lowe, P.A.R. Shane, and J.A. Westgate. 2007. Tephrochronology, p. 2869–2898. *In* S.A. Elias (ed.) Encyclopaedia of Quaternary science. Elsevier, London, U.K.

Alloway, B.V., V.E. Neall, and C.G. Vucetich. 1992. Particle size analyses of late Quaternary allophane-dominated andesitic deposits from New Zealand. Quat. Int.13–14:167–174.

Arai, Y., D.L. Sparks, and J.A. Davis. 2005. Arsenate adsorption mechanism at the allophane–water interface. Environ. Sci. Technol. 39:2537–2544.

Arnalds, O. 2004. Volcanic soils of Iceland. Catena 56:3–20.

Arnalds, O. 2008. Andosols, p. 39–46. *In* W. Chesworth (ed.) Encyclopaedia of soil science. Springer, Dordrecht, The Netherlands.

Arnalds, O. 2010. Dust sources and deposition of aeolian materials in Iceland. Icelandic Agric. Sci. 23:3–21.

Arnalds, O., F. Bartoli, P. Buurman, E. Garcia-Rodeja, H. Oskarsson, and G. Stoops (eds.). 2007. Soils of volcanic regions in Europe, 644pp. Springer, Berlin, Germany.

Arnalds, O., C.T. Hallmark, and L.P. Wilding. 1995. Andisols from four different regions of Iceland. Soil Sci. Soc. Am. J. 59:161–169.

Arnalds, O., and J. Kimble. 2001. Andisols of deserts in Iceland. Soil Sci. Soc. Am. J. 65:1778–1786.

Bakker, L., D.J. Lowe, and A.G. Jongmans. 1996. A micromorphological study of pedogenic processes in an evolutionary soil sequence formed on late Quaternary rhyolitic tephra deposits, North Island, New Zealand. Quat. Int. 34–36:249–261.

Barton, L., L.A. Schipper, G.F. Barkle, M. McLeod, T.W. Speir, M.D. Taylor, A.C. McGill, A.P. van Schaik, N.B. Fitzgerald, and S.P. Pandey. 2005. Land application of domestic effluent onto four soil types: Plant uptake and nutrient leaching. J. Environ. Qual. 34:635–643.

Basile, A., G. Mele, and F. Terribile. 2003. Soil hydraulic behaviour of a selected benchmark soil involved in the landslide of Sarno 1998. Geoderma 117:331–346.

Basile-Doelsch, I., R. Amundson, W.E.E. Stone, C.A. Masiello, J.Y. Bottero, F. Colin, D. Borschneck, and J.D. Meunier. 2005. Mineralogical control of organic carbon dynamics in a volcanic ash soil on La Réunion. Eur. J. Soil Sci. 56:689–703.

Bigham, J.M., R.W. Fitzpatrick, and D.G. Schulze. 2002. Iron oxides, p. 323–366. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA Book Series No. 7. SSSA, Madison, WI.

Brigatti, M.F., E. Galan, and B.K.G. Theng. 2006. Structures and mineralogy of clay minerals, p. 19–86. *In* F. Bergaya, B.K.G. Theng, and G. Lagaly (eds.) Handbook of clay science. Developments in clay science 1. Elsevier, Amsterdam, The Netherlands.

Broquen, P., J.C. Lobartin, F. Candan, and G. Falbo. 2005. Allophane, aluminium, and organic matter accumulation across a bioclimatic sequence of volcanic ash soils of Argentina. Geoderma 129:167–177.

Buol, S.W., R.J. Southard, R.C. Graham, and P.A. McDaniel. 2003. Soil genesis and classification. 5th edn. Blackwell–Iowa State Press, Ames, IA.

Candan, F., and P. Broquen. 2009. Aggregate stability and related properties in NW Patagonian Andisols. Geoderma 154:42–47.

Chadwick, O.A., R.T. Gavenda, E.F. Kelly, K. Ziegler, C.G. Olson, W.C. Elliott, and D.M. Hendricks. 2003. The impact of climate on the biogeochemical functioning of volcanic soils. Chem. Geol. 202:195–223.

Childs, C. 1992. Ferrihydrite: A review of structure, properties and occurrence in relation to soils. Z. Pflanzenernahr. Bodenkd. 154:1–7.

Childs, C.W., N. Matsue, and N. Yoshinaga. 1991. Ferrihydrite in volcanic ash soils of Japan. Soil Sci. Plant Nutr. 37:299–311.

Churchman, G.J. 2000. The alteration and formation of soil minerals by weathering, p. F3–F76. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Churchman, G.J., and D.J. Lowe. 2012. Alteration, formation, and occurrence of minerals in soils. *In* Y. Li and M.E. Sumner (eds.) Handbook of soil science. 2nd edn. Taylor & Francis, Boca Raton, FL.

Colman, S.M., and D.P. Dethier. 1986. An overview of the rates of chemical weathering, p. 1–17. *In* S.M. Colman and D.P. Dethier (eds.) Rates of chemical weathering of rocks and minerals. Academic Press, Orlando, FL.

Cornforth, I. 1998. Practical soil management. Lincoln University Press, Canterbury, and Whitirea Publishing, Wellington, New Zealand.

Cronin, S.J., V.E. Neall, and A.S. Palmer. 1996. Investigation of an aggrading paleosol developed into andesitic ring-plain deposits, Ruapehu volcano, New Zealand. Geoderma 69:119–135.

Cullen, S.J., C. Montagne, and H. Ferguson. 1991. Timber harvest trafficking and soil compaction in western Montana. Soil Sci. Soc. Am. J. 55:1416–1421.

Dahlgren, R.A., M. Saigusa, and F.C. Ugolini. 2004. The nature, properties, and management of volcanic soils. Adv. Agron. 82:113–182.

De Paepe, P., and G. Stoops. 2007. A classification of tephra in soils. A tool for soil scientists, p. 119–128. *In* O. Arnalds, F. Bartoli, P. Buurman, H. Oskarsson, G. Stoops, and E. Garcia Rodeja (eds.) Soils of volcanic regions in Europe. Springer, Dordrecht, The Netherlands.

Drees, L.R., L.P. Wilding, N.E. Smeck, and A.L. Senkayi. 1989. Silica in soils: Quartz and disordered silica polymorphs, p. 913–974. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA Book Series No. 1. SSSA, Madison, WI.

Drijber, R.A., and L.E. Lowe. 1990. Nature of humus in Andosols under differing vegetation in the Sierra Nevada, Mexico. Geoderma 47:221–231.

Dubroeucq, D., D. Giessert, and P. Quantin. 1998. Weathering and soil forming processes under semi-arid conditions in two Mexican volcanic-ash soils. Geoderma 86:99–122.

Farmer, V.C., W.J. McHardy, F. Palmieri, A. Violante, and P. Violante. 1991. Synthetic allophanes formed in calcareous environments: Nature, conditions of formation, and transformations. Soil Sci. Soc. Am. J. 55:1162–1166.

Fieldes, M., and K.W. Perrott. 1966. The nature of allophane in soils. Part 3. A rapid field and laboratory test for allophane. N.Z. J. Sci. 9:623–629.

Gérard, M., S. Caquineau, J. Pinheiro, and G. Stoops. 2007. Weathering and allophane neoformation in soils developed on volcanic ash in the Azores. Eur. J. Soil Sci. 58:496–515.

Goldin, A., W.D. Nettleton, and R.J. Engel. 1992. Pedogenesis on outwash and glacial marine drift, northwestern Washington. Soil Sci. Soc. Am. J. 56:1545–1552.

Grange, L.I., and N.H. Taylor. 1932. The occurrence of bush sickness on the volcanic soils of the North Island. Part 2A. The distribution and field characteristics of bush-sick soils. N.Z. DSIR Bull. 32:21–35.

Harsh, J. 2000. Poorly crystalline aluminosilicate clays, p. F169–F182. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Harsh, J., J. Chorover, and E. Nizeyimana. 2002. Allophane and imogolite, p. 291–322. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA Book Series No. 7. SSSA, Madison, WI.

Henriet, C., N. De Jaeger, M. Dore, S. Opfergelt, and B. Delvaux. 2008. The reserve of weatherable primary silicates impacts the accumulation of biogenic silicon in volcanic ash soils. Biogeochemistry 90:209–223.

Hiradate, S., T. Nakadai, H. Shindo, and T. Yoneyama. 2004. Carbon source of humic substances in some Japanese volcanic ash soils determined by carbon stable isotopic ratio, $\delta^{13}C$. Geoderma 119:133–141.

Hiradate, S., and S.-I. Wada. 2005. Weathering processes of volcanic glass to allophane determined by ^{27}Al and ^{29}Si solid-state NMR. Clays Clay Miner. 53:401–408.

Hochella, M.F., Jr. 2008. Nanogeoscience: From origins to cutting-edge applications. Elements 4:373–379.

Hodder, A.P.W., B.E. Green, and D.J. Lowe. 1990. A two-stage model for the formation of clay minerals from tephra-derived volcanic glass. Clay Miner. 25:313–327.

Hodder, A.P.W., T.R. Naish, and D.J. Lowe. 1996. Towards an understanding of thermodynamic and kinetic controls on the formation of clay minerals from volcanic glass under various environmental conditions, p. 1–11. *In* S.G. Pandalai (ed.) Recent research developments in chemical geology. Research Signpost, Trivandrum, India.

Ildefonse, P., R.J. Kirkpatrick, B. Montez, G. Calas, A.M. Flank, and P. Lagarde. 1994. ^{27}Al MAS NMR and aluminum X-ray absorption near edge structure study of imogolite and allophanes. Clays Clay Miner. 42:276–287.

Inoue, Y., J. Baasansuren, M. Watanabe, H. Kame, and D.J. Lowe. 2010. Interpretation of pre-AD 472 Roman soils from physicochemical and mineralogical properties of buried tephric paleosols at Somma Vesuviana ruin, southwest Italy. Geoderma 152:243–251.

IUSS Working Group WRB. International Union of Soil Science Working Group WRB. 2006. World reference base for soil resources 2006. World Soil Resources Reports No. 103. FAO, Rome, Italy.

Jambor, J.L., and J.E. Dutrizac. 1998. Occurrence and constitution of natural and synthetic ferrihydrite, a widespread iron oxyhydroxide. Chem. Rev. 98:2549–2585.

Johnson-Maynard, J.L., P.A. McDaniel, D.E. Ferguson, and A.L. Falen. 1997. Chemical and mineralogical conversion of Andisols following invasion by bracken fern. Soil Sci. Soc. Am. J. 61:549–555.

Jongmans, A.G., F. Van Oort, P. Buurman, A.M. Jaunet, and J.D.J. van Doesburg. 1994. Morphology, chemistry and mineralogy of isotropic aluminosilicate coatings in a Guadeloupe Andisol. Soil Sci. Soc. Am. J. 58:501–507.

Jongmans, A.G., P. Verburg, A. Nieuwenhuyse, and F. van Oort. 1995. Allophane, imogolite, and gibbsite in coatings in a Costa Rican Andisol. Geoderma 64:327–342.

Joussein, E., S. Petit, G.J. Churchman, B.K.G. Theng, D. Right, and B. Devlvaux. 2005. Halloysite clay minerals—A review. Clay Miner. 40:383–426.

Kimsey, M.J., J. Moore, and P.A. McDaniel. 2008. A geographically weighted regression analysis of Douglas-fir site index in north central Idaho. Forest Sci. 54:356–366.

Kirkman, J.H., and W.J. McHardy. 1980. A comparative study of the morphology, chemical composition and weathering of rhyolitic and andesitic glass. Clay Miner. 15:165–173.

Kleber, M., C. Mikutta, and R. Jahn. 2004. Andosols in Germany—Pedogenesis and properties. Catena 56:67–83.

Kondo, R., C.W. Childs, and I. Atkinson. 1994. Opal phytoliths in New Zealand. Manaaki Whenua Press, Lincoln, New Zealand.

Lowe, D.J. 1986. Controls on the rates of weathering and clay mineral genesis in airfall tephras: A review and New Zealand case study, p. 265–330. *In* S.M. Colman and D.P. Dethier (eds.) Rates of chemical weathering of rocks and minerals. Academic Press, Orlando, FL.

Lowe, D.J. 1995. Teaching clays: From ashes to allophane, p. 19–23. *In* G.J. Churchman, R.W. Fitzpatrick, and R.A. Eggleton (eds.) Clays: Controlling the environment. CSIRO Publishing, Melbourne, Australia.

Lowe, D.J. (ed.). 2008. Guidebook for pre-conference North Island Field Trip A1 'Ashes and Issues,' November 28–30, 2008, p. 1–194. Australian and New Zealand 4th Jt Soils Conf., Massey University, Palmerston North. New Zealand Society of Soil Science Lincoln, New Zealand.

Lowe, D.J. 2011. Tephrochronology and its application: A review. Quat. Geochron. 6:107–153.

Lowe, D.J., and D.J. Palmer. 2005. Andisols of New Zealand and Australia. J. Integr. Field Sci. 2:39–65.

Lowe, D.J., and H.J. Percival. 1993. Clay mineralogy of tephras and associated paleosols and soils, and hydrothermal deposits, North Island. Guidebook for New Zealand pre-conference field trip F1, p. 1–110. 10th Int. Clay Conf., Adelaide, Australia.

Lowe, D.J., and P.J. Tonkin. 2010. Unravelling upbuilding pedogenesis in tephra and loess sequences in New Zealand using tephrochronology, p. 34–37. *In* Proc. IUSS 19th World Soil Congress, Symposium 1.3.2 Geochronological techniques and soil formation, Brisbane, Australia August 1–6, 2010. Published on DVD and online at: http://www.iuss.org

McDaniel, P.A., and K.W. Hipple. 2010. Mineralogy of loess and volcanic ash eolian mantles in Pacific Northwest (USA) landscapes. Geoderma 154:438–446.

McDaniel, P.A., M.A. Wilson, R. Burt, D. Lammers, T.D. Thorson, C.L. McGrath, and N. Peterson. 2005. Andic soils of the inland Pacific Northwest, USA: Properties and ecological significance. Soil Sci. 170:300–311.

McGlone, M., J.M. Wilmshurst, and H.M. Leach. 2005. An ecological and historical review of bracken (*Pteridium esculentum*) in New Zealand, and its cultural significance. N.Z. J. Ecol. 29:165–184.

McLeod, M., J. Aislabie, J. Ryburn, and A. McGill. 2008. Regionalizing potential for microbial bypass flow through New Zealand soils. J. Environ. Qual. 37:1959–1967.

Michaelson, G.J., and C.L. Ping. 1987. Effect of P, K and liming on soil pH, Al, Mn, K, matter yield and forage barley dry matter yield and quality for a newly cleared Cryorthod. Plant Soil 104:155–161.

Michaelson, G.J., and C.L. Ping. 1990. Mehlich 3 extractable P of Alaska soils as affected by P fertilizer application. Appl. Agric. Res. 5:255–260.

Michel, F.M., L. Ehm, S.M. Antao, P.L. Lee, P.J. Chupas, G. Liu, D.R. Strongin, M.A.A. Schoonen, B.L. Phillips, and J.B. Parise. 2007. The structure of ferrihydrite, a nanocrystalline material. Science 316:1726–1729.

Mitchell, J.K., and K. Soga. 2005. Fundamentals of soil behaviour. 3rd edn. Wiley, Hoboken, NJ.

Mizota, C., and L.P. van Reeuwijk. 1989. Clay mineralogy and chemistry of soils formed in volcanic materials in diverse climatic regions. Int. Soil Ref. Inf. Centre Monogr. 2:1–185.

Nanzyo, M. 2002. Unique properties of volcanic ash soils. Global Environ. Res. 6:99–112.

Nanzyo, M. 2007. Introduction to studies on volcanic ash soils in Japan and international collaboration. J. Integr. Field Sci. 4:71–77.

Nanzyo, M., and T. Takahashi. 2005. Changes in elemental composition with andisolization. J. Integr. Field Sci. 2:83–87.

Navarette, I.A., K. Tsutsuki, R. Kondo, and V.B. Asio. 2008. Genesis of soils across a late Quaternary volcanic landscape in the humid tropical island of Leyte, Philippines. Aust. J. Soil Res. 46:403–414.

Neall, V.E. 1977. Genesis and weathering of Andosols in Taranaki, New Zealand. Soil Sci. 123:400–408.

Neall, V.E. 2006. Volcanic soils, p. 1–24. *In* W. Verheye (ed.) Land use and land cover. Encyclopaedia of life support systems (EOLSS). EOLSS Publishers with UNESCO, Oxford, U.K. Available online at: http://www.eolss.net

Newnham, R.M., D.J. Lowe, and P.W. Williams. 1999. Quaternary environmental change in New Zealand: A review. Prog. Phys. Geogr. 23:567–610.

Parfitt, R.L. 1990. Allophane in New Zealand—A review. Aust. J. Soil Res. 28:343–360.

Parfitt, R.L. 2009. Allophane and imogolite: Role in soil biogeochemical processes. Clay Miner. 44:135–155.

Parfitt, R.L., and C.W. Childs. 1988. Estimation of forms of Fe and Al: A review, and analysis of contrasting soils by dissolution and Moessbauer methods. Aust. J. Soil Res. 26:121–144.

Parfitt, R.L., and B. Clayden. 1991. Andisols—The development of a new order in Soil Taxonomy. Geoderma 49:181–198.

Parfitt, R.L., and T. Henmi. 1982. Comparison of an oxalate extraction method and an infrared spectroscopic method for determining allophane in soil clays. Soil Sci. Plant Nutr. 28:183–190.

Parfitt, R.L., M. Russell, and G.E. Orbell. 1983. Weathering sequence of soils from volcanic ash involving allophane and halloysite, New Zealand. Geoderma 29:41–57.

Parfitt, R.L., and M. Saigusa. 1985. Allophane and humus–aluminum in Spodosols and Andepts formed in the same volcanic ash beds in New Zealand. Soil Sci. 139:149–155.

Parfitt, R.L., M. Saigusa, and J.D. Cowie. 1984. Allophane and halloysite formation in a volcanic ash bed under differing moisture conditions. Soil Sci. 138:360–364.

Pewe, T.L. 1975. Quaternary geology of Alaska. U.S. Geological Survey Professional Paper 835.

Ping, C.-L. 2000. Volcanic soils, p. 1259–1270. *In* H. Sigurdsson et al. (eds.) Encyclopedia of volcanoes. Academic Press, San Diego, CA.

Ping, C.-L., S. Shoji, and T. Ito. 1988. The properties and classification of three volcanic ash derived pedons from Aleutian Islands and Alaska Peninsula. Soil Sci. Soc. Am. J. 52:455–462.

Ping, C.-L., S. Shoji, T. Ito, T. Takahashi, and J.P. Moore. 1989. Characteristics and classification of volcanic-ash-derived soils in Alaska. Soil Sci. 148:8–28.

Prado, B., C. Duwig, C. Hildago, D. Gómez, H. Yee, C. Prat, M. Esteves, and J.D. Etchevers. 2007. Characterization, functioning and classification of two volcanic soil profiles under different land uses in central Mexico. Geoderma 139:300–313.

Qafoku, N.P., E. Van Ranst, A. Noble, and G. Baert. 2004. Variable charge soils: Their mineralogy, chemistry and management. Adv. Agron. 84:159–215.

Quantin, P. 2004. Volcanic soils of France. Catena 56:95–109.

Reynolds-Vargas, J., D.D. Richter, and E. Bornemisza. 1994. Environmental impacts of nitrification and nitrate adsorption in fertilized andisols in the Valle Central, Costa Rica. Soil Sci. 157:289–299.

Rieger, S., D.B. Schoepherster, and C.E. Furbush. 1979. Exploratory soil survey of Alaska. USDA-SRC. U.S. Government Printing Office, Washington, DC.

Ross, C.W., M.S. Watt, R.L. Parfitt, R. Simcock, J. Dando, G. Coker, P.W. Clinton, and M.R. Davis. 2009. Soil quality relationships with tree growth in exotic forests in New Zealand. For. Ecol. Manage. 258:2326–2334.

Ryan, M.C., G.R. Graham, and D.L. Rudolph. 2001. Contrasting nitrate adsorption in Andisols of two coffee plantations. J. Environ. Qual. 30:1848–1852.

Saigusa, M., and N. Matsuyama. 2004. Distribution and agronomic aspects of non-allophanic andisols. Soil Sci. Plant Nutr. 50:334–335.

Schlichting, E. 1988. Physical properties of Andisols conducive to treading damage. *In* D.I. Kinloch, S. Shoji, F.H. Beinroth, and H. Eswaran (eds.) Proc. 9th lnt. Soil Classif. Workshop USDA-SMSS, Washington, DC.

Schwertmann, U. 2008. Iron oxides, p. 363–369. *In* W. Chesworth (ed.) Encyclopaedia of soil science. Springer, Dordrecht, The Netherlands.

Schwertmann, U., and R.M. Taylor. 1989. Iron oxides, p. 379–438. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA Book Series No. 1. SSSA, Madison, WI.

Shikazono, N., A. Takino, and H. Ohtani. 2005. An estimate of dissolution rate constant of volcanic glass in volcanic ash soil from the Mt. Fuji area, central Japan. Geochem. J. 39:185–196.

Shoji, S., T. Ito, M. Saigusa, and I. Yamada. 1985. Properties of nonallophanic Andosols from Japan. Soil Sci. 140:264–277.

Shoji, S., M. Nanzyo, and R.A. Dahigren. 1993. Volcanic ash soils: Genesis, properties and utilization. Dev. Soil Sci. 21:1–288.

Shoji, S., M. Nanzyo, and T. Takahashi. 2006. Factors of soil formation: Climate. As exemplified by volcanic ash soils, p. 131–149. *In* G. Certini and R. Scalenghe (eds.) Soils: Basic concepts and future challenges. Cambridge University Press, Cambridge, U.K.

Shoji, S., and T. Takahashi. 2002. Environmental and agricultural significance of volcanic ash soils. Global Environ. Res. 6:113–135.

Sigurgeirsson, M.A., O. Arnalds, S.E. Palsson, B.H. Howard, and K. Gudnason. 2005. Radiocaesium fallout behaviour in volcanic soils in Iceland. J. Environ. Radioact. 79:39–53.

Simonson, R.W. 1979. Origin of the name "Ando soils." Geoderma 22:333–335.

Singleton, P.L., M. McLeod, and H.J. Percival. 1989. Allophane and halloysite content and soil solution silicon in soils from rhyolitic volcanic material, New Zealand. Aust. J. Soil Res. 27:67–77.

Smith, G.D. 1978. The Andisol proposal—A preliminary proposal for reclassification of the Andepts and some Andic subgroups, p. 1–165. *In* M.L. Leamy, D.I. Kinloch, and R.L. Parfitt (eds.) International Committee on Andisols—Final report. USDA Soil Management Support Services Technical Monograph No. 20.

Smith, R.T., D.J. Lowe, and I.C. Wright. 2006. Volcanoes. Te Ara—The encyclopedia of New Zealand. Updated April 16, 2007. New Zealand Ministry for Culture and Heritage, Wellington, New Zealand. Available online at: http://www.TeAra.govt.nz/EarthSeaAndSky/NaturalHazardsAndDisasters/Volcanoes/en

Soil Survey Staff. 1975. Soil Taxonomy—A basic system of soil classification for making and interpreting soil surveys. USDA-SCS agricultural handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1990. Keys to Soil Taxonomy. 4th edn. SMSS Technical Monograph No. 6. Blacksburg, VA.

Soil Survey Staff. 1999. Soil Taxonomy—A basic system of soil classification for making and interpreting soil surveys. 2nd edn. USDA-NRCS agricultural handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2011. National Cooperative Soil characterization Database. Available online at: http://ssldata.nrcs.usda.gov. Accessed July 8, 2011.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS, Washington, DC.

Southard, S.B., and R.J. Southard. 1991. Mineralogy and classification of Andic soils in northeastern California. Soil Sci. Soc. Am. J. 53:1784–1791.

Stevens, K.F., and C.G. Vucetich. 1985. Weathering of upper Quaternary tephras in New Zealand. Part 2. Clay minerals and their climatic interpretation. Chem. Geol. 53:237–247.

Stoops, G. 2007. Micromorphology of soils derived from volcanic ash in Europe: A review and synthesis. Eur. J. Soil Sci. 58:356–377.

Sumner, M.E., and K. Hylton. 1993. A diagnostic approach to solving soil fertility problems in the tropics, p. 215–234. *In* J.K. Syers and D.L. Rimmer (eds.) Soil science and sustainable land management in the tropics. CAB International, Wallingford, U.K.

Takahashi, T., R. Dahlgren, and P. van Suteren. 1993. Clay mineralogy and chemistry of soils formed in volcanic materials in the xeric moisture regime of northern California. Geoderma 59:131–150.

Takahashi, T., Y. Ikeda, K. Fujita, and M. Nanzyo. 2006. Effect of liming on organically complexed aluminum of nonallophanic Andosols from northeastern Japan. Geoderma 130:26–34.

Takahashi, T., and S. Shoji. 2002. Distribution and classification of volcanic ash soils. Global Environ. Res. 6:83–97.

Takesako, H., D.J. Lowe, G.J. Churchman, and D. Chittleborough. 2010. Holocene volcanic soils in the Mt Gambier region, South Australia, p. 47–50. *In* Proc. IUSS 19th World Soil Congress, Symposium 1.3.1 Pedogenesis: Ratio and range of influence, Brisbane, August 1–6, 2010. Published on DVD and online at: http://www.iuss.org

Takesako, H., and S. Muranaka. 2006. Palaeoenvironmental change in central Japan: A ca. 33-ka record based on physico-chemical analysis of a multilayered sequence of tephra deposits and andic paleosols, p. 81–92. *In* H. Takesako (ed.) Proc. Int. Symp. Volcanic-Ash Soils and Field Workshop in the Mt. Fuji Area. Meiji University (Ikata), Kawasaki, Japan.

Theng, B.K.G., and G. Yuan. 2008. Nanoparticles in the soil environment. Elements 4:395–399.

Torrance, J.K. 1992. Discussion on "Sensitivity to remoulding of some volcanic ash soils in New Zealand," by D. Jacquet. Eng. Geol. 32:101–105.

Ugolini, F.C., and R.A. Dahlgren. 1991. Weathering environments and occurrence of imogolite/allophane in selected Andisols and Spodosols. Soil Sci. Soc. Am. J. 55:1166–1171.

Ugolini, F.C., and R.A. Dahlgren. 2002. Soil development in volcanic ash. Global Environ. Res. 6:69–81.

Vacca, A., P. Adamo, M. Pigna, and P. Violante. 2003. Genesis of tephra-derived soils from the Roccamonfina volcano, south central Italy. Soil Sci. Soc. Am. J. 67:198–207.

Van Ranst, E., S.R. Utami, and J. Shamshuddin. 2002. Andisols on volcanic ash from Java Island, Indonesia: Physico-chemical properties and classification. Soil Sci. 167:68–79.

Wada, K. (ed.). 1986. Ando soils in Japan. Kyushu University Press, Fukuoka, Japan.

Wada, K. 1989. Allophane and imogolite, p. 1051–1087. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA Book Series No. 1. SSSA, Madison, WI.

Warkentin, B.P., and T. Maeda. 1980. Physical and mechanical characteristics of Andisols. *In* B.K.G. Theng (ed.) Soils with variable charge. New Zealand Society of Soil Science, Lower Hutt, New Zealand.

Waychunas, G.A., and H. Zhang. 2008. Structure, chemistry, and properties of mineral nanoparticles. Elements 4:381–387.

White, G.N., and J.B. Dixon. 2002. Kaolin–serpentine minerals, p. 389–414. *In* J.B. Dixon and D.G. Schulze (eds.) Soil mineralogy with environmental applications. SSSA Book Series No. 7. SSSA, Madison, WI.

Wilding, L.P. 2000. Classification of soil, p. E175–E183. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Wolff-Boenisch, D., S.R. Gislason, E.H. Oelkers, and C.V. Putnis. 2004. The dissolution rates of natural glasses as a function of their composition at pH 4 and 10.6, and temperatures from 25 to 74°C. Geochim. Cosmochim. Acta 68:4843–4858.

33.4 Entisols

L.C. Nordt

M.E. Collins

H.C. Monger

D.S. Fanning

33.4.1 Introduction

Entisols are mineral soils that commonly exhibit minimal pedogenic alteration (Soil Survey Staff, 1999, 2010). This taxonomic order was established to preserve the early Russian "azonal" concept for soils with undeveloped profiles crossing bioclimatic zones (Smith, 1986). Typically described as A-C, A-Cr, or A-R profiles, Entisols retain many physical and chemical properties of the associated parent materials. Most A horizons classify as ochric epipedons, but with Ap horizons classifying as anthropic in some human-disturbed lands and as agric in some agricultural lands. Occasionally, Entisols have histic epipedons if affiliated with sulfidic materials in coastal regions with persistently high water tables. Highly disturbed soils from human activity may have fragments of former diagnostic horizons classified as Entisols.

Subsurface B horizons may be described to indicate a pathway of pedogenesis (e.g., Bw), but the properties must not meet the minimum requirements of any subsurface diagnostic horizon (e.g., cambic). Exceptions are the recognition of a weathered albic horizon as long as genetically related illuvial horizons (e.g., argillic and spodic) do not occur within 2 m of the surface, and cryoturbation in Gelisols that prevents diagnostic horizons from forming. Despite minimal horizonation, however, Entisols must develop physical and chemical properties capable of supporting plants. This characteristic distinguishes Entisols from nonsoil areas such as rock outcrop, water, ice, or badlands.

33.4.2 Distribution

Entisols most often occur where the influence of the soil-forming factors retards pedogenesis. Landforms where Entisols form are marshes and estuaries, modified urban areas, sand dunes, floodplains, and arid regions with steep or rocky slopes (Figure 33.22). These landforms are associated with wetlands, inert sandy parent materials, rapid rates of aggradation or erosion, and intensive human activity (Buol et al., 1997). Because of the geographic extent of these environments, Entisols are the most abundant of all soil orders covering approximately 16% of the world's ice-free land mass (Soil Survey Staff, 1998). Up to 60% of Entisols occur in temperate and tropical regions, and many mapped in circumarctic regions have been replaced by the Gelisol order.

Entisols are replete in buried contexts throughout the Phanerozoic (see Birkeland, 1999). Recognition of Entisols is important for understanding Quaternary landscape evolution and preservation potentials of the buried archaeological record (Holliday, 2004). Paleo-Entisols in the pre-Quaternary rock record serve as sequence stratigraphic markers (McCarthy and Plint, 1998; Atchley et al., 2004), enhance interpretations of pedofacies relations (Kraus, 2002), provide inferences regarding colonizing ecosystems (Retallack, 1994), and serve as repositories of paleobotanical and paleontological information (see Retallack, 2001).

33.4.3 Formation

Ripening and melanization are the principle soil-forming processes that initiate the development of surface diagnostic horizons and Entisols (Buol et al., 1997). Grossman (1983) reports data showing that bulk densities in alluvial settings initially increase because of frequent wetting and drying cycles, which leads to compaction. This process is similar to ripening as freshly deposited sediments from running water become dehydrated, but is followed by ped formation and increased porosity (Pons and van der Molen, 1973). Melanization contributes organic materials to the parent material by way of plant and animal activity promoting structural development that is accompanied by increased porosity and minor release
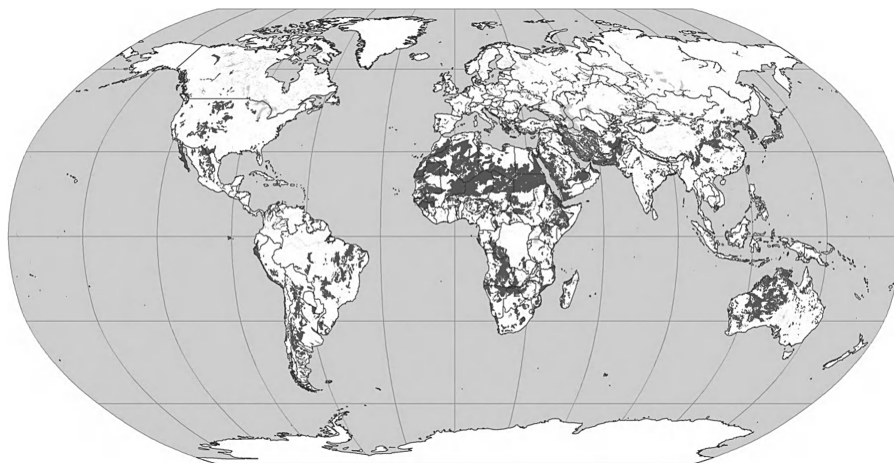
FIGURE 33.22 Global distribution of Entisols. (Courtesy of USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC, 2010.)
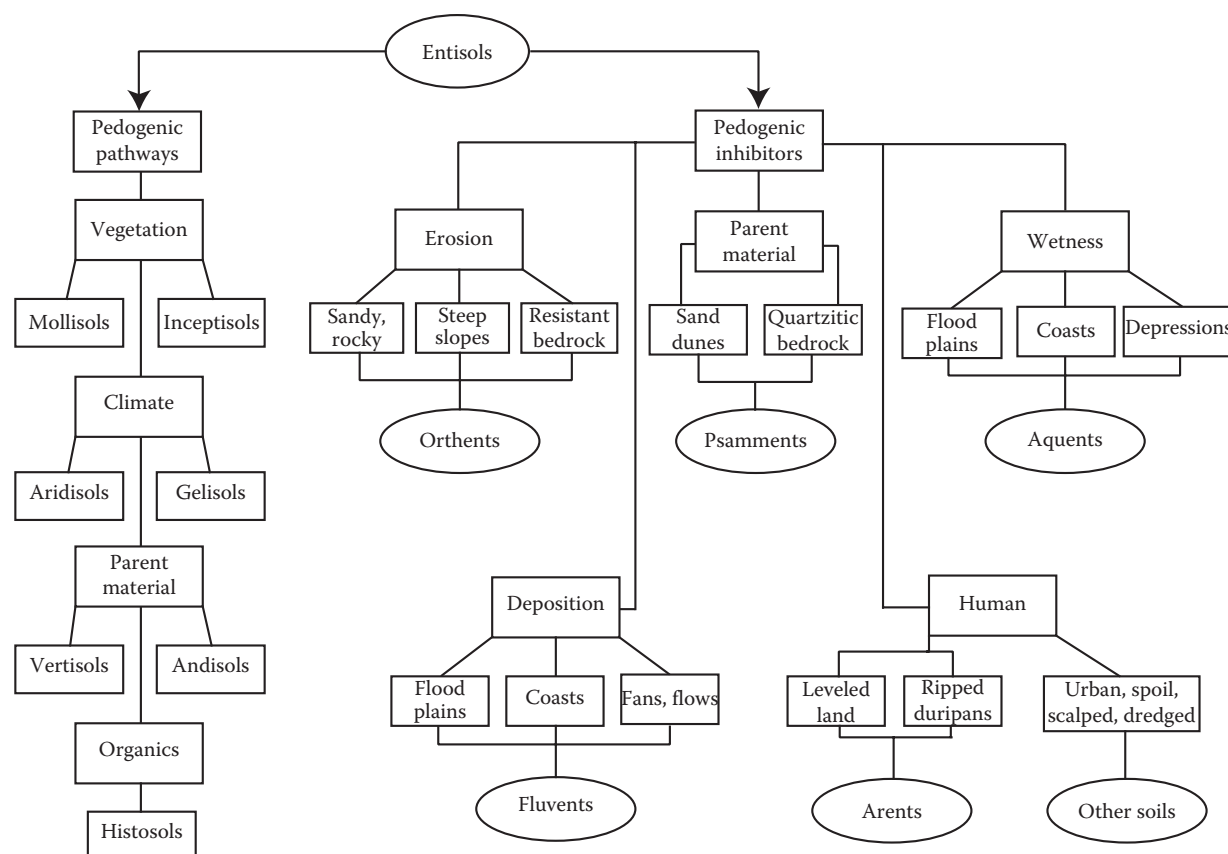
FIGURE 33.23 Flow diagram illustrating pedogenic pathways of Entisols to other soil orders and pedogenic inhibitors conditioning the formation of Entisol suborders.

and leaching of soluble constituents. In resistant bedrock, this process increases porosity as rock material is weathered and replaced by organic material. The processes of ripening and melanization in the formation of Entisols can proceed as long as mollic or umbric epipedons or subsurface diagnostic horizons do not form.

Entisols serve as the precursor to all other soil orders during their evolutionary history (Figure 33.23) (Soil Survey Staff, 1999, 2010). In grasslands, Mollisols with mollic epipedons form from Entisols when steady state has been attained in the accumulation of organic matter. In forest settings, ochric epipedons and Inceptisols may form because of relatively low, subsurface biomass production. Entisols develop into Aridisols outside of floodplains in areas with aridic soil moisture regimes. Vertisols form from Entisols in shrink-swell parent materials where wet–dry cycles are prevalent, and Andisols form from Entisols in parent materials with large quantities of volcanic ash in cool and humid climates. Histosols may form from Entisols with the accumulation of significant amounts of organic matter in association with high water tables and anaerobic conditions.

33.4.4 Suborders

In contrast to most other soil orders, suborders of Entisols are not defined by precipitation or temperature regime. Rather, Entisol suborders are designed primarily to differentiate conditions

that impede subsurface horizon development (Smith, 1986; Soil Survey Staff, 1999, 2010). The five Entisol suborders are Aquents, Arents, Psamments, Fluvents, and Orthents (Figures 33.23 and 33.24). The great group and subgroup lower taxonomic levels recognize extragrades and intergrades to other soil moisture regimes and diagnostic conditions (Table 33.18).

33.4.4.1 Aquents

Aquents are wet Entisols. They exist widely in the United States, except in Arizona and New Mexico, with the largest concentrations in southern Florida (Soil Survey Staff, 1990). Worldwide, Aquents are not extensive, covering approximately 0.1% of the Earth's land. Areas where Aquents occur include salt marshes of coastal tidal areas, freshwater marshes and swamps, alluvial flats and backswamps of floodplains, and outwash plains that receive new deposits of sediment at frequent intervals. Depending on where Aquents are located, some can be used for agricultural production when drained. Many undrained areas of Aquents are used for pasture or wildlife habitat.

Aquents are defined as having aquic conditions and sulfidic materials within 50 cm of the mineral soil surface, or permanent saturation and a reduced matrix below a depth of 25 cm from the mineral soil surface, or texture and color requirements if a densic, lithic, or a paralithic contact, or the presence of aquic conditions for some time in some years in a layer between 40 and 50 cm from the
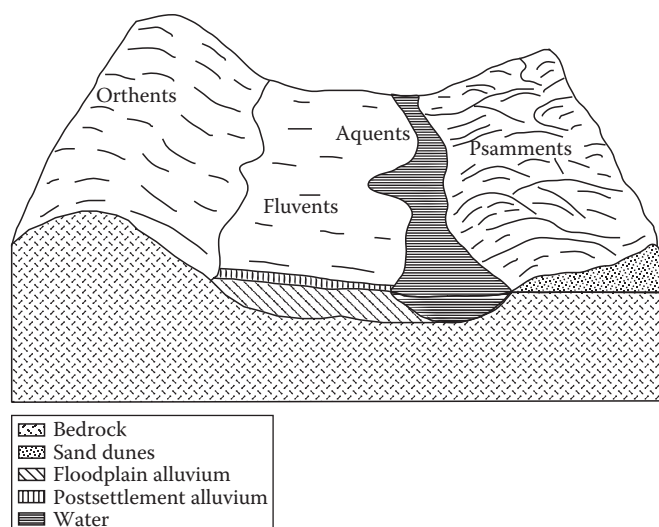
FIGURE 33.24 Block diagram illustrating typical landscape positions where Entisol suborders form: Orthents on steep slopes associated with bedrock, Fluvents in floodplains that include postsettlement alluvium, Aquents in poorly drained riparian conditions, and Psamments in sand dunes. Arents not shown.

mineral soil surface (Soil Survey Staff, 1998). Aquents can range in texture from sandy to clayey. One common feature of all Aquents is their saturation with water at or near the surface for much of the year, which presents many engineering and land use challenges.

Wetness in Aquents is normally inferred by the color distribution generated by the redox state of Fe. In most Aquents, this process produces redoximorphic features (Fe concentrations or depletions) within the upper 50 cm of the surface. Redoximorphic features are most common in nonsandy Aquents. The sandy Aquents (Psammaquents) often lack redoximorphic features because of the very low Fe content of the quartz-rich parent materials. Many Aquents have hydric soil indicators (thin layer of muck on the soil surface, strong sulfidic odor, or a gray and reduced soil matrix in the upper 15 cm of the soil; USDA-NRCS, 1996). Florida law and county listings elsewhere show that some Aquents (Sulfaquents, Hydraquents, and Psammaquents) are considered as hydric soils (Florida Statutes, 1995).

33.4.4.1.1 Lower Taxonomic Levels

Aquents are associated with landscapes that accumulate soil water. They exist in cryic to isohyperthermic soil temperature regimes. Sulfaquents and Hydraquents are located along coastal areas in

TABLE 33.18 Listing of Suborder, Great Groups, and Subgroups in the Entisol Order

Suborder	Great Group	Subgroup
Aquents	Sulfaquents	Haplic, Histic, Thapto-Histic, Typic
	Hydraquents	Sulfic, Sodic, Thaptic-Histic, Typic
	Cryaquents	Aquandic, Typic
	Psammaquents	Lithic, Sodic, Spodic, Humaqueptic, Mollic, Typic
	Fluvaquents	Sulfic, Vertic, Thapto-Histic, Aquandic, Aerie, Humaqueptic, Rhodic, Mollic, Typic
	Epiaquents	Aerie, Humaqueptic, Mollic, Typic
	Endoaquents	Sulfic, Lithic, Sodic, Aerie, Humaqueptic, Mollic, Typic
Arents	Ustarents	Haplic
	Xerarents	Sodic, Duric, Alfic, Haplic
	Torriarents	Sodic, Duric, Haplic
	Udarents	Alfic, Ultic, Mollic, Haplic
Psamments	Cryopsamments	Lithic, Aquic, Oxyaquic, Vitrandic, Spodic, Lamellic, Typic
	Torripsamments	Lithic, Vitrandic, Haploduridic, Ustic, Xeric, Rhodic, Typic
	Quartzipsamments	Lithic, Aquodic, Aquic, Oxyaquic, Ustoxic, Udoxic, Plinthic, Lamellic Ustic, Lamellic, Ustic, Xeric, Spodic, Typic
	Ustipsamments	Lithic, Aquic, Oxyaquic, Aridic, Lamellic, Rhodic, Typic
	Xeropsamments	Lithic, Aquic Durinodic, Aquic, Oxyaquic, Vitrandic, Durinodic, Lamellic, Dystric, Typic
	Udipsamments	Lithic, Aquic, Oxyaquic, Spodic, Lamellic, Plaggeptic, Typic
Fluvents	Cryofluvents	Andic, Vitrandic, Aquic, Oxyaquic, Mollic, Typic
	Xerofluvents	Vertic, Aquandic, Andic, Vitrandic, Aquic, Oxyaquic, Durinodic, Mollic, Typic
	Ustifluvents	Aquertic, Torrertic, Vertic, Anthraquic, Aquic, Oxyaquic, Aridic, Udic, Mollic, Typic
	Torrifluvents	Ustertic, Vertic, Vitrixerandic, Vitrandic, Aquic, Oxyaquic, Duric Xeric, Duric, Ustic, X, Anthropic, Typic
	Udifluvents	Aquertic, Vertic, Andic, Vitrandic, Aquic, Oxyaquic, Mollic, Typic
Orthents	Cryorthents	Lithic, Vitrandic, Aquic, Oxyaquic, Lamellic, Typic
	Torriorthents	Lithic Ustic, Lithic Xeric, Lithic, Xerertic, Ustertic, Vertic, Vitrandic, Aquic, Oxyaquic, Duric, Ustic, Xeric, Typic
	Xerorthents	Lithic, Vitrandic, Aquic, Oxyaquic, Durinodic, Dystric, Typic
	Ustorthents	Aridic Lithic, Lithic, Torrertic, Vertic, Anthraquic, Aquic, Oxyaquic, Durinodic, Vitritorrandic, Vitrandic, Aridic, Udic, Vermic, Typic
	Udorthents	Lithic, Vitrandic, Aquic, Oxyaquic, Vermic, Typic

Source: Soil Survey Staff. 1998. Global soil regions. USDA-NRCS, Lincoln, NE.

tidal marshes. Sulfaquents, as the name implies, have sulfidic materials. Hydraquents are thixotropic soils, meaning that they cannot support appreciable overburden pressures (n value >0.7). Psammaquents have a texture of loamy fine sand or coarser in all layers within the control section for the family particle size class. Fluvaquents occur on floodplains and are defined by the amount of organic C and its distribution in the soil. Cryaquents form in the colder climatic areas of Alaska and other higher latitudes that receive significant snowfalls, and soil development is slow.

33.4.4.1.2 Genesis

In association with wet conditions, Aquents form where pedogenetic development does not occur or is very slow. The soil must be saturated for long periods, and in some areas, organic matter can accumulate to create a histic epipedon. Saturation in Aquents can be episaturated (perched water table) or endosaturated (apparent water table). The time needed for Aquents to develop can be short. On floodplains and other areas experiencing rapid deposition, Aquents commonly have stratified layers of sand, silt, clay, and organic matter. Aquents forming on floodplains may also have buried histic epipedons. Aquents with high clay content may have vertic properties.

Aquents from Florida (Pellicer series: fine, smectitic, nonacid, hyperthermic Typic Sulfaquents), Egypt (Edko series: fine-loamy, mixed, nonacid, hyperthermic Mollic Fluvaquents), and Alaska (Beluga series: coarse-loamy, mixed, superactive, nonacid, Typic Cryaquents) are selected to represent Aquents from different geographic areas and to reflect different genetic pathways. The Pellicer soils form in clayey marine sediments in tidal marshes along the Atlantic coastal areas of Florida (Carlisle et al., 1981). These soils are very poorly drained and subject to daily tidal flooding. The representative Aquent from Egypt has anthric saturation and redoximorphic features in the form of brownish Fe accumulations. The soil was sampled in a depressional area south of Lake Burullus near Resheed City in northern Egypt (Rasheed et al., 1992). The Edko soil developed in recent fluviolacustrine deposits in an aridic climate. The Beluga soils are extensive on floodplains and alluvial fans in southern Alaska and form in stratified alluvium and colluvium derived from soft clayey shale and sandstone. These soils are poorly drained and saturated with water at the surface for portions of the year.

33.4.4.1.3 Physical and Chemical Properties

Physical properties such as particle size distribution and chemical properties such as organic C content, pH, extractable bases, and CEC_7 are presented for the Pellicer and Edko soils (Table 33.19). The sand, silt, and clay content of the Pellicer varies considerably within the profile, reflecting the different conditions under which the materials were deposited. Organic C content of this soil is high in the surface horizon as a result of the accumulation of organic materials under wet conditions. The Pellicer soils are high in S content (1%–4%), and this is reflected in the extremely low pH values (<3.5).

The sand, silt, and clay content of the Edko soil is fairly evenly distributed through the profile. In contrast to the Pellicer soils, the pH of the Edko soil is high (≥8.4) due to the high Na content.

33.4.4.1.4 Subaqueous Soils

An amendment to *Soil Taxonomy* to better classify subaqueous soils has been developed by Dr. Mark Stolt, Professor of Pedology at the University of Rhode Island (Soil Survey Staff, 2010). The formative element Wass is derived from the German (Swiss) word "wasser" for water and is used for subaqueous Histosols and Entisols. Six great groups within Wassents are proposed keying out in the order: Frasiwassents, Psammowassents, Sulfiwassents, Hydrowassents, Fluviwassents, and Haplowassents. The depth range of the water column where these soils occur is not known but a depth of 2.5 m below the surface has been set for soil survey inventory. Excluded are soils in areas where the water column is too thick (>2.5 m) for the growth of rooted plants. Subaqueous soils occur in protected estuarine coves, bays, inlets, and lagoons.

Only recently have pedologists considered these substrates to be soil material. Consequently, very few subaqueous soils have been characterized, and relationships between subaqueous soils and associated subtidal landforms have not been extensively studied. Recent subaqueous soil projects have been undertaken or are in progress along the Gulf coast in Florida and Texas and the Atlantic Coast in Connecticut, Delaware, Florida, Maine, Maryland, Massachusetts, New Hampshire, and Rhode Island. Work along the southeast Atlantic Coast of Florida (Fischler, 2006) has included mapping and classification of subaqueous soils near Cedar Key, Florida (Ellis, 2006), quantifying human activities that either destroy or create an environment in which it is difficult to reestablish the fragile seagrass vegetation, and studying the biogeochemical characteristics of subaqueous soils as related to aquatic vegetation in a river flowing into the Gulf of Mexico (Saunders, 2007).

33.4.4.2 Arents

Arents are former soils physically disturbed by the actions of humans (Soil Survey Staff, 1999, 2010). Arents are restricted to those soils "which have, in one or more layers between 25 and 100 cm from the mineral soil surface, 3% or more by volume fragments of diagnostic horizons that are not arranged in any discernible order (Soil Survey Staff, 1999, 2010)." Two examples of the transformation of preexisting soils into Arents follow.

Duripans were ripped by heavy equipment to improve the soils for the growth of almond nut trees in California (R.W. Simonson, personal communication, based on observations in the 1960s). According to Dr. Simonson, the ripping was possible only with moderately developed duripans, whereas thick, strongly developed duripans were impossible to rip. The value of the land was greatly improved by the ripping, such that successful commercial almond production became possible. As a result of the duripan ripping, the great group classification of the soils changed from Durixeralfs to Xerarents.

Another example of Arent soil formation occurred in western Tennessee where formerly gullied land was leveled with heavy equipment for agricultural purposes. Fragments of argillic horizons from the original Alfisols permitted the new soils to be identified as Arents after the land disturbance. Thus, Hapludalfs (i.e., Memphis soil series) were reclassified to Udarents following the leveling.

TABLE 33.19 Soil Characterization Data for the Pellicer Clay (Carlisle et al., 1981) and Edko Soil (Rasheed et al., 1992) of the Aquent Suborder

Horizon	Depth (cm)	Very Coarse Sand	Coarse Sand	Medium Sand	Fine Sand	Very Fine Sand	Silt	Clay	pH (H$_2$O)	OC[a] (%)	Ca[b]	Mg	Na	K	CEC$_7$	Base Saturation (%)
					(%)							(cmol$_c$ kg^{-1})				
Pellicer Clay (Typic Sulfaquents)																
A	0–51	0.1	0.5	2.1	3.5	9.3	23.6	60.9	2.7	14.3	11.7	24.3	27.7	1.0	129.6	50
C1	51–89	0.1	1.1	15.9	21.6	11.3	16.6	33.4	3.2	7.1	6.8	15.7	20.7	1.1	78.6	56
C2	89–140	0.1	2.6	38.7	34.1	10.9	7.5	6.1	3.1	0.8	1.3	3.3	5.0	0.2	15.4	63
Edko Soil (Mollic Fluvaquents)																
A1	0–6	—	4.6	—	37.4	—	26.4	31.6	8.6	3.3	12.0	16.0	64.6	1.1	—	—
A2	6–23	—	4.0	—	29.1	—	36.0	30.9	8.4	1.6	14.0	24.0	67.2	0.9	—	—
C	23–56	—	0.4	—	29.9	—	34.8	29.9	8.6	0.9	10.0	14.0	65.2	4.5	—	—
2 C	56–76	—	0.4	—	29.9	—	34.8	29.9	8.6	0.9	10.0	14.0	65.2	4.5	—	—

Source: Soil Survey Staff. 2011. National Cooperative Soil Characterization Database. Available online at http://ssldata.nrcs.usda.gov (Accessed July 8, 2011.)

[a] OC is organic carbon.

[b] Ammonium acetate extractable Ca, Mg, Na, K, and CEC (cation exchange capacity), buffered at pH 7.

33.4.4.2.1 Soils in Human-Deposited or Human-Exposed Parent Materials

Humans affect soils in many ways. These soils form in mine spoil, dredge spoil, and other recently deposited urban and industrially related materials, as well as in bomb craters in war zones (Hupy and Schaetzl, 2006). Some Entisols have been brought into existence where humans have deposited or exposed soil parent material by earth-moving operations. Some of these operations include highway construction, surface mining, landfill construction, or dredging.

In many soils with human-deposited or human-exposed parent materials, fragments of diagnostic horizons are not identifiable; thus, a classification of Arents would not be appropriate. Such soils commonly have been classified in the suborder of Orthents and subdivided into great groups according to the soil moisture regime (i.e., into Udorthents in the eastern United States; Soil Survey Staff, 1999, 2010). This classification is insufficient for many soil scientists because it does not reflect the role of humans in the genesis of the soils (Short et al., 1986a; Strain and Evans, 1994). Although these soils are commonly thought to be highly variable, studies have shown degrees of variability similar to that of some soils in naturally deposited parent materials (Short et al., 1986b).

Many soils forming in materials deposited by bulldozers or other earth-moving equipment have an irregular distribution of organic matter with depth. In theory, these soils would classify as Fluvents at the suborder taxonomic level (Soil Survey Staff, 1999, 2010). In contrast to Fluvents, however, human-disturbed soils are not stratified by natural processes, are not flooded in most years, and have a higher density. Despite the irregular organic matter depth distribution, these soils behave more like Orthents than Fluvents and have been forced, in some soil surveys (Smith, 1976), into Orthents by ignoring the irregular organic matter depth distribution.

Because of the inadequacies of the present classification system, new suborders and/or subgroups of Entisols have been proposed for soils in human-deposited or human-exposed parent materials (Fanning and Fanning, 1989; Chapters 24, 28, and Appendix 3). An international committee for anthropogenic soils (ICOMANTH) has shown considerable interest in adding new taxonomic classes. The committee, however, has not determined at which taxonomic level human-influenced soils should be recognized. For example, Kosse (1988) and Strain and Evans (1994) have proposed classifying these soils at the order level as Anthosols. Regardless, the Arent suborder discussed above established a precedence for recognition at the suborder level. Fanning and Fanning (1989) have supported this viewpoint.

West Virginia workers (Sencindiver, 1977; Smith and Sobek, 1978) proposed the suborder Spolents (for spoil) for soils formed in materials deposited from surface mining operations. During development of the soil survey of Washington, DC, Smith (1976) proposed a suborder of Urbents for soils formed in miscellaneous urban fill that had artifacts such as brick or concrete in the particle size control section (generally between a depth of 25 and 100 cm). The Garbents suborder was proposed for soils formed in garbage of an organic nature within a depth of 2 m from the soil surface, such that the soils would readily subside

and generate methane under anaerobic conditions (Fanning and Fanning, 1989; Fanning, 1991). Fanning and Fanning (1989, Chapter 24) proposed special diagnostic soil characteristics (i.e., urbic, garbic, spolic, and dredged material; and the scalped land surface) for taxonomic classification. These characteristics were employed to propose special subgroup classes, such as Urbic Garbic Udorthents for "sanitary landfills where the cover material qualified as urbic materials" (Fanning and Fanning, 1989, Appendix 3). It was proposed that soils on scalped land surfaces (cuts from an engineering perspective) could be placed in Scalpic subgroups.

Soils created by dredging operations commonly qualify as Sulfaquents immediately after sulfidic dredged materials have been deposited (Fanning and Fanning, 1989). Under aerobic conditions at dredged material sites, sulfuric horizons commonly form from sulfidic materials in a few weeks or months. This causes the soils to become Sulfaquepts, a great group of Inceptisols. Thus, the period that such soils remain as Entisols, before becoming a soil of a different order, may be very brief.

With the present strong interest in improving knowledge about the classification of human-influenced soils, much progress may be expected in the near future. This is needed as soil surveys of lands with human-influenced soils are becoming more widespread. In ongoing urban surveys, many of the problems associated with the soils discussed in this section will be encountered. For example, sulfide-bearing clays have been used as landfill cover materials, which can lead to the formation of acid sulfate soils on landfills (Kargbo et al., 1993; Fanning and Burch, 1997). A more resolved classification system can enhance future interpretations and mapping techniques of soil surveys in these areas. Further, it has been suggested (Fanning, 1990) that a discipline of pedotechnology is needed to improve the design of soils brought into existence through the action of humans.

33.4.4.3 Psamments

Psamments are Entisols with sandy parent materials. They typically show little or no evidence of soil development with horizon sequences of AC or A/AC-C. An exception, however, are areas (e.g., central Florida ridge) where Psamments represent intensely weathered albic horizons with argillic and spodic horizons forming at depths greater than 2 m below the surface. Consequently, Psamments generally have an ochric epipedon, but lack any diagnostic subsurface horizons within the upper 2 m depth.

Psamments cover 3.4% of the Earth's land. Major desert regions, such as in northern Africa (Sahara), Saudi Arabia, western Australia, and southern Mongolia (Gobi), have Psamments. In the United States, these soils exist in all states. They are extensive in the Central Ridge of Florida, in the Sandhills in western Nebraska, in the sandsheet in south Texas, in central Wisconsin, and throughout the arid and semiarid areas of the western United States on the leeward sides of pluvial lakes.

Generally, Psamments have low amounts of plant nutrients, low water-holding capacity, and rapid permeability. Irrigation is normally necessary to maintain economical crop production. With management, however, they can be used for citrus

production, for growing sand pines, and even for real estate development. In arid and semiarid regions, Psamments are ecologically fragile, and disturbance to rangeland vegetation growing on them can initiate feedback processes leading to severe desertification (Monger and Bestelmeyer, 2006).

Natural vegetation associated with Psamments in Florida consists of bluejack oak (*Quercus incana*), blackjack oak (*Quercus marilandica*), turkey oak (*Quercus laevis*), longleaf threeawn (*Aristida stricta*), and bluestem (*Andropogon* sp.). In the major cattle-growing states of Texas and Nebraska, Psamments are used for spring and summer grazing of native grasses. In central Wisconsin, truck crops are an important use.

33.4.4.3.1 Lower Taxonomic Levels

Psamments are restricted to soils that have a texture of loamy fine sand or coarser in the control section (family particle size class) and <35% by volume of particles >2 mm diameter (Soil Survey Staff, 1999, 2010). The latter distinguishes Psamments from sandy Orthents. Sandy loam lamellae are permitted in Psamments, as long as cumulatively they do not meet the requirements of an argillic horizon.

Psamments occur in a wide range of climates. Five of the six great groups reflect a climatic/geographic bias. Only Quartzipsamments do not indicate a geographic location, but rather the dominant mineralogy, quartz. Psamments exist in cryic or pergelic (Cryopsamments) soil temperature regimes but also are located in aridic (Torripsamments), ustic (Ustipsamments), xeric (Xeropsamments), and udic (Udipsamments) soil moisture regimes.

Subsurface features present in some Psamments include a lithic contact (Lithic), aquic or oxyaquic conditions, lamellae (Argic), plinthite, or durinodes. Thus, subsoil features vary greatly, and the variation is noted in the subgroup taxonomic name. Examples include Lithic Cryopsamments, Plinthic Quartzipsamments, and Aquic Durinodic Xeropsamments.

At the subgroup taxonomic level, climatic/geographic conditions are also described. As an example, Aridic Ustipsamments are located in several soil temperature regimes, but must have a period of time in which the soil is dry in all parts. The length of time that the soil must be dry is defined by the soil temperature regime.

33.4.4.3.2 Genesis

Genetically, Psamments develop in areas high in sand content (>85%). The parent material can be geologically young, as in eolian or alluvial materials or old, as in sandy residual bedrock. Sand dunes of either marine (Florida) or eolian (Saudia Arabia or Nebraska) origin are common parent materials. Some Psamments form in materials weathered from sandstone bedrock (e.g., South Dakota or Montana). Depending on the mineralogy of the sand, Psamments (Quartzipsamments) can be very resistant to weathering.

33.4.4.3.3 Representative Psamments

Psamments from Florida (Astatula series: hyperthermic, uncoated, Typic Quartzipsamments), Nebraska (Valentine series: mixed, mesic Typic Ustipsamments), and Saudi Arabia (Torripsamments) were selected to represent different geographic areas and, therefore, different genetic pathways. But morphologically, these soils are very similar and they are excessively drained.

Astatula, Valentine, and the Torripsamments from Saudi Arabia have thick C horizons with thin A horizons. If an A horizon is present, it is light colored and <15 cm thick. The A horizon is subject to continual mixing by wind action. Astatula soils form in eolian and marine sands and are extensive in peninsular Florida on upland slopes that range from 0% to 30%. Valentine soils form in sand dunes in the Sandhills of Nebraska on slopes that range from 0% to 60%. The sand dunes generally are oriented in a northwest to southeast direction (prevailing wind direction). In addition to Nebraska, Valentine soils occur extensively in Kansas, Montana, New Mexico, South Dakota, Texas, and Wyoming (>2 million ha). The Torripsamments from Saudi Arabia form in sand dunes and also in alluvial sandy deposits on plains and stream terraces. The dunes range in height from <2 to >10 m.

33.4.4.3.4 Physical and Chemical Properties of Representative Psamments

Physical and chemical properties are presented for the Astatula in Florida and for the Valentine in Nebraska (Table 33.20). Astatula soils (Quartzipsamments) have low organic C content and are strongly acid (pH < 5.5) and strongly leached as a consequence of high rainfall (1270 mm year^{-1}). Torripsamments from Saudi Arabia with <100 mm year^{-1} precipitation (data not shown have an accumulation of bases and are alkaline in reaction (pH > 8.5)). The sand size fraction of the Astatula and Valentine series are dominated by fine sand. Clay content of these Psamments is <5%.

33.4.4.4 Fluvents

Fluvents are Entisols that form in alluvium of rapidly aggrading floodplains, fans, deltas, and in some cases mudflows. World coverage of Fluvents is 2.2%, principally within major river valleys (see Lindbo, 1997).

When rates of deposition in alluvial settings approximate rates of pedogenesis, A-C soil profiles develop. In many instances, the A horizon is an ochric epipedon, but in others the surface horizon may be bedded (Soil Survey Staff, 1999, 2010). Alluvial processes prevent the formation of subsurface diagnostic horizons because of variations in flood magnitude and frequency leading to fluctuations in organic carbon content with depth. This occurs because greater quantities of allogenic organic matter tends to be associated with clayey rather than sandy flood deposits (Grossman, 1983). To taxonomically capture these conditions, organic carbon content in Fluvents must decrease irregularly with depth, or be above 0.2% at a depth of 125 cm.

Fluvents are differentiated from other Entisol suborders by (1) having textures finer than loamy very fine sand and less than 35% rock fragments (Psamments); (2) occurring on slopes less than 25% and without a densic, lithic, or paralithic contact within 25 cm (Orthents); (3) not having aquic conditions (Aquents); (4) having less than 3% fragments of diagnostic horizons in no discernable order (Arents); and (5) having any soil temperature regime that excludes the presence of gelic materials (Soil Survey Staff, 1999, 2010).

TABLE 33.20 Soil Characterization Data for the Astatula Fine Sand (Sodek et al., 1990) and Valentine Fine Sand (Soil Survey Staff, 1966) of the Psamment Suborder

Horizon	Depth (cm)	Very Coarse Sand	Coarse Sand	Medium Sand	Fine Sand	Very Fine Sand	Silt	Clay	pH[a] (H₂O)	OC[b] (%)	Ca[c]	Mg	Na	K	CEC₇	Base Saturation (%)
				(%)							(cmol₍c₎ kg⁻¹)					(%)
Astatula Fine Sand (Typic Quartzipsamments)																
A	0–13	—	0.1	18.8	76.7	2.1	0.9	1.4	5.5	0.5	0.5	0.1	0.0	0.0	2.7	23
C1	13–64	—	0.1	18.4	76.5	2.4	1.6	1.5	5.2	0.2	0.0	0.0	0.0	—	1.5	5
C2	64–175	—	0.1	19.7	75.8	2.2	0.6	1.6	5.0	0.1	0.0	0.0	0.0	—	0.7	5
C3	175–203	—	0.1	19.4	76.2	2.0	0.6	1.7	5.0	0.0	0.0	0.0	0.0	—	0.9	6
Valentine Fine Sand (Typic Ustipsamments)																
A	0–10	0.2	4.2	12.7	50.2	24.7	4.1	3.9	6.0	0.1	3.6	0.6	0.1	0.3	5.9	78
AC	10–25	0.1	4.6	13.3	55.1	21.6	2.2	3.1	6.2	0.5	2.1	0.4	0.1	0.3	3.7	78
C1	25–45	—	3.6	11.2	56.2	24.0	1.8	3.2	6.5	0.2	2.0	0.6	0.1	0.1	0.1	88
C2	45–81	—	4.4	18.6	58.9	14.7	1.0	2.4	6.7	0.1	1.5	0.2	0.1	0.2	2.4	79
C3	81–137	0.1	3.8	10.9	59.1	22.1	1.3	2.9	6.8	0.0	1.8	0.8	0.1	0.1	2.8	100

Source: Modified from Eswaran, H., P.F. Reich, J.M. Kimble, F.H. Beinroth, E. Padmanabhan, and P. Moncharoen. 2000. Global carbon stocks, p. 15–25. *In* R. Lal, J. Kimble, A. Mtimet, H. Eswaran, and H. Scharpenseel (eds.) Global climate change and pedogenic carbonates. Lewis Publishers, Boca Raton, FL.

[a] Determined in a 1:1 H₂O solution for the astatula fine sand and in a 1:5 Cl solution for the valentine fine sand.

[b] OC is organic carbon.

[c] Ammonium acetate extractable Ca, Mg, Na, K, and CEC (cation exchange capacity), buffered at pH 7.

33.4.4.4.1 Other Taxonomic Levels

Great groups of the Fluvents are differentiated by either soil temperature or soil moisture regime (Soil Survey Staff, 1999, 2010). Fluvents occur in udic (Udifluvents), ustic (Ustifluvents), xeric (Xerofluvents), and aridic (Torrifluvents) moisture regimes (see Sidhu et al., 1994). They also occur in cryic temperature regimes (Cryofluvents) and in colder areas without permafrost (Gelifluvents). Fluvents form in all other soil temperature regimes (frigid, mesic, thermic, and hyperthermic), including those with minimal annual temperature change (iso).

At the subgroup taxonomic level, typical Entisols intergrade to the Andisol (andic), Mollisol (mollic), Aridisol (aridic), and Vertisol (vertic) orders, and to other soil moisture and temperature regimes not used at the great group categorical level. For additional information on the family taxonomic classification of Entisols, see the Soil Survey Staff (1999, 2010).

33.4.4.4.2 Climate, Vegetation, and Topography

Fluvents are soils in disequilibrium with the regional influence of climate, vegetation, and in some cases topography. Thus, Fluvents can occur in most climates associated with most vegetation communities and on a variety of alluvial slopes. Climate can, however, indirectly influence the chemical and physical properties of Fluvents. Alluvium transported into floodplains, fans, deltas, or mudflows is derived from erosion of soils within the surrounding drainage basin where properties are more likely to reflect regional climate conditions. These preconditioned sediments can expedite the transformation of Entisols to other soil orders.

Flood frequency and magnitude in many areas appear to be increasing because of urbanization and other changes in land use practices. Many workers have observed this hydrological phenomenon where Inceptisols, Mollisols, and Vertisols of inactive or slowly aggrading floodplains are being buried by Historic sediments coined "postsettlement alluvium" (Daniels and Jordan, 1966; Scully and Arnold, 1981; Grossman, 1983; Knox, 1987; Phillips, 1997; Gomez et al., 2004; Carson, 2006; Lecce, 2007; Rustomji, 2007). If the thickness of postsettlement alluvium, or alluvium deposited by any process, buries a soil by more than 50 cm, these soils are classified as Fluvents because of minimal development. In most cases, if a recent flood deposit is less than 50 cm thick, it is treated as a mapping phase (Soil Survey Staff, 1999, 2010).

33.4.4.4.3 Time

Ripening processes lead to the formation of Fluvents within a few hundred years after drainage from the construction of polders in the Netherlands (Pons and van der Molen, 1973). This is consistent with a compilation of data from quasi-stable floodplains globally (Buol et al., 1997) showing that ochric epipedons commonly form within several hundred years.

Under conditions of landscape stability, Entisols may quickly evolve into other soil orders. In subhumid climates of central Texas and Oregon, for example, Entisols transform into Mollisols and Inceptisols, respectively, within 600 years (Parsons et al., 1970; Nordt and Hallmark, 1990). Further, Entisols transform into Inceptisols and Vertisols within 400 years in slowly aggrading floodplain sediments of the Brazos River in central Texas (Waters and Nordt, 1995). In some humid regions, however, mollic epipedons and Mollisols can develop in as little as 100 years (Ruhe et al., 1975) whereas in other humid regions it takes Entisols 500–600 years to develop into Inceptisols (Bilzi and Ciolkosz, 1977; Scully and Arnold, 1981). Further, Fluvents transform into Inceptisols on levee and point bar facies of the lower Mississippi River valley in less than 600–800 years (Aslan and Autin, 1998). In a subhumid climate of Australia, Walker and Coventry (1976) reported that floodplains that flooded once each year maintained soil development in the Entisol stage, whereas when flood frequencies were on the order of 1–10 years, Mollisols began forming. Fluvents in arid regions can persist for a thousand years in low carbonate parent material and for up to several thousand years in carbonatic parent material (Gile, 1975). Beyond these temporal thresholds, subsurface diagnostic horizons and Aridisols begin to form.

33.4.4.4.4 Parent Material

The distribution of alluvial facies in a floodplain, fan, or deltaic system is dependent on stream type (Chorely et al., 1985), which strongly influences the distribution of pedofacies (see Autin and Aslan, 2001, for example). Sedimentological packages of mixed load, meandering streams in humid regions typically display a fining upward sequence consisting of lower channel, intermediate point bar, and upper floodbasin or levee facies. In some cases, Fluvents form entirely within one of these facies and do not exhibit a noticeable textural change within the solum. If a Fluvent develops in a gradational point bar to floodbasin sequence, for example, a fining upward textural profile will be inherited. In other situations, reverse grading may be encountered. For example, a levee may be deposited over a floodbasin facies, producing a Fluvent with a coarsening upward textural distribution. In addition, suborders of Entisols other than Fluvents may form in floodplains. Psamments sometimes form in sandy channel and point bar deposits and Aquents in poorly drained topographic positions. In contrast to meandering streams, bedload streams or streams near mountain fronts will have a preponderance of Psamments and coastal streams a preponderance of Aquents.

Figure 33.25 is a pedostratigraphic column illustrating some of these fluvial geomorphic principles. The stratigraphic column displays two fining upward point bar to overbank sequences. The first terminates in a buried Fluvent with relict rooting patterns. The Fluvent formed in the upper sedimentological package is a mappable Entisol in the modern floodplain. However, it is buried by postsettlement alluvium. As depicted, the modern buried Fluvent is taxonomically classified, with the postsettlement alluvium characterized as a mapping phase because its thickness is <50 cm (see also Figure 33.24).

33.4.4.4.5 Physical and Chemical Properties

Many alluvial soils are inherently fertile and support some of the most important agricultural centers of the world (Edelman and Van Der Voorde, 1963; Radwanski, 1968; Sidhu et al., 1977). In fact, one-half of Entisols in alluvial settings support up to 1/3 of
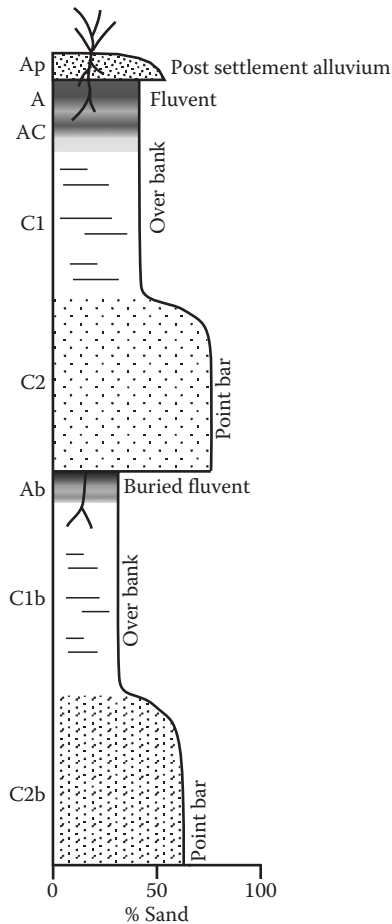
FIGURE 33.25 Pedostratigraphic diagram demonstrating fluvial geomorphic principles in relation to a buried and modern Fluvents in a floodplain setting covered by postsettlement alluvium.

the human population for food production (Buol et al., 1997). In the tropics, however, upland soils tend to be depleted in nutrients such that their alluvial counterparts produce low-fertility floodplain soils (Edelman and Van Der Voorde, 1963). This presents even greater food challenges to agricultural production in tropical regions.

As further examples, a floodplain Udifluvent and Torrifluvent display A-C profiles with textural and organic carbon stratification with depth (Table 33.21). There are similarities in clay, silt, and sand content between the two profiles, but the organic carbon content, largely detrital, is greater in the Udifluvent. Consequently, CEC_7 is slightly greater in the Udifluvent than the Torrifluvent. Perhaps the greatest difference in the two profiles is that the Udifluvent is carbonate free with the Torrifluvent having 11%–12% $CaCO_3$ equivalence. This results in a neutral pH and base saturation of ~60%–80% in the udic floodplain soil and a pH of between 8.1 and 8.5 and base saturation of 100% in the torric floodplain soil. The neutral pH of the Udifluvent would lead to greater plant availability of a number of nutrients such as iron and phosphorous (Brady and Weil, 2002). These two examples exemplify the variety of properties possible in Fluvents, largely inherited from the parent material. Soils surrounding the Udifluvent from Missouri

are Alfisols and Inceptisols (Larsen, 2002) with Aridisols and other carbonate-rich soils dominating the drainage basin of the Torrifluvent in southern New Mexico (Gile et al., 1981).

33.4.4.5 Orthents

Orthents are Entisols that do not fit the other suborder categories (e.g., they are the "other Entisols"). They are not water saturated (Aquents), the result of deep mixing or other disturbances by humans (Arents), dominated by sandy textures (Psamments), or stratified recent water-deposited sediments (Fluvents). Instead, most soils that key into this category are on recent erosional surfaces where all former diagnostic horizons have been truncated (Figure 33.26). Other Orthents have formed in recent landslides, mudflows, glacial deposits, piedmont slope deposits, and young terraces, or are in areas of recent solifluction, volcanic deposits, thin regolith over hard rocks, or weakly cemented rock, such as shale (Soil Survey Staff, 1999, 2010).

Orthents occupy approximately 10.5% of the Earth's land surface. They are more extensive than the combined area of the other Entisol suborders, which together occupy an area of about 5.7%. Orthents also have the most taxonomic subdivisions, reflecting lithologic properties inherited from exposed regolith, as well as properties related to climate, depth to bedrock, wetness, shrink-swell, and bioturbation (Soil Survey Staff, 1999, 2010). They occur in any climate and under any vegetation. Orthents in arid and semiarid climates are commonly used for rangeland, wildlife habitat, and irrigated agriculture. In humid regions, Orthents are used for pasture, hay production, cropland, and as forests.

33.4.4.5.1 Other Taxonomic Levels

Orthents are subdivided at the great group level based on the occurrence of udic, ustic, xeric, or aridic moisture regimes, or the cryic temperature regime (Soil Survey Staff, 1999, 2010). At the subgroup level, Orthents are divided according to many features that convey extragrade properties of the soil or to denote transitional intergrades to other taxonomic categories. Extragrades include lithic, lamellic, and vermic properties. Intergrades include durinodic (intergrade to Durids), oxyaquic (intergrade to Aquic subgroups), and aquic (intergrade to Aquents).

33.4.4.5.2 Occurrence and Genesis

Orthents are common in mountainous regions with arid and semiarid climates. An example of these Orthents is illustrated in Figure 33.27 as a block diagram of stepped geomorphic surfaces bordering a major floodplain mapped as Fluvents. Mountains and hills contain Orthents and areas of rock outcrop as the result of rapid erosion and truncation of former diagnostic horizons that were subsequently deposited downslope as colluvium. Widespread erosion of late Pleistocene mountain soils began in the middle Holocene around 7000 years ago in the southwestern United States (Gile et al., 1981). Such sediment makes up the mid-to-late Holocene stepped surface. The elevation of this surface was controlled by the former and slightly higher base level of the river floodplain. A recent lowering of the base level has caused the modern arroyo network to downcut into the Orthents of the Holocene deposits.

TABLE 33.21 Characterization Data of a Udifluvent (Phelps County, Missouri) and Torrifluvent (Dona Ana County, New Mexico) (Soil Survey Staff, 2008) of the Fluvent Suborder

Horizon	Depth (cm)	Sand	Silt	Clay	OC[a]	CEC_7[b] (cmol$_c$ kg^{-1})	$CaCO_3$ (%)	pH H_2O
		(%)						
Udifluvent								
A	0–13	89	8	3	0.39	2.9	0	6.9
C1	13–23	62	28	10	1.15	8.5	0	7.1
C2	23–43	51	36	13	1.05	9.1	0	7.3
C3	43–58	57	33	10	0.57	6.2	0	7.3
C4	58–81	69	24	7	0.48	4.5	0	7.4
C5	81–89	95	3	2	0.06	1.1	0	7.2
C6	89–117	97	1	2	0.02	0.7	0	7.0
C7	117–132	94	3	3	0.08	1.1	0	7.0
C8	132–152	98	0	2	0.01	0.8	0	7.1
Torrifluvent								
Ap1	0–10	69	20	11	0.67	8.2	11	8.1
Ap2	10–31	73	18	9	0.44	6.6	11	8.3
C1	31–51	81	12	7	0.28	5.2	11	8.4
C2	51–71	75	19	6	0.30	5.6	11	8.4
C3	71–107	83	13	4	0.17	4.2	11	8.4
C4	107–152	84	12	4	0.15	3.9	12	8.5

[a] OC is organic carbon.

[b] Ammonium acetate extractable Ca, Mg, Na, K, and CEC (cation exchange capacity) buffered at pH 7.
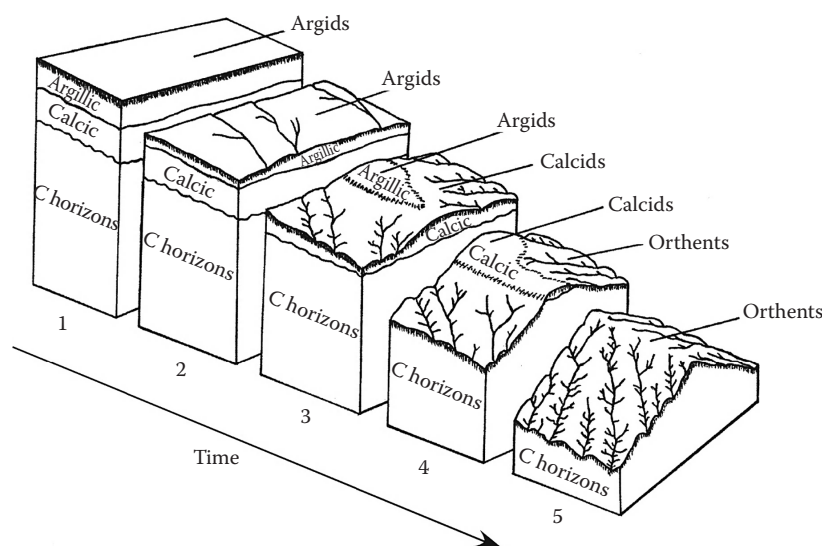
FIGURE 33.26 Formation of Orthents by the truncation of diagnostic horizons. Block diagram illustrates an Aridisol with argillic and calcic horizons that are progressively eroded, thus converting the suborders from Argids to Calcids to Orthents. (Modified from Gile, L.H., J.W. Hawley, and R.B. Grossman. 1981. Soil and geomorphology in the basin and range area of southern New Mexico—Guidebook to the desert project. New Mexico Bureau of Mines and Mineral Resources Memoir 39. New Mexico Bureau of Mines and Mineral Resources, Socorro, NM.)

Another location for Orthents in this landscape is the backslope of scarps (risers) associated with the stepped geomorphic surfaces (Figure 33.27). These erosional surfaces are truncating diagnostic horizons of an Aridisol and exhuming underlying sediments beneath the late Pleistocene stepped surface (tread). The rate of erosion on this geomorphic surface is greater than the rate of pedogenesis, and, thus, these soils have no diagnostic horizons other than the ochric epipedon. Sediments generated by erosion on the backslope are deposited on the footslope as unstratified colluvium, which is an example of a geomorphic surface having linked erosional and constructional components. Orthents associated with the erosional component are the same age as Orthents associated with the constructional component because pedogenesis would have started at the same time for both.

Once landscapes have stabilized the type of parent material can affect the rate at which diagnostic horizons form, and,
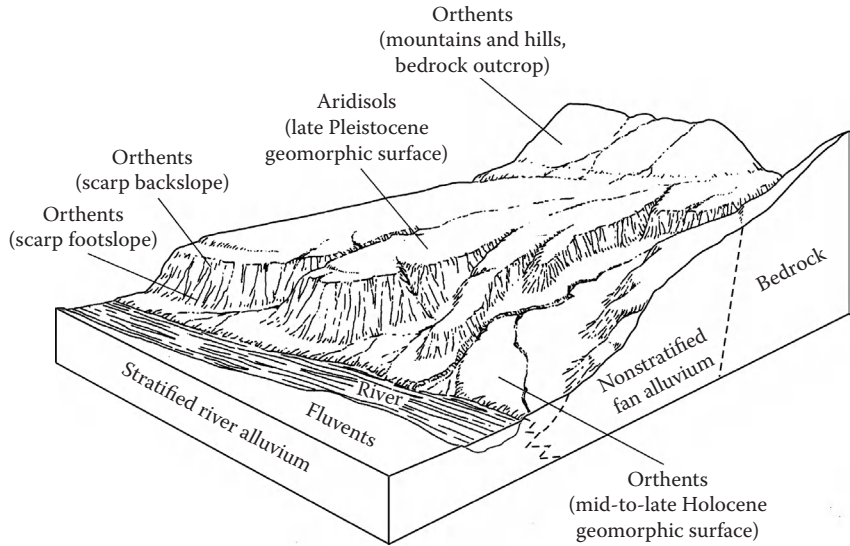
FIGURE 33.27 Block diagram of stepped geomorphic surfaces bordering a major floodplain illustrating the erosional and depositional occurrence of Orthents. (Modified from Gile, L.H., and R.B. Grossman. 1979. The desert project soil monograph. Doc. PB80-135304. National Technical Information Service, Springfield, VA.)

thus, the length of time a soil remains an Orthent. First, parent materials that contain large amounts of detrital carbonate can curtail the formation of argillic horizons by keeping clay flocculated and less mobile (Gile et al., 1981). Second, parent materials containing plutonic rocks with weatherable minerals and cleavage, such as granites, disintegrate along mineral contacts and produce argillic horizons at a rate faster than parent materials containing volcanic rock, such as rhyolite. This rock type, for example, resists weathering because they lack planes of weakness along which water can enter.

33.4.4.5.3 Physical and Chemical Properties

In contrast to soils classified according to their pedogenically produced diagnostic horizons, the physical and chemical properties of Orthents are mainly inherited from their parent materials. As a result, the range of Orthent properties is wide and reflects the properties of the unconsolidated material in which they occur, combined with bioclimatic factors. Coarse-silty Udorthents in Iowa, for example, have A horizons with organic carbon contents as high as 1.2% and cation exchange capacities of 13.6 cmol$_c$ kg^{-1} (Soil Survey Staff, 1975). In contrast, sandy-skeletal Torriorthents in New Mexico have A horizons with organic carbon contents as low as 0.4% and cation exchange capacities of 7.5 cmol$_c$ kg^{-1} (Gile and Grossman, 1979).

33.4.5 Carbon Stocks

Entisols store approximately 353 Pg (10^{15} g) of total soil carbon (TSC), which is 14.2% of all soil carbon globally (Table 33.22). Of this, 90 Pg (5.8%) of the global soil carbon pool is contained in the organic fraction and 263 Pg (27.7%) in the inorganic fraction. Of the TSC within the Entisol order, 75% is held in the soil inorganic carbon (SIC) pool and 25% in the soil organic carbon (SOC) pool (see also Wilding et al., 2002). This distribution

reflects the presence of calcium carbonate in both pedogenic forms and as detritus in the fine-earth fraction inherited from the parent material (Rabenhorst et al., 1984; Nordt et al., 1998; Kraimer and Monger, 2009). Further, there is more SIC stored in Entisols than any other soil order except Aridisols (Eswaran et al., 2000). In contrast, Entisols store a smaller proportion of the SOC carbon pool globally than most other soil orders.

Orthents store the most TSC (10% globally) of the Entisol suborders (Table 33.22). The SIC pool is high because of the widespread distribution of Orthents in arid lands. After Orthents, Fluvents store more TSC than Psamments and Aquents. The low accumulation of SOC in Aquents is misleading in that their areal extent is limited, even though great groups and subgroups may contain histic and mollic intergrades.

Sequestration rates of atmospheric CO_2 in Entisols are low because of geomorphic processes that limit pedogenesis and the

TABLE 33.22 Total Soil Carbon, Soil Organic Carbon, and Soil Inorganic Carbon Stocks by Entisol Suborder in Petagrams (Pg) and Percent

Suborder	TSC		SOC		SIC	
	Pg	%	Pg	%	Pg	%
Aquents	2	0.1	2	0.2	0	0.0
Psamments	41	1.6	8	0.5	33	3.5
Fluvents	60	2.4	33	2.1	27	2.8
Orthents	250	10.0	47	3.0	203	21.4
Total	353	14.2	90	5.8	263	27.7

Source: Modified from Eswaran, H., P.F. Reich, J.M. Kimble, F.H. Beinroth, E. Padmanabhan, and P. Moncharoen. 2000. Global carbon stocks, p. 15–25. *In* R. Lal, J. Kimble, A. Mtimet, H. Eswaran, and H. Scharpenseel (eds.) Global climate change and pedogenic carbonates. Lewis Publishers, Boca Raton, FL.

Arents undetermined.

incorporation of carbon. However, if processes that limit pedogenesis are removed, more organic carbon could be stored in surface horizons during the process of melanization. As with any parent material rich in calcium carbonate, the formation of pedogenic carbonate-carbon would not result in a net accumulation of atmospheric CO_2 (see Nordt et al., 2000). However, Entisols forming in parent materials with weatherable calcium silicate minerals could in the process of forming pedogenic carbonate sequester atmospheric CO_2 at a greater rate than currently.

33.4.6 Conclusions

Entisols are the most widespread soil order in the world forming in a variety of regions reflecting parent material characteristics and geomorphic processes that limit development to A-C or A-R profiles. They typically have ochric epipedons and no subsurface diagnostic horizons. Entisols provide engineering and agricultural challenges in areas that are wet (Aquents or Wassents), sandy (Psamments), flooded (Fluvents), steep or rocky (Orthents), or disturbed by humans (Arents). In contrast, Entisols, especially Torrifluvents, have played an important role in the origins of agricultural centers of early civilizations, such as those along the Nile, Euphrates, Tigris, Indus, and Huang He rivers.

With quasi-landscape stability in parent materials containing weatherable minerals on relatively shallow slopes, Entisols quickly form into one of the other taxonomically defined soil orders accompanied by a variety of physical and chemical changes that includes increased carbon sequestration and storage. Further, Entisols hold more TSC than any other soil order except Aridisols. Fossil Entisols are common in the geologic record as Paleosols and provide information about paleoenvironmental conditions. Entisols capture the essence of challenging pedological problems for the future in response to an ever increasing population and other global change issues related to land use.

Acknowledgment

References

Aslan, A., and W.J. Autin. 1998. Holocene flood-plain soil formation in the southern lower Mississippi Valley: Implications for interpreting alluvial paleosols. Geol. Soc. Am. Bull. 110:433–449.

Autin, W.J., and A. Aslan. 2001. Alluvial pedogenesis in Pleistocene and Holocene Mississippi River deposits: Effects of relative sea-level change. Geol. Soc. Am. Bull. 113:1456–1466.

Atchley, S., L. Nordt, and S. Dworkin. 2004. Alluvial sequence stratigraphy and sea level change: An example from the K/T transition of the Tornillo Basin, Big Bend National Park, West Texas. J. Sediment. Res. 74:392–405.

Bilzi, A.F., and E.J. Ciolkosz. 1977. Time as a factor in the genesis of four soils developed in recent alluvium in Pennsylvania. Soil Sci. Soc. Am. J. 41:122–127.

Birkeland, P. 1999. Soils and geomorphology. 3rd edn. Oxford University Press, New York.

Brady, N.C., and R.R. Weil. 2002. The nature and properties of soils. 13th edn. Prentice Hall, Upper Saddle River, NJ.

Buol, S.W., F.D. Hole, R.J. McCracken, and R. Southard. 1997. Soil genesis and classification. 3rd edn. Iowa State University Press, Ames, IA.

Carlisle, V.W., C.T. Hallmark, F. Sodek, III, R.E. Caldwell, L.C. Hammond, and V.E. Berkheiser. 1981. Characterization data for selected Florida soils. Soil Science Research Report No. 81-1. University of Florida, Gainesville, FL.

Carson, E. 2006. Hydrologic modeling of flood conveyance and impacts of historic overbank sedimentation on West Fork Black's Fork, Uinta Mountains, northeastern Utah, USA. Geomorphology 75:368–383.

Chorely, R.J., S.A. Schumm, and D.E. Sugden. 1985. Geomorphology. Methuen and Co., London, U.K.

Daniels, R.B., and R.H. Jordan. 1966. Physiographic history and the soils, entrenched stream systems, and gullies. USDA Technical Bulletin No. 1348, Harrison County, IA.

Edelman, C.H., and P.K.J. Van Der Voorde. 1963. Important characteristics of alluvial soils in the tropics. Soil Sci. 95:258–263.

Ellis, L.R. 2006. Subaqueous pedology: Expanding soil science to near-shore subtropical marine habitats. Ph.D. Dissertation. University of Florida. Gainesville, FL.

Eswaran, H., P.F. Reich, J.M. Kimble, F.H. Beinroth, E. Padmanabhan, and P. Moncharoen. 2000. Global carbon stocks, p. 15–25. *In* R. Lal, J. Kimble, A. Mtimet, H. Eswaran, and H. Scharpenseel (eds.) Global climate change and pedogenic carbonates. Lewis Publishers, Boca Raton, FL.

Fanning, D.S. 1990. Pedotechnology-soil genetic engineering: How and why soils scientists should be involved. Soil Surv. Horiz. 31:29–32.

Fanning, D.S. 1991. Human-influenced and disturbed soils: Overview with emphasis on classification, p. 3–14. *In* C.V. Evans (ed.) Proc. Conf. Univ. of New Hampshire. Department of Natural Resources, University of New Hampshire, Durham, NH.

Fanning, D.S., and S.N. Burch. 1997. Acid sulphate soils and some associated environmental problems, p. 145–158. *In* K. Auerswald, H. Stanjek, and J.M. Bigham (eds.) Soils and environments. Advances in geoecology. Vol. 30. Catena Verlag, Reiskirchen, Germany.

Fanning, D.S., and M.C.B. Fanning. 1989. Soil: Morphology, genesis, and classification. John Wiley & Sons, New York.

Fischler, K.C. 2006. Observations and characterization of subaqueous soils and seagrasses in a recently constructed habitat in the Indian River Lagoon, Florida. M.S. Thesis. University of Florida. Gainesville, FL.

Florida Statutes. 1995. Wetlands delineation, Chapters 373–4211. Ratification of Chapter 17–340, Florida administrative code on delineation of the landward extent of wetlands and surface waters. State of Florida, Tallahassee, FL.

Gile, L.H. 1975. Holocene soils and soil-geomorphic relations in an arid region of southern New Mexico. Quat. Res. 5:321–360.

Gile, L.H., and R.B. Grossman. 1979. The desert project soil monograph. Doc. PB80-135304. National Technical Information Service, Springfield, VA.

Gile, L.H., J.W. Hawley, and R.B. Grossman. 1981. Soil and geomorphology in the basin and range area of southern New Mexico—Guidebook to the desert project. New Mexico Bureau of Mines and Mineral Resources Memoir 39. New Mexico Bureau of Mines and Mineral Resources, Socorro, NM.

Gomez, B., H.L. Brackley, D.M. Hicks, H. Neff, and K.M. Rogers. 2004. Organic carbon in floodplain alluvium: Signature of historic variations in erosion processes associated with deforestation, Waipaoa River basin, New Zealand. J. Geophys. Res. 27:F04011.

Grossman, R.B. 1983. Entisols, p. 55–90. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy, II. The soil orders. Elsevier, Amsterdam, the Netherlands.

Holliday, V.T. 2004. Soils in archaeological research. Oxford University Press, Oxford, U.K.

Hupy, J.P., and R.J. Schaetzl. 2006. Introducing "bombturbation," a singular type of soil disturbance and mixing. Soil Sci. 171:823–836.

Kargbo, D.M., D.S. Fanning, H.I. Inyang, and R.W. Duell. 1993. Environmental significance of acid sulfate "clays" as waste covers. Environ. Geol. 22:218–226.

Knox, J.C. 1987. Historical valley floor sedimentation in the Upper Mississippi Valley. Ann. Assoc. Am. Geogr. 77:224–244.

Kosse, A. 1988. Anthrosols: Proposal for a new order. Agron. Abstr. 80:260.

Kraimer, R.A., and H.C. Monger. 2009. Carbon isotopic subsets of soil carbonate—Particle size comparison of limestone and igneous parent materials. Geoderma 150:1–9.

Kraus, M.J. 2002. Basin-scale changes in floodplain paleosols: Implications for interpreting alluvial architecture. J. Sediment. Res. 72:500–509.

Larsen, S.E. 2002. Soil survey of Phelps County, Missouri. USDA-NRCS, U.S. Government Printing Office, Washington, DC.

Lecce, S. 2007. Spatial patterns of historical overbank sedimentation and floodplain evolution, Blue River, Wisconsin. Geomorphology 18:265–277.

Lindbo, D.L. 1997. Entisols–fluvents and fluvaquents: Problems recognizing aquic and hydric conditions in young, flood plain soils, p. 133–152. *In* M.J. Vepraskas and S.W. Sprecher (eds.) Aquic conditions and hydric soils: The problem soils. SSSA Special Publication No. 50. Madison, WI.

McCarthy, P.J., and A.G. Plint. 1998. Recognition of interfluve sequence boundaries: Integrating paleopedology and sequence stratigraphy. Geology 26:387–390.

Monger, H.C., and B.T. Bestelmeyer. 2006. The soil-geomorphic template and biotic change in arid and semiarid ecosystems. J. Arid Environ. 65:207–218.

Nordt, L.C., and C.T. Hallmark. 1990. Soils—Geomorphology tour guidebook. Departmental Technical Report 90-7. Department of Soil and Crop Sciences. Texas A&M University, College Station, TX.

Nordt, L.C., C.T. Hallmark, L.P. Wilding, and T.W. Boutton. 1998. Quantifying pedogenic carbonate accumulations using stable carbon isotopes. Geoderma 82:115–136.

Nordt, L.C., L.P. Wilding, and L.R. Drees. 2000. Pedogenic carbonate transformations in leaching soil systems: Implications for the global C cycle, p. 43–64. *In* R. Lal, J. Kimble, A. Mtimet, H. Eswaran, and H. Scharpenseel (eds.) Global climate change and pedogenic carbonates. Lewis Publishers, Boca Raton, FL.

Parsons, R.B., C.A. Balster, and A.O. Ness. 1970. Soil development and geomorphic surfaces, Willamette Valley, Oregon. Soil Sci. Soc. Am. Proc. 34:485–491.

Phillips, J. 1997. Human agency, Holocene sea level, and floodplain accretion in coastal plain rivers. J. Coastal Res. 13:854–866.

Pons, L.J., and W.H. van der Molen. 1973. Soil genesis under dewatering regimes during 1000 years of polder development. Soil Sci. 116:228–234.

Rabenhorst, M.C., L.P. Wilding, and L.T. West. 1984. Identification of pedogenic carbonates using stable carbon isotope and microfabric analysis. Soil Sci. Soc. Am. J. 48:125–132.

Radwanski, S.A. 1968. Field observations of some physical properties in alluvial soils of arid and semi-arid regions. Soil Sci. 106:314–316.

Rasheed, M.A., F.B. Labib, and Th.K. Ghabour. 1992. The wet soils of Egypt, p. 206–211. *In* J.M. Kimble (ed.) Proc. 8th Int. Soil Correl. Meet. (VIII ISCOM): Charact. classif. util. wet soils. USDA-SCS, National Soil Survey Center, Lincoln, NE.

Retallack, G.J. 1994. A pedotype approach to latest Cretaceous and earliest Tertiary paleosols in eastern Montana. Geol. Soc. Am. Bull. 106:1377–1397.

Retallack, G.J. 2001. Soils of the past: An introduction to paleopedology. 2nd edn. Blackwell Science, Oxford, U.K.

Ruhe, R.V., T.E. Fenton, and L.L. Ledesma. 1975. Missouri River history, flood plain construction, and soil formation in southwestern Iowa. IA Agric. Home Econ. Exp. Stn. Res. Bull. 580:38–791.

Rustomji, P. 2007. Alluvial sedimentation rates from southeastern Australia indicate post-European settlement landscape recovery. Geomorphology 90:73–90.

Saunders, T.J. 2007. Multi-scale analysis of benthic biogeochemical properties and processing in a spring-fed river and estuary. Ph.D. Dissertation. University of Florida. Gainesville, FL.

Scully, R.W., and R.W. Arnold. 1981. Holocene alluvial stratigraphy in the upper Susquehanna River basin, New York. Quat. Res. 15:327–344.

Sencindiver, J.A. 1977. Classification and genesis of minesoils. Ph.D. Dissertation. University of West Virginia. Morgantown, WV.

Short, J.R., D.S. Fanning, M.S. McIntosh, J.E. Foss, and J.C. Patterson. 1986a. Soils of the mall in Washington, DC: I. Statistical summary of properties. Soil Sci. Soc. Am. J. 50:699–705.

Short, J.R., D.S. Fanning, M.S. McIntosh, J.E. Foss, and J.C. Patterson. 1986b. Soils of the mall in Washington, DC: II. Genesis, classification, and mapping. Soil Sci. Soc. Am. J. 50:705–710.

Sidhu, P.S., R. Kumar, and B.D. Sharma. 1994. Characterization and classification of Entisols in different soil moisture regimes of Punjab. J. Indian Soc. Soil Sci. 42:633–640.

Sidhu, P.S., J.L. Sehgal, and N.S. Randhawa. 1977. Elemental distribution and associations is some alluvium-derived soils of the Indo-Gangetic plain of Punjab (India). Pedology 27:225–235.

Smith, H. 1976. Soil survey of the District of Columbia. USDA-SCS. U.S. Government Printing Office, Washington, DC.

Smith, G.D. 1986. The Guy Smith interviews: Rationale for concepts in soil taxonomy. Soil Management Support Services Technical Monograph No. 11. U.S. Government Printing Office, Washington, DC.

Smith, R.M., and A.A. Sobek. 1978. Physical and chemical properties of overburden, spoils, wastes, and new soils, p. 149–172. *In* F.W. Schaller and P. Sutton (eds.) Reclamation of drastically disturbed lands. ASA, Madison, WI.

Sodek, F., III, V.W. Carlisle, M.E. Collins, L.C. Hammond, and W.G. Harris. 1990. Characterization data for selected Florida soils. Soil Science Research Report No. 90-1. University of Florida, Gainesville, FL.

Soil Survey Staff. 1966. Soil descriptions for some soils of Nebraska. USDA-SCS. Soil Survey Investigation Report No. 5. USDA-SCS, Lincoln, NE.

Soil Survey Staff. 1975. Soil Taxonomy: A basic system of making and interpreting soil surveys. USDA-NRCS agricultural handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1990. Soil series of the United States, including Puerto Rico and the U.S. Virgin Islands. USDA-NRCS Publication No. 1483. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1998. Global soil regions. USDA-NRCS, Lincoln, NE.

Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd edn. USDA-NRCS agricultural handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2011. National Cooperative Soil Characterization Database. Available online at: http://ssldata.nrcs.usda.gov (Accessed on 7/8/2011).

Strain, M.R., and C.V. Evans. 1994. Map unit development for sand- and gravel-pit soils in New Hampshire. Soil Sci. Soc. Am. J. 58:147–155.

USDA-NRCS. U.S. Department of Agriculture-Natural Resource Conservation Service. 1996. Field indicators of hydric soils in the United States, Version 3.2. USDA-NRCS, Fort Worth, TX.

Walker, P.H., and R.J. Coventry. 1976. Soil profile development in some alluvial deposits of eastern New South Wales. Aust. J. Soil Res. 14:305–317.

Waters, M.R., and L.C. Nordt. 1995. Late Quaternary floodplain history of the Brazos River in east-central Texas. Quat. Res. 43:311–319.

Wilding, L.P., L.C. Nordt, and J.M. Kimble. 2002. Soil inorganic carbon: Global stocks, p. 709–713. *In* R. Lal (ed.) Encyclopedia of soils. Marcel-Dekker, New York.

33.5 Inceptisols

Wayne H. Hudnall

33.5.1 Introduction

Inceptisols serve two important functions in *Soil Taxonomy*. First, they group soils with incipient soil development (L. *Inceptum*, beginning), and second, they ensure that the classification system is fully bifurcated, that is, all soils fall within some taxa of the classification system. Because of the latter, Smith (1986) referred to Inceptisols as the wastebasket order. They are the repository for all soils that do not meet class differentiae of other orders. While the central concept of Inceptisols is that of soils in cool, very warm humid and subhumid regions that have weekly developed subsoil horizon (cambic) and a light colored surface horizon (Ochric), the order contains a wide variety of soils (Soil Survey Staff, 1998). In some areas, they contain soils with minimal development, while in other areas they represent soils with well-expressed diagnostic horizons that fail the criteria of other orders. All soils that have a plaggen epipedon are Inceptisols (Soil Survey Staff, 1999).

Inceptisols, because of their broad and inclusive nature, served as the precursor for new soil orders (Andisols and Gelisols) and new taxa in other orders. For example, shrink/swell soils with aquic conditions and those with cryic soil temperature regimes previously placed within Inceptisols are now Vertisols (Aquerts and Cryerts). Inceptisols with isomesic or warmer *iso* temperature regimes, previously called Tropepts, are now distributed among other great groups and families of Inceptisols with *iso* temperature regimes. Inceptisols as markers of past cultural habitats and anthropogenic activities (Plaggepts) are now termed Anthrepts, and Ochrepts, a previous suborder of Inceptisols, have now been incorporated into Inceptisols suborders reflecting soil moisture bias. Subaqueous soils formed in freshwater will be classified as Wassepts. Table 33.23 illustrates the dynamics of these changes and how suborders of Inceptisols have evolved with increased knowledge over the past 40 years.

The central concept of Inceptisols includes soils that have undergone modifications of the parent material by structural development and alteration sufficient to differentiate them from Entisols. Pedogenesis may include eluviations (losses of constituents); the translocation of clay, iron, silica, aluminum, carbonates, bases, and organic matter; and the formation of redoximorphic features by aquic (hydromorphic) conditions. Inceptisols occur in all known climates, except under aridic and pergelic conditions, and have many kinds of diagnostic horizons and epipedons. Hence, while the central concept expresses soils with incipient development, Inceptisols are an order with inordinate diversity. No attempt will be made herein to consider their full diversity. Rather, examples of the great range in physical, chemical, biological, and mineralogical attributes of this order pertinent to land use and behavior will be illustrated. For a more extensive coverage of these attributes, the reader is referred to online resources (http://ssldata.nrcs.usda.gov/).

TABLE 33.23   Inceptisols Suborders Reflecting the Evolution of Soil Taxonomy

7th Approximation (1960)[a]	Soil Taxonomy (1975)[b]	Soil Taxonomy (1998)[c]	Keys to Soil Taxonomy (2010)[d]
Andepts	Andepts	Aquepts	Anthrepts
Aquepts	Aquepts	Ochrepts	Aquepts
Ochrepts	Ochrepts	Plaggepts	Cryepts Gelepts
Umbrepts	Plaggepts	Tropepts	Udepts
	Tropepts	Umbrepts	Ustepts
	Umbrepts		Xerepts

Source: Soil Survey Staff. 2010. Keys to Soil Taxonomy. 11th edn. U.S. Government Printing Office, Washington, DC.

[a] Soil Survey Staff (1960).

[b] Soil Survey Staff (1975).

[c] Soil Survey Staff (1998). Soil

[d] Soil Survey Staff (2010).

33.5.2 Distribution

The worldwide distribution of Inceptisols is illustrated in Figure 33.28. Because these soils lack a sharply focused central concept, they occur indiscriminately globally. Where Inceptisols reflect youthfulness, they occur in high-gradient mountainous regions, along major river systems as deltaic/fluvial plains, and as soils developed from carbonate-rich bedrocks or sediments in positions of geomorphic instability. Where Inceptisols serve to bifurcate *Soil Taxonomy* there in no particular pattern to their occurrence, but they are commonly dispersed among more strongly developed Alfisols, Ultisols, and Oxisols in tropical and subtropical regions (Figure 33.28).

33.5.3 Formation of Inceptisols

Because soils placed in Inceptisols are varied and represent many different pedogenic processes and combinations of soil-forming factors, examples will be given that typify central concepts or exhibit unique characteristics.

33.5.3.1 Parent Material

Foss et al. (1983) state that Inceptisols develop on geologically young sediment or landscapes and/or under environmental conditions that inhibit soil development (e.g., coarse-textured siliceous deposits that are resistant to weathering or parent materials within cool climates that inhibit pedogenesis). Given the above constraints, parent materials include almost all types of igneous, metamorphic, and sedimentary materials (residuum, alluvium, colluvium, loess/eolian, glacial drift, etc.). Inceptisols are excluded from soils having a sandy texture throughout because it is believed that pedological features indicative of eluviations, weathering, translocation, and transformation (e.g., development of color Bw horizons) take a very short time to form in subsoils with such low specific surface area coarse-textured materials. The kinds and arrangement of horizons, and their chemical, physical, and mineralogical properties, are controlled to a major degree by the kinds of parent material given from which Inceptisols have developed. For example, fine-textured parent materials give rise to clayey Inceptisols, acidic parent materials result in base-poor (Dystr)
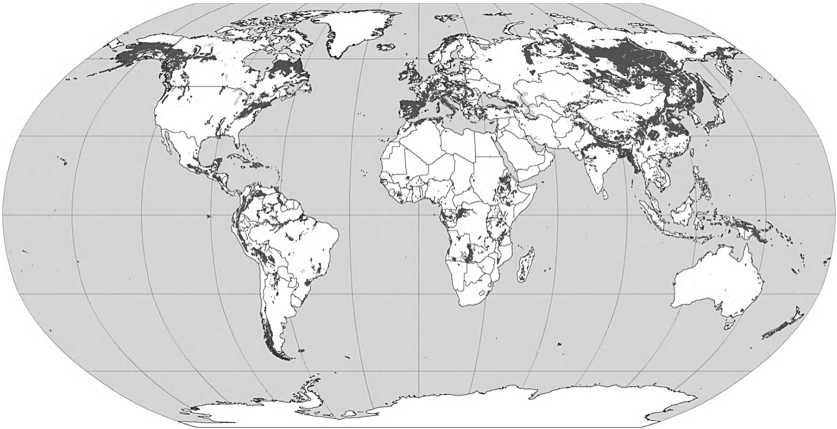
FIGURE 33.28   Global distribution of Inceptisols. (Courtesy of USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC, 2010.)

Inceptisols, and calcareous parent materials form neutral or alkaline, base-rich (Eutr) Inceptisols. Depth, color, organic carbon content, and drainage frequently reflect the combination of geomorphic position, landscape instability, and nature of the parent materials.

Table 33.24 illustrates the diversity in selected physical, chemical, and mineralogical properties of soils developed from six different parent materials [e.g., Mississippi River alluvium (*Fluvaquentic Epiaquepts*); noncalcareous glacial outwash (*Aeric Fragiaquepts*); Permian red bed shales (*Vertic Calciustepts*); basic igneous rock (Oxic Haplustepts); colluviums weathered from sandstone, siltstone, and shale (Typic Dystrudepts); and Coastal Plains alluvium (Typic Sulfaquepts)]. Abridged attributes of

these soils follow with detailed morphological, physical, chemical, and mineralogical data found online at Soil Survey Data Mart.

33.5.3.1.1 Fluvaquentic Epiaquepts (Conciene Series)

These soils developed from loamy alluvial sediments that have weathered into soils with moderately expressed cambic horizons. Their nearly level to depressional landforms favors restricted drainage, especially under humid climates. Under slightly better drained conditions, soils developed in this alluvium may form argillic horizons in a relatively short time (hundreds to several thousand years). This is because the eluviation–illuviation process proceeds relatively rapidly in such permeable, base-rich, medium textured parent sediments.

TABLE 33.24 Physical and Chemical Properties of Selected Pedons in the Inceptisols Order

Horizon	Depth (cm)	Clay	Sand (%)	OC	Extr. Fe	CEC/Clay	WRD (cm cm⁻¹)	CEC (cmol$_c$ kg⁻¹)	BS	CaCO$_3$ (%)	pH$_{H_2O}$	COLE (cm cm⁻¹)
Conciene Series (S88-LA-047—001) (Fine, Silty, Mixed, Superactive, Hyperthermic, Fluvaquentic Epiaquepts), LA												
Ap1	0–14	13.8	20.4	1.94	0.6	1.25	0.25	17.3	100	—	6.8	0.024
Bg2	51–78	15.9	17.6	0.30	0.4	0.91	0.22	14.5	100	1	7.7	0.009
Bssgb	180–215	50.4	4.3	0.71	1.0	0.69	0.19	34.9	100	1	7.5	0.116
Vernon Soil Series (S80-TX-253—001) (Fine, Mixed, Semiactive, Thermic, Vertic Calciustepts), TX												
A1	0–6	31.5	23.7	1.46	0.9	0.65	—	20.6	100	7	8.3	—
Bca2	45–65	40.5	13.8	0.21	1.7	0.34	0.08	14.8	100	23	8.5	0.042
Cr2	160–190	52.9	5.2	0.04	1.9	0.48	—	25.5	100	11	8.7	—
Kahana Series (S89-HI009—004) (Very Fine, Kaolinitic, Semiactive, Isohyperthermic Oxic Haplustepts), HI												
Ap	0–38	71.9	3.3	1.20	5.7	0.21	—	14.8	26	—	4.7	—
Bo2	56–87	59.8	7.1	0.45	5.9	0.23	—	14.0	71	—	6.2	—
Bo4	120–155	48.3	12.1	0.33	5.9	0.28	—	13.7	80	—	6.3	—
Volusia Series (S86-NY-025—001) (Fine-Loamy, Mixed, Mesic, Aeric Fragiaquepts), NY												
Ap	0–20	—	—	4.50	2.0	—	—	22.5	44	—	5.4	—
Bw	20–29	—	—	0.95	1.8	—	—	11.3	33	—	5.5	—
Eg	29–42	—	—	0.27	1.2	—	—	6.3	30	—	5.2	—
2Bx3	90–145	—	—	0.16	1.5	—	—	10.9	68	—	6.2	—
Fedscreek Series (S83-KY-195—018) (Coarse-Loamy, Mixed, Semiactive, Mesic, Typic Dystrudepts), KY												
A	0–9	13.5	38.0	2.17	1.2	0.75	0.28	10.1	37	—	5.0	1.3
Bw2	42–76	14.7	43.6	0.20	1.2	0.34	0.17	5.0	20	—	4.8	0.4
BC	122–152	17.4	36.8	0.43	1.5	0.41	0.13	7.2	31	—	4.9	0.5
Gimhae Series (S85-FN-515—012) (Fine-Silty, Mixed, Active, Mesic, Typic Sulfaquepts), Korea												
Ap1	0–9	33.2	4.6	1.97	1.5	0.34	0.19	11.3	80	—	5.1	0.2
Bg1	15–32	34.4	5.2	1.59	1.9	0.38	—	13.2	42	—	4.0	—
Bg3	66–125	19.4	29.2	1.64	0.9	0.58	0.28	11.3	100	—	3.6	1.4
Vidhas Soil (S94-FN-120—009) (Fine-Silty, Mixed, Superactive, Thermic, Fluventic Haploxerepts), Albania												
Ap	0–25	21.4	12.4	0.98	0.9	0.75	0.17	16.1	100	20	8.1	0.6
Bw1	49–83	31.1	0.9	0.59	0.9	0.66	0.18	20.5	100	20	8.3	1.6
Bw2	83–116	59.8	1.9	0.69	1.3	0.52	0.09	31.2	100	15	8.2	1.4
Moran Series (S85-CO-051—005) (Loamy-Skeletal, Mixed, Superactive, Humic Dystrocryepts), CO												
Oi	8–0	—	—	16.6	—	—	—	—	—	—	—	—
A1	0–19	17.5	46.1	5.07	2.3	1.33	—	23.3	33	—	4.8	—
A2	19–64	13.1	54.0	3.36	2.3	1.40	—	18.4	24	—	5.2	—
Bw	64–98	14.5	50.7	3.05	2.3	1.32	—	19.2	22	—	5.2	—

Source: Soil Survey Staff. 2011. National Cooperative Soil Characterization Database. Available online at http://ssldata.nrcs.usda.gov Accessed 7/8/2011.
OC, organic C; WRD, water retention difference; BS, base saturation; COLE, coefficient of linear expandability.

33.5.3.1.2 *Vertic Calciustepts (Vernon Series)*

The calcareous clays and shales of the Permian red beds restrict water percolation and profile development. Well-drained soils that develop in these sedimentary materials have a calcic or salic horizon whereby lithogenic carbonates and soluble salts have been translocated to subsoils and precipitated as pedogenic products at the terminus of the wetting front. The depth of carbonate and salt translocation, degree of calcic or salic horizon expression, and thickness of the cambic horizons are conditioned by the landform position, texture, and permeability of the parent materials.

33.5.3.1.3 *Oxic Haplustepts (Kahana Series)*

These well-drained, clayey soils have developed in basic igneous rocks in Hawaii. Under a warm, tropical environment, the weathering of nonresistant mafic-rich minerals and fine grained to amorphous materials has been rapid. An oxic-like (sesquioxide rich, subactive clay) cambic subsoil has developed. On a stable landform position, under an ustic moisture regime, subsequent weathering over time will result in an Oxisol with low activity clay.

33.5.3.1.4 *Aeric Fragiaquepts (Volusia Series)*

These soils have developed from weakly or noncalcareous glacial outwash of stratified sand, silt, and gravel. Upon weathering, these poorly sorted materials form loamy pans that have a brittle or fragic rupture upon deformation (fragipans). Weathering discontinuities in these materials enhanced by textural stratification are believed to result in a weak cementation between closely packed skeleton grains (Smeck et al., 1989). The bonding agents between the grains are believed to be hydrous oxides of Al and Si and/or translocated clay coatings or bridges. Restricted drainage would complement transport discontinuities in these soils to form weathering discontinuities.

33.5.3.1.5 *Typic Dystrudepts (Fedscreek Series)*

These well-drained soils form under udic moisture regimes in loamy colluviums weathered from sandstone, siltstone, and shale. The parent materials consist of acid channery-sized skeletal materials. Because the chemical and physical buffering capacity of the colluvium is low, and the parent materials are siliceous clasts, very acid conditions develop upon weathering. Profile development within solum, however, is minimal because the preweathered clasts are resistant to further decomposition and disintegration. Hence, release of clay and other products of weathering for subsequent translocation and transformation are limited. Again, the nature of the parent material has been pivotal in governing the intensity, modes, and mechanisms of pedogenesis in these coarse-textured, highly permeable, acid upland soils.

33.5.3.1.6 *Typic Sulfaquepts (Gimhae Series)*

These soils are common along coastlines with tidal influence. Here, the parent materials are medium textured alluvial deposits, usually associated with deltas or depositional regions of rivers that empty into the ocean. These wet environments are favorable for the accumulation of metal sulfides (e.g., pyrite), which upon drainage (natural or artificial) oxidize to form sulfuric acid and other acid sulfate weathering products (Van Breeman, 1982). Sulfaquepts are important soils because of their land use, distribution, and behavior are conditioned by their acidic character.

33.5.3.2 Climate

Inceptisols occur in subhumid to humid climates from equatorial to tundra regions where, if they have permafrost within 2 m, they are replaced by Gelisiols (Soil Survey Staff, 1998, 2010). Inceptisols cannot have an aridic (torric) moisture regime unless they have an Anthropic epipedon, which causes them to become Anthrepts.

Soil temperature and moisture are used at the suborder level in *Soil Taxonomy* to reflect major controls on current soil-forming processes. Such controls influence not only rates, modes, and mechanisms of physical and chemical weathering, but also biological and geomorphic processes that influence relief, landscape stability, topography, and soil drainage. Because Inceptisols form over a wide range of climatic conditions, excluding aridic and pergelic, the climatic impact on Inceptisols is similar to that of most other orders.

Soil moisture regime is used as a differentiation for four Inceptisols suborders (Aquepts, Udepts, Ustepts, and Xerepts), soil temperature regime for one suborder (Cryepts), and past human cultural impacts for one suborder (Anthrepts). For Cryepts, it was judged that temperature constraints on pedogenesis, use and management were more important than moisture constraints, even though both are important; the soil temperature regime was used as a differentiation at the family level for all orders except Gelisols and those soils that have a cryic soil temperature regime. Anthrepts have no particular climatic bias and are of limited geographical extent.

Aquepts form under aquic or excess wetness conditions (saturation with water, periodic reduction, and formation of contemporaneous redoximorphic features, redox concentrations and/or depletions). Udepts form under better drained, strongly leaching environments, where precipitation exceeds evapotranspiration. These soils generally lack salts and carbonates in the solum and have undergone moderate to strong weathering. Generally, summer periods of soil moisture deficit are insufficient to negatively impact mesophytic agricultural crops. Ustepts are formed under drier climates with significant periods of summer soil moisture deficit. Precipitation is less than evapotranspiration, at least during significant periods of the summer. They commonly exhibit incomplete leaching, weak to moderate weathering intensities, and pedogenic salts and/or carbonates in the lower solum. Moisture deficit is sufficiently severe to limit crop growth during summer months. Xerepts are similar to Ustepts but are formed in climates where winter precipitation is dominant. These soils are common in Mediterranean regions of Europe and along the Pacific coastal zones of western United States. The Cryepts serve as an interface between the cold soils of high latitudes

with pergelic conditions (permafrost within 2 m) and cold soils of temperate region or orographic mountainous zones without permafrost within 2 m.

In Table 33.24, the climate effects on soil classification and corresponding properties are illustrated for a few pedons. Vertic Calciustepts (*Vernon series*) and Oxic Haplustepts (*Kahana series*) are examples of soils with an ustic moisture regime and contrasting cambic horizon composition, development, and chemical characteristics. The Calciustepts are incompletely leached, have carbonates throughout the pedon, and have a clay mineral composition that favors shrink/swell activity with changing moisture content. Comparison between Fluventic Haploxerepts (*Vidhas series*) and Humic Dystrocryepts (*Moran series*) reflecting contrasting temperature and precipitation regimes due to elevation differences. Although parent materials are different, climate is the major determinant governing increased organic matter contents in the Humic Dystrocryepts, which formed under udic moisture and cryic temperature regimes about 1050 m elevation. This is compared to the Fluventic Haploxerepts formed under xeric moisture and thermic temperature regimes at an elevation of about 90 m. The Dystrocryepts have accumulated much greater organic carbon contents in the Histic layer (Oi) overlaying and umbric epipedon and a low base subsoil. In contrast, the Haploxerepts have accumulates much less organic matter in the ochric epipedon, and the presence of carbonates distributed throughout the profile implies weak leaching potentials. Climate, in this example, is considered an intensity factor whereby the rate of soil development increases as temperature and rainfall increase, but decreases as either decreases. Limit soil moisture during periods of optimal soil weathering limits profile development in the Haploxerepts. In contrast, the Dystrocryepts have optimal moisture for pedogenesis, but the soil temperature is limiting.

33.5.3.3 Topography

The topography associated with many Inceptisols is a sloping to strongly sloping landscape subject to erosion rates that are equal to or greater than soil formation rates. Many Inceptisols occur on hillsides of major mountains throughout the world. Profile developments here are constrained by unstable landforms and cool climates. Youthful Inceptisols are common also on nearly level fluvial, lacustrine, and deltaic plains (or associated depressions) where profile development is limited by age of the sediments and/or shallow groundwater tables. Coastal estuaries are also probable locations for Inceptisols, especially those where metallic sulfides and sulfates are likely to accumulate. Subaqueous condition favor the formation of the cambic horizon of the soils formed before subsidence and sea level rise. Processes that are active for subaqueous conditions that form cambic horizons have not been extensively studied.

33.5.3.4 Vegetation

It is difficult to isolate vegetation as an independent soil-forming factor governing the synthesis of Inceptisols because vegetation is dependent on climate and topography. Interdependency of these factors is easily demonstrated by comparing the Humic Dystrocryepts (*Moran series*) and Vertic Calciustepts (*Vernon series*) of Table 33.24. Dystrocryepts support mountain meadow and alpine vegetative communities such as Kobres (*Koberesia* sp.), Asters (*Aster* sp.), and Yarrow (*Achillea* sp.) under cold, moist climates. Under hotter, drier conditions of the Calciustepts, native vegetation is mixed, midgrass prairie consisting of buffalograss (*Buchloe dactyloides*), Grama grasses [blue (*Bouteloua gracilis*), hairy (*B. hirsuta*), and sideoats (*B. curtipendula*)], and tobosa grass (*Hilaria mutica*). Primary productivity, accumulation of soil organic matter, and leaching potentials are governed by effective soil moisture and temperature regimes. Effects on soil properties are remarkably evident in these two soils (Table 33.24). Primary productivity for the Dystrocryepts is limited by soil temperature regime while a seasonal soil moisture deficit limits primary productivity for Calciustepts. Higher organic matter accumulation, low turnover rates, and moderate to strong leaching potentials (acid subsoils) are consequences in Dystrocryepts. In contrast, lower organic matter contents, high turnover rates, and less effective leaching potentials (calcareous solum) occur in Calciustepts.

33.5.3.5 Time

Many examples are available in the literature to demonstrate that time is a soil-forming factor (Jenny, 1941, 1980). Some might argue that factors other than time have more influence on soil formation. While this may be true in some cases, time in combination with the intensity factor (climate) has been responsible for many attributes of Inceptisols. Inceptisols are usually considered immature soils, but some have advanced stages of development and are just short of completing their evolutionary cycle from Entisols to Oxisols. For example, Oxic Haplustepts are similar to Oxisols but either have oxic conditions too deep in the soil (greater than 150 cm) or have CEC charge characteristics just a little too high for oxic horizon requirements—a prerequisite for all Oxisols. Upon erosion and truncation, some Inceptisols have soil properties at depths associated with old stable landforms prior to dissection. Further, several taxonomic classes of Inceptisols have morphological features that could be interpreted either as youthful or an advanced stage of pedogenesis. Arenic Eutrudepts is one such example in which the soils may have developed from recent sandy sediments or may be residuum from weakly consolidated sandstone on old stable landforms.

33.5.4 Morphology

Inceptisols have many kinds of diagnostic horizons and epipedons, but the most common sequence of horizons is an ochric or umbric epipedon overlying a cambic or fragipan horizon. However, a cambic horizon is not required if an umbric, histic, or plaggen epipedon is present or if there is a fragipan, duripan, placic, calcic, petrocalcic, gypsic, petrogypsic, salic, or sulfuric horizon (Soil Survey Staff, 1998, 2010). They can also have anthropic and mollic epipedons but not argillic, kandic, or natric horizons unless buried below recent sediments. Oxic horizons are permitted only if the upper boundary is deeper than

150 cm. Spodic horizons are permitted only if they are less than 10 cm thick or if the upper boundary is deeper than 50 cm.

It is clear from the above discussion that the number and kinds of processes active in Inceptisols and the definition of Inceptisols are unavoidably complex because the order serves dual functions in grouping soils of minimal pedogenic development and bifurcating *Soil Taxonomy*. In summary, they are youthful soils in rejuvenated landscapes but pedogenetically advanced on more stable landforms. Their attributes are highly variable and seldom uniquely definitive.

33.5.5 Pedogenic Processes

Pedogenic processes common to Inceptisols have been briefly considered under Section 33.5.3. Such processes are converged in greater detail in Chapters 30 and 31. The processes germane to Inceptisols are combinations and complexes of subprocesses and reactions found in most other soil orders. The nature of these processes, including their terminology, is reviewed by Marbut (1921), Kellogg (1936), Byers et al. (1938), Simonson (1959), Buol et al. (1980), Arnold (1983), Fitzpatrick (1986), and Fanning and Fanning (1989).

Biological processes including establishment of macro and micro fauna and flora, microbial and fungal mineralization, decomposition (mummification) and accumulation of organic matter, biotic transformations (e.g., ammonification, nitrification, denitrification, and nitrogen fixation), and pedoturbation are important in epipedon and cambic horizon genesis. Chemical processes including hydration, hydrolysis, solution, mineral synthesis, and oxidation/reduction reactions are responsible for both epipedon and cambic horizon synthesis. Physical processes contribute to the formation of pedogenic structure (obliteration of at least 50% of the "rock" structure) and transport of materials within the solum. They include aggregation, expansion/contraction (shrink/swell), freeze–thaw, and mass transport by solution or suspension (translocation).

Formation of a cambic horizon requires conversion of rock structure to soil structure (peds) and evidence of at least one of the following: oxidation/reduction (aquic conditions); neoformations of clay; translocation of salts, carbonates, clay, and/or organic–metal complexes; or liberation of free iron oxides. The cambic horizon is a prerequisite for most Inceptisols, but exceptions occur for certain epipedons or subsurface diagnostic features (Section 33.5.3).

Illuviation and/or neosynthesis of clay and organic–metal complexes cannot be of such magnitude that argillic, natric, kandic, or spodic horizons are formed, unless buried. Processes responsible for formation of oxic horizons include desilication, neoformation of low activity clays, and/or residual concentration of secondary oxides and oxyhydroxides of iron and aluminum (Kellogg, 1936; Sivarajasingham et al., 1962; Cline, 1975; Jenny, 1980). Acid sulfate soils are formed by processes termed sulfidization and sulfurization. These processes are responsible for the accumulation of metal sulfides in soils and sediments, oxidation of these sulfides, formation of sulfuric acid, ferrolysis, soil acidification, and mineral dissolution (Kittrick et al., 1982; Van Breeman, 1982; Fanning and Fanning, 1989).

33.5.6 Classification

In the 1938 classification scheme (Baldwin et al., 1938) and in most subsequent revisions in the United States prior to *Soil Taxonomy*, Inceptisols were included in a number of great soil groups including Ando, Hydro Humi Latosols, Humic Gley, Tundra, Half Bog, Sols Bruns Acids, Brown forest, Regosols, Lithosols, Aluvial, and numerous other minor groups (Foss et al., 1983). Even at that time, this was a very extensive and diverse taxa of soils. In the Canadian soil classification system, the wet Inceptisols (Aquepts) were classified as Gleysols but most of the better drained members were Brunisols. In the legend of the FAO/UNESCO soil map of the world (FAO-UNESCO, 2003), most well-drained Inceptisols would be classed as Cambisols or Phaeozems if the base saturation is low. The wet Inceptisols would be classed mostly as Gleysols except those with acid sulfate conditions, which would be grouped with Fluvisols.

Inceptisols suborders, great groups, and subgroups are represented in Table 33.25. Rationale for these taxa has been considered previously in this chapter and in Chapter 32. The following discussion of suborders has been taken directly from *Soil Taxonomy* (1998) to illustrate the general nature and characteristics of Inceptisols suborders.

33.5.6.1 Anthrepts

These are the Inceptisols that are the more less freely drained that have either an anthropic or a plaggen epipedon. Most of these soils have been used as cropland or places of human occupation for many years. They can have almost any temperature regime, and almost any vegetation. Most of them have a cambic horizon. Only two subgroups are recognized at this time, those with plaggen epipedons indicative of ancient cultural sites (*Plaggenanthrepts*), and those with anthropic epipedons reflecting other long-term human impacts on soils such as irrigated farmlands or housing areas. Most of these soils are in Eurasia or Northern Africa.

33.5.6.2 Aquepts

These are the wet Inceptisols. Their natural drainage is poor or very poor and, if they have not been artificially drained, groundwater is at or near the soil surface at some time during normal years but typically not during at all seasons. They mostly have a gray to black surface horizon and gray subsurface horizon with redox concentrations that begin at a depth of less than 50 cm. A few have a brownish surface horizon that is less than 50 cm thick.

Most Aquepts have developed in late Pleistocene or younger deposits in depressions, nearly level plains, or floodplains. They occur from the equator to latitudes with discontinuous permafrost. The common features of most of these soils are the gray and rusty colors of redoximorphic features at a depth of 50 cm or less and the shallow groundwater or artificial drainage. They may have almost any particle size class except fragmental, and any reaction class, any temperature regime except pergelic, and almost any

TABLE 33.25 Listing of Suborders, Great Groups, and Subgroups for Inceptisols Order

Suborder	Great Groups	Subgroups
Aquepts	Sulfaquepts	Salidic, Hydraquentic, Typic
	Petraquepts	Histic Placic, Placic, Plinthic, Typic
	Halaquepts	Vertic, Aquandic, Duric, Aeric, Typic
	Fragiaquepts	Aeric, Humic, Typic
	Cryaquepts	Sulfic, Histic Lithic, Lithic, Vertic, Histic, Aquandic, Fluvaquentic, Aeric Humic, Aeric, Humic, Typic
	Vermaquepts	Sodic, Typic
	Humaquepts	Hydraquentic, Histic, Aquandic, Cumulic, Fluvaquentic, Aeric, Typic
	Epiaquepts	Vertic, Aquandic, Fluvaquentic, Fragic, Aeric, Humic, Mollic, Typic
	Endoaquepts	Sulfic, Lithic, Vertic, Aquandic, Fluvaquentic, Fragic, Aeric, Humic, Mollic, Typic
Anthrepts	Plagganthrepts	Typic
	Haplanthrepts	Typic
Cryepts	Eutocryepts	Humic Lithic, Lithic, Andic, Vitrandic, Aquic, Oxyaquic, Lamellic, Xeric, Ustic, Humic, Typic
	Dystrocryepts	Humic Lithic, Lithic, Andic, Vitrandic, Aquic, Oxyaquic, Lamellic, Spodic, Xeric, Ustic, Humic, Typic
Gelepts	Humigelepts	Lithic, Andic, Aquic, Oxyaquic, Fluventic, Turbic, Eutric, Typic
	Dystrogelepts	Lithic, Andic, Aquic, Fluventic, Turbic, Eutric, Typic
	Haplogelepts	Lithic, Andic, Aquic, Fluventic, Turbic, Eutric, Typic
Ustepts	Durustepts	Typic
	Calciustepts	Lithic Petrocalcic, Lithic, Torrertic, Vertic, Petrocalcic, Gypsic, Aquic, Aridic, Udic, Typic
	Dystrustepts	Lithic, Andic, Vitrandic, Aquic, Fluventic, Oxic, Humic, Typic
	Haplustepts	Aridic Lithic, Lithic, Udertic, Torrertic, Vertic, Andic, Vitrandic, Anthraquic, Aquic, Oxyaquic, Oxic, Lamellic, Torrifluventic, Udifluventic, Fluventic, Gypsic, Haplocalcidic, Calcic, Udic, Calcic, Aridic, Dystric, Udic, Typic
Xerepts	Durixerepts	Aquandic, Andic, Vitrandic, Aquic, Entic, Typic
	Calcixerepts	Lithic, Vertic, Petrocalcic, Sodic, Vitrandic, Aquic, Typic
	Fragixerepts	Andic, Vitrandic, Aquic, Humic, Typic
	Dystroxerepts	Humic Lithic, Lithic, Aquandic, Andic, Vitrandic, Anthraquic, Fragiaquic, Fluvaquentic, Aquic, Oxyaquic, Fragic, Fluventic Humic, Fluventic, Humic, Typic
	Haploxerepts	Humic Lithic, Lithic, Vertic, Aquandic, Andic, Vitrandic, Gypsic, Aquic, Lamellic, Fragic, Fluventic, Calcic, Typic
Udepts	Sulfudepts	Typic
	Durudepts	Aquandic, Andic, Vitrandic, Aquic, Typic
	Fragiudepts	Andic, Vitrandic, Aquic, Humic, Typic
	Calciustepts	Lithic Petrocalcic, Lithic, Torrertic, Vertic, Petrocalcic, Gypsic, Aquic, Aridic, Udic, Typic
	Dystrudepts	Lithic, Andic, Vitrandic, Aquic, Fluventic, Oxic, Humic, Typic

Source: Soil Survey Staff. 2010. Keys to Soil Taxonomy. 11th edn. U.S. Government Printing Office, Washington, DC.

vegetation. Most of them have a cambic horizon, and some have a fragipan. It is possible that some have a plaggen epipedon.

Table 33.25 lists the nine great groups of Aquepts. These great groups are based upon limiting features or horizons that impact use, management, and behavior. Specifically, these include soils with acid sulfuric horizons within 50 cm of the surface associated with oxidation of metal sulfides (*Sulfaquepts*); soils with a restrictive cemented subsurface horizon that forms a continuous phase within 100 cm of the soil surface (*Petraquepts*); salty (salic) or alkali (natric) subsoil horizons (*Halaquepts*); soils with fragipan (*Fragiaquepts*); soils with a cryic temperature regime (*Cryaquepts*); soils with strongly bioturbated layers by macrofauna such as crayfish, worms, and mammals (*Vermaquepts*); soils with darkened, organic-enriched surface horizons (*Humaquepts*); soils with a perched water table (*Epiaquepts*); soils that are saturated from a groundwater source (*Endoaquepts*).

33.5.6.3 Cryepts

Cryepts are the cold Inceptisols of high mountains or high latitudes. They cannot have permafrost. The vegetation is mostly conifers or mixed conifers and hardwood trees. Few of them are cultivated. These soils may be formed in loess, drift, alluvium, or solifluction deposits, mostly late Pleistocene or Holocene in age. They commonly have a thin, dark brownish ochric epipedon and a brownish cambic horizon. Some have bedrock within 100 cm of the surface. In the United States, these soils are moderately extensive in the high mountains of the West as well as other mountainous areas of southern Alaska and the world.

Two great groups of Cryepts are recognized (Table 33.25). Those that are calcareous or have high base saturation are *Eutocryepts*, and Cryepts with low base saturation are *Dystrocryepts*.

33.5.6.4 Udepts

Udepts are mainly the more or less freely drained Inceptisols that have a udic or perudic moisture regime. They formed on nearly level to steep surfaces mostly of late Pleistocene or Holocene age. Some, where the soil moisture regime is perudic, formed in older deposits. Most them had or now have a forest vegetation, but some have shrub or grass. A few have been formed from

Mollisols by truncation of the mollic epipedon (*Eutrudepts*), mostly under cultivation. Most of them have an ochric or umbric epipedon and a cambic horizon with low base saturation (*Dystrudepts*). These were the Sols Bruns Acides of earlier classification schemes (*Fragiudepts*), or a duripan (*Durudepts*). In the United States, Udepts are most extensive on the Appalachian Mountains, the Allegheny Plateau, and the West Coast; they also occur extensively in Eurasia.

33.5.6.5 Ustepts

Ustepts are the more or less freely drained Inceptisols that have a ustic moisture regime. They have dominantly summer precipitation or an isomesic, hyperthermic, or warmer temperature regime. Many of these soils are calcareous at a shallow depth and have a Bk or calcic horizon (*Calciustepts*). A few have formed from Mollisols by truncation of the mollic epipedon, mostly under cultivation. Most of them have an ochric or umbric epipedon and a cambic horizon (*Haplustepts*). Some have a duripan (*Durustepts*), especially in areas where a labile silica source is associated with pyroclastic deposits or volcanic ash. Ustepts with low base saturation (*Dystrustepts*) occur in areas with polycycled preweathered acidic sediments or outcrops of acid bedrocks common in West Africa and isolated areas of the United States. The native vegetation commonly was grass, but some supported tress and savannas. Ustepts are of moderate extent in the United States. They are most common on the Great Plains mostly in Montana, Texas, and Oklahoma. They occur extensively in ustic sectors of the Americas, Eurasia, West Africa, and several island countries.

33.5.6.6 Xerepts

Xerepts are the more or less freely drained Inceptisols that have a xeric moisture regime, dominantly winter precipitation. They have a frigid, mesic, or thermic temperature regime. They formed mostly in Pleistocene or Holocene deposits or on steep slopes. Many of these soils are calcareous at a shallow depth and have a Bk, a calcic, or a petrocalcic horizon (*Calcixerepts*). Others have low base saturation (*Dystrixerepts*). Most of them have an ochric or umbric epipedon and a cambic horizon (*Haploxerepts*). Some have a duripan (*Durxerepts*) and a few have a fragipan (*Fragixerepts*). The native vegetation was commonly coniferous forest on those with a thermic temperature regime. These soils are of moderate extent in the United States. They are most common near the West Coast in the States of California, Oregon, Washington, Idaho, and Utah. They are the major soils of the Mediterranean regions of Eurasia and Northern Africa.

33.5.7 Physical, Chemical, and Mineralogical Properties

Because of the diverse nature of Inceptisols, few generalized statements can be made about their physical, chemical, biological, and mineralogical properties. Their properties are nearly as inclusive as all the other soil orders collectively. Inceptisols span the global regions from intensively weathered to minimally developed soils. For example, they are acidic to alkaline in reaction, weak to strongly physic-chemically buffered, have low to high organic matter contents, low to high hydraulic transport functions, low to high water retention values, fertile to infertile in nutrient status, etc. The *Soil Survey Manual* (Soil Survey Staff, 1993) identifies class ranges for the above classes and for many other soil attributes described and measured. Table 33.24 illustrates the kind of magnitude of diversity vested in many of these Inceptisols attributes. The database online for Inceptisols (http://soildatamart.nrcs.usda.gov/) further documents this aspect.

The reader is referred to discussions of other orders in Sections 33.2 through 33.13 for physical, chemical, biological, and mineralogical attributes likely to be associated with Inceptisols. For example, those Inceptisols with andic surface materials are similar to Andisols; those with vertic shrink/swell features to Vertisols; those with calcic, gypsic, salic, petrocalcic, and petrogypsic horizons to Aridisols; those with translocation and/or neosynthesis of clay or organic–metal colloids to Alfisols, Ultisols, and Spodosols; those with high sesquioxide contents and/or low activity clay mineral suites to Oxisols; and those with cryic temperature regimes to Gelisols.

The sand and silt mineralogy commonly consists of quartz, mica, and feldspars with minor components of weatherable heavy minerals (opaque). The opaque minerals in soils derived from basic bedrocks are feldspars, hematite, magnetite, ilmenite, and rutile. Inceptisols rich in sesquioxides (Oxic subgroups) comprise oxyhydroxides of Fe, Al, and Ti including gibbsite, hematite, goethite, boehmite, rutile, and anatine. Those soils associated with drier climates (Ustepts and Xerepts) contain soluble salts (e.g., $NaCO_3$, $NaSO_4$, and $NaCl$), gypsum and carbonates (calcite and dolomite). Inceptisols with andic and vitrandic materials (pyroclastics) contain amorphous or short-range order minerals such as allophone, ferrihydrite, and glasses. Phyllosilicate clay minerals range from smectite, mica, kaolinite, chlorite, vermiculate, and mixed layer assemblages to rather unique suites dominated by serpentine and glauconite. In acid sulfate soils (e.g., Sulfaquepts and Sulfudepts), jarosite is a common constituent that marks very acid conditions associated with metal sulfide oxidation. Pedogenic gypsum is also common in these soils.

33.5.8 Management

For management purposes, Inceptisols can be subdivided into three land uses, namely forestry, pasture production, and agronomic cropping.

33.5.8.1 Forestry

Most Inceptisols under forested land use occur in mountainous regions on slopes ranging from 3% to 90%. On steep terrains, management systems other than natural regrowth are environmentally unacceptable and practically impossible. Most forested Inceptisols have carbon contents sufficiently high as long as the surface is not eroded, that indigenous nutrient recycling is sufficient to sustain growth without external fertilizer amendments. On less sloping terrain, many large, commercial timber

companies find fertilizer amendments, especially phosphorous applications, to be economically beneficial. Management techniques to control competition from unwanted species and disease suppression commonly involve controlled burns or physical removal of dead or unwanted species. Harvesting methods depend on the slope gradient; they range from surface skidding to aerial removal by helicopter or cable lifts. Harvest schemes may involve clear cutting or selective harvesting. Damages from erosion and compaction of harvest operations on less sloping Inceptisols are commonly ameliorated with surface tillage operations on less sloping Inceptisols are commonly ameliorated with surface tillage operations to break up surface crusts, compacted zones and to establish soil and water conservation buffers.

33.5.8.2 Pasture Production

Most of the Inceptisols under pasture management are planted to improved pastures that respond to N, P, and K fertilization, especially if the pasture is intensely managed for animal or forage production. Areas, which occur under native range, or under shifting traditional agriculture as in developing countries, should be managed using best management practices for stocking rate, pasture rotation, and fallow period. These vary from one region to another, not only in terms of soil conditions but also in terms of the ability of the operator to provide investment inputs. Commonly under traditional agriculture, the inputs are minimum and the soil dictates the best management and utilization practices to be followed. Inceptisols with an ustic or xeric moisture regime are limited by insufficient rainfall. Irrigation may be used to supplement natural precipitation on some Inceptisols, but the cost of this investment for improved pastures may make it economically unattractive. However, many examples of irrigated improved pastures occur in Europe and the United States where the economics are favorable and adequate high quality aquifers are available. Surface or subsurface drainage may be required to remove surface water or lower the water table for optimal forage production, especially for Aquepts.

33.5.8.3 Agronomic Cropping

Inceptisols are the major soils on which agronomic crops are produced in some parts of the world. Aquepts, Cryepts, Udepts, Ustepts, and Xerepts are very productive and valuable agricultural land resources, if properly managed. Crops produced depend on climate (length of growing season, dependability of precipitation, seasonal periods of soil moisture excess or deficit, photoperiodism, degree days, etc.), but because of Inceptisols diversity, most major food, feed, and fiber crops are included. Irrigation is used on vegetable, citrus, and other important cash or specialty crops when summer periods of moisture deficit become extreme. The highly weathered Inceptisols within the tropics are used for sugarcane, pineapple, cotton, and some upland rice production with many of these crops being produced under irrigation. Most of these soils respond to N, P, and K fertilizers and some soils require lime in order to neutralize acidity and sustain production.

Cryepts are used mostly for cereal crops because of the short growing season, photoperiodism, and few number of degree days. Sometimes these soils are used for cool season vegetables and forage production. Where these soils interface with Gelisols, they are constrained by cold soil temperatures, freeze–thaw heaving, and erosion induced by wind, water, and mass movement (solifluction). They may be under alpine meadow.

Aquepts used for agronomic production usually require surface or internal drainage but depending on soil temperature, a wide range of crops are grown including cotton, sugarcane, rice, and corn in the subtropical and tropical climates and corn, soybeans, sorghum, and some cereal crops in temperate climates. Sulfaquepts (and Sulfudepts) require specialized management because when they are drained mainly for lowland rice production, the sulfuric horizon produces large amounts of sulfuric acid. Because the reactions involved were not well understood when most of these soils were drained, vast wastelands were created. Management schemes, which reduce acid production, have been developed by Ponnamperuma et al. (1973), Cisse et al. (1993), Van Breeman (1982), and Coly (1996). Reclamation and management practices for acid sulfate soils have been proposed by Van Breeman (1982), Rimwanich and Suebsuri (1983), and Coly (1996). The formation of a sulfuric horizon should be prevented or minimized by controlling the depth of the water table upon drainage (keeping the soils saturated to prevent pyrite oxidation). Sulfaquepts can be reclaimed by inducing reducing conditions within the sulfuric horizon and/or flushing the soil to remove the sulfuric acid. The flushing can be accomplished using a combination of saline and freshwater to remove the soluble salts. The addition of $CaCO_3$ is required to neutralize any acidity produced, and $CaSO_4$ is required when saline water has been used for reclamation to remove excess sodium (Cisse et al., 1933; Coly, 1996).

References

Arnold, R.W. 1983. Concepts of soils and pedology, p. 1–21. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy I. Concepts and interactions. Elsevier Science Publishers, Amsterdam, the Netherlands.

Baldwin, M., C.E. Kellogg, and J. Thorp. 1938. Soil classification, p. 897–1001. *In* G. Hambidge (ed.) Soils and men. 1938 Yearbook of agriculture. USDA Government Printing Office, Washington, DC.

Buol, S.W., F.D. Hole, and R.J. McCracken. 1980. Pedogenic process: Internal, soil-building processes. Soil genesis and classification. 2nd edn. IA State Press, Ames, IA.

Byers, H.G., C.E. Kellogg, and J. Thorp. 1938. Soil classification, p. 948–978. *In* G. Hambidge (ed.) Soils and men. 1938 Yearbook of agriculture. USDA Government Printing Office, Washington, DC.

Cisse, S., W.H. Hudnall, and A.A. Szogi. 1993. Influence of land and reclamation on coastal acid sulfate soils in the Republic of Guinea, p. 101–112. *In* J.M. Kimble (ed.) Proc. 8th Int. Soil Manage. Workshop. Utilization of soil survey information for sustainable land use. USDA-NSSC, Lincoln, NE.

Cline, M.C. 1975. Origin of the term latosol. Soil Sci. Soc. Am. Proc. 39:162.

Coly, L. 1996. Management of acid sulfate saline soils in Casamance, Senegal. M.S. Thesis. Louisiana State University. Baton Rouge, LA.

Fanning, D.S., and M.C.B. Fanning. 1989. Soil morphology, genesis, and classification. John Wiley & Sons, New York.

FAO-UNESCO. 2003. Map of world Soil Resources. FAO-UNESCO, Rome. Available online at: http://www.fao.org/ag/agl/agll/wrb/wrbmaps/htm/soilres.htm.Accessed July 8, 2011.

Fitzpatrick, E.A. 1986. An introduction to soil science. 2nd edn. Longman Scientific and Technical, Longman Group U.K. Limited, Essex, U.K.

Foss, J.E., F.R. Morman, and S. Reiger. 1983. Inceptisols, p. 355–381. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy. II. The soil orders. Elsevier Science Publishers, Amsterdam, the Netherlands.

Jenny, H. 1941. Factors of soil formation. McGraw-Hill, New York.

Jenny, H. 1980. The soil resource: Origin and behavior. Springer Verlag, New York.

Kellogg, C.E. 1936. Development and significance of the great group in the United States. USDA Miscellaneous Publication No. 229. USDA, Washington, DC.

Kittrick, J.A., D.S. Fanning, and L.R. Hossner (eds.). 1982. Acid sulfate weathering. SSSA Special Publication No. 10. SSSA, Madison, WI.

Marbut, C.F. 1921. The contribution of soil survey to soil science. Soc. Prom. Agric. Sci. Proc. 41:116–142.

Ponnamperuma, F.N., T. Attanandana, and G. Beye. 1973. Amelioration of three acid sulfate soils for lowland rice, p. 391–406. *In* H. Dost (ed.) Acid sulfate soils. Proc. Int. Symp. Vol. 2. ILRI, Wageningen, the Netherlands.

Rimwanich, S., and B. Suebsuri. 1983. Nature and management of problem soils in Thailand, p. 1–23. *In* International seminar on ecology and management of problem soils in Asia. Kasetsart University, Bangkok, Thailand.

Simonson, R.W. 1959. Outline of generalized theory of soil genesis. Soil Sci. Am. Proc. 23:152–156.

Sivarajasingham, S., L.T. Alexander, J.G. Cady, and M.G. Cline. 1962. Laterite. Adv. Agron. 14:1–60.

Smeck, N.E., J.L. Thompson, L.D. Norton, and M.J. Shipitalo. 1989. Weathering discontinuities: A key to fragipan formation, p. 99–112. *In* N.E. Smeck and E. Ciolkosz (eds.) Fragipans: Their occurrence, classification, and genesis. SSSA Special Publication No. 24. SSSA, Madison, WI.

Smith, G.D. 1986. The Guy Smith interviews: Rationale for concepts in soil taxonomy. SMSS Technical Monograph No. 11. USDA-SMSS and USDA-SCS, Washington, DC.

Soil Survey Staff. 1960. Soil Classification, a comprehensive system, 7th approximation. USDA-SCS, U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1975. Soil Taxonomy. Basic system of soil classification for making and interpreting soil surveys. USDA-SCS agricultural handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1993. Soil survey manual. USDA agricultural handbook no. 18. USDA, Washington, DC.

Soil Survey Staff. 1998. Soil taxonomy. USDA-NRCS, Washington, DC. Available online at: http://soils.usda.gov/technical/classification/taxonomy/Accessed March 3, 2009.

Soil Survey Staff. 2010. Keys to Soil Taxonomy. 11th edn. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2011. National Cooperative Soil characterization Database. Available online at http://ssldata.nrcs.usda.gov. Accessed July 8, 2011.

Van Breeman, N. 1982. Genesis, morphology and classification of acid sulfate soils in coastal plains, p. 98–105. *In* J.A. Kittrick, S. Fanning, and L.R. Hossner (eds.) Acid sulfate weathering. SSSA Special Publication No. 10. SSSA, Madison, WI.

33.6 Gelisols

J.G. Bockheim

C. Tarnocai

33.6.1 Introduction

Gelisols, which are the permafrost-affected soils, constitute the 12th and newest soil order. They comprise 18 million km^2 or about 13% of the Earth's land surface and occur in the Arctic, Subarctic, Boreal, Antarctic, Subantarctic, and some alpine regions under cold continental, subhumid or semiarid, and arid conditions (Bockheim et al., 1994). They are either unvegetated or support continuously vegetated tundra, subarctic and boreal forest, and some alpine tundra. Gelisols are of global concern because they contain many protected areas, support numerous indigenous populations who depend on the land and surrounding oceans for sustenance, and may be subject to considerable impacts from human development (oil, coal and gas exploration, and mining) and global warming.

Gelisols are defined as soils having permafrost within 100 cm of the soil surface, or gelic materials within 100 cm of the soil surface and permafrost within 200 cm of the soil surface (Soil Survey Staff, 2010). Gelic materials, in turn, are seasonally or perennially frozen mineral or organic soil materials that have evidence of cryoturbation (frost churning), ice segregation, or cracking from cryodesiccation. Gelic materials contrast with other kinds of materials, such as andic or spodic materials, in being defined entirely on the basis of physical and thermal characteristics, rather than chemical properties.

A representative Gelisol is shown in Figure 33.29. Gelic materials occur in both the active layer and the upper part of the permafrost as evidenced by cryoturbation, denoted in soil descriptions by the subscript jj (y in the Canadian system).

33.6.2 Permafrost and the Occurrence of Gelisols

Gelisols only occur in areas containing permafrost within 100 cm of the soil surface if the soil is not cryoturbated or 200 cm of the surface if the soil is cryoturbated. Permafrost is defined
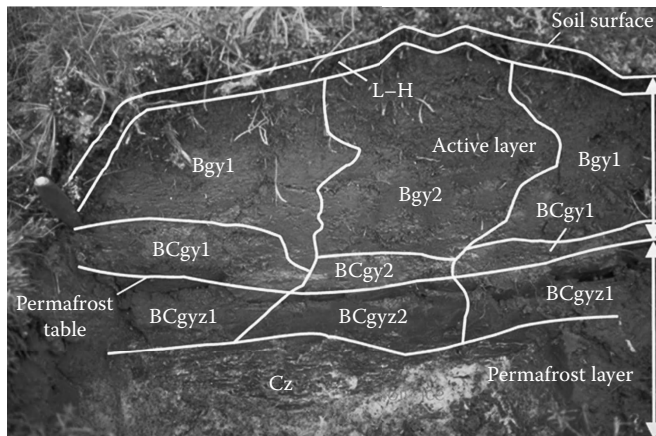
FIGURE 33.29 An aquaturbel developed on an earth hummock in northern Canada. The Canadian soil horizon nomenclature is used, where L–H = Oi–Oa, y = jj, and z = j in *Soil Taxonomy*.

here as a thermal condition in which a material (including soil materials) remains below 0°C for 2 or more years in succession. Permafrost may be ice-cemented (designated in soil descriptions with the subscript fm) or in the case of insufficient interstitial water, dry (designated as f). In the frozen layer, a variety of ice lenses, vein ice, segregated ice crystals, and ice wedges may be evident. An important consideration is that the permafrost is in dynamic equilibrium with the environment.

Permafrost, which comprises about 13% of the Earth's surface and 24% of the northern hemisphere, is differentiated into four zones on land in the circumpolar regions: continuous (91%–100% cover), discontinuous (51%–90% cover), sporadic (10%–50% cover), and isolated (<10% cover) patches (Figure 33.30).

The distribution and thickness of permafrost are influenced by natural surface features, such as snow and vegetation cover, topography, and bodies of water, but as Figure 33.30 shows, are most affected by regional climate.

33.6.3 Description and Distribution of Gelisols

There are three suborders within the Gelisol order (Soil Survey Staff, 2010): Histels, Turbels, and Orthels that are differentiated on the basis of organic matter content and for mineral soils whether or not there is cryoturbation. Distribution of Gelisols is depicted in Figure 33.31. Suborders, great groups, and subgroups are shown in Table 33.26.
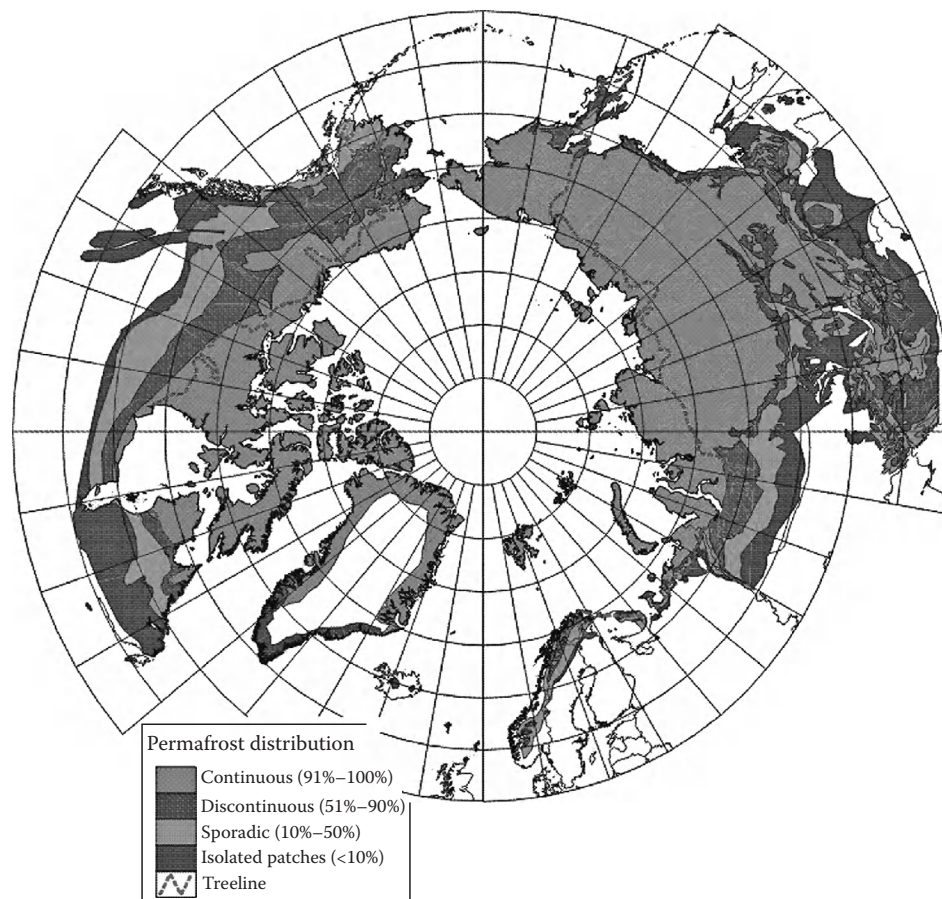
FIGURE 33.30 Distribution of permafrost in the northern hemisphere. (From Brown, J., O.J. Ferrians, Jr., J.A. Heginbottom, and E.S. Melnikov. 1997. Circumarctic map of permafrost and ground ice conditions, 1:10 000 000 scale map. U.S. Geological Survey International Permafrost Association. U.S. Geological Survey, Washington, DC.)
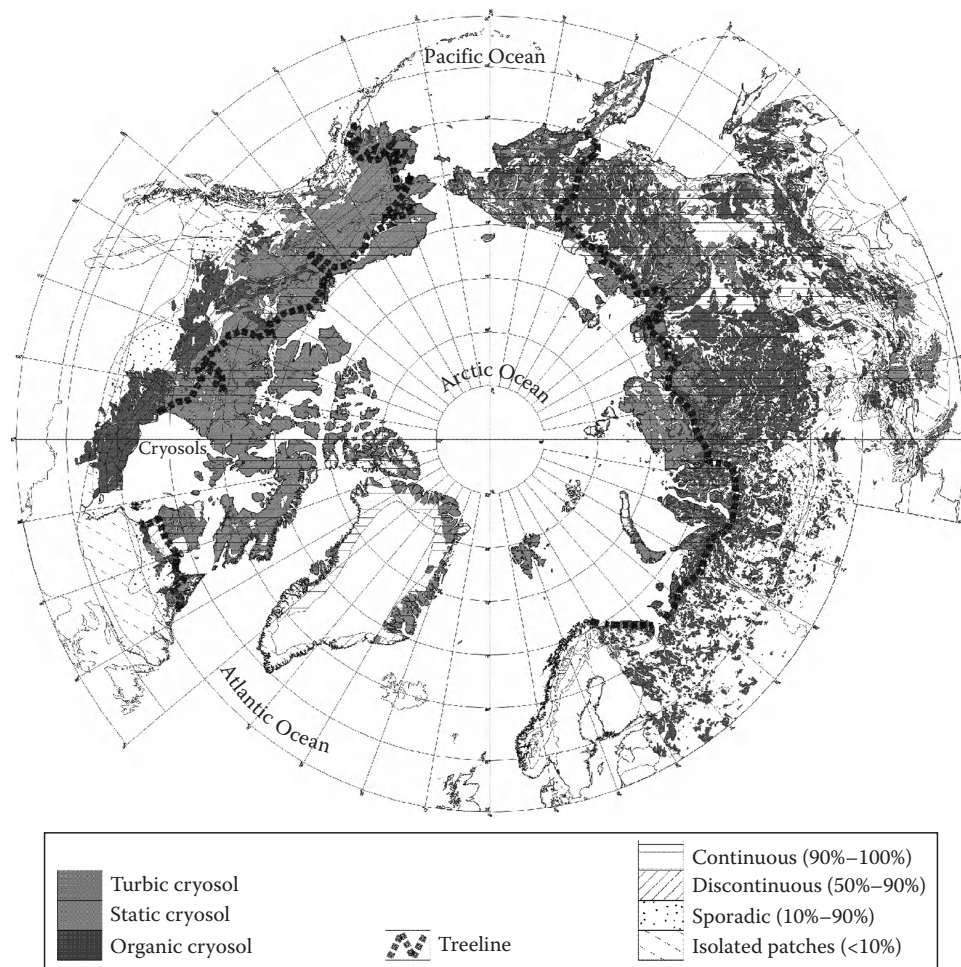
FIGURE 33.31 Distribution of Gelisols by dominant suborder in the circumarctic region. (From Tarnocai, C., D. Swanson, J. Kimble, and G. Broll. 2007. Northern circumpolar carbon database. Digital Database, Research Branch, Agriculture and Agri-Food Canada, Ottawa, Canada. With permission.)

33.6.3.1 Histels

Histels are Gelisols that have 80% or more (by volume) of organic materials from the soil surface to a depth of 50 cm, or to a restricting layer (Figure 33.32a). Histels otherwise meet the requirements of a Histosol except for the presence of permafrost within the upper 100 cm. Histels comprise 15% of the Gelisols in the circumarctic (Tarnocai et al., 2007; Figure 33.31). Most of these soils are located in the Mackenzie River Valley and the Hudson Bay lowlands of Canada. In general, Histels mostly occur in the Boreal, Subarctic, and Low Arctic regions. Histels are commonly associated with palsas (a peaty permafrost mound containing a core of alternating layers of segregated ice and peat or mineral soil), peat plateau, peat hummocks, and low- and high-centered lowland polygons. These soils have been discussed fully by Zoltai and Tarnocai (1971), Tarnocai (1972, 1973), Zoltai and Tarnocai (1974, 1975), Everett (1979), and Bockheim et al. (2004).

Histels are divided into five great groups, including the Folistels, Glacistels, Fibristels, Hemistels, and Sapristels. With the exception of the Glacistels, these great groups follow the suborders in the Histosol order (Section 34.2). The Glacistels contain a glacic layer within 50 cm of the surface that is 30 cm or more thick and contains 75% or more ice (by volume).

33.6.3.2 Turbels

Turbels are mineral soils that occur in areas with patterned ground. Patterned ground is a general term for any ground surface with a discernibly ordered, more or less symmetrical, morphological pattern of ground and, where present, vegetation (Washburn, 1980). These soils show marked influence of cryoturbation (Figure 33.32b) and occur throughout the circumpolar regions. They are the dominant suborder and account for more than 57% of the Gelisols mapped in the circumarctic (Figure 33.31). Field and laboratory data for these soils are contained in publications by Douglas and Tedrow (1960), Tedrow et al. (1968), Tedrow (1970), Zoltai and Tarnocai (1974), Pettapiece (1975), Tarnocai and Smith (1992), Bockheim and Hinkel (2007), and Simas et al. (2008).

Most horizons and layers within the active layer of Turbels are strongly affected by cryoturbation. Turbels commonly contain irregular or broken horizons, involutions, organic matter that

TABLE 33.26   Listing of Suborders, Great Groups, and Subgroups of the Gelisol Order

Suborder	Great Group	Subgroups
Histels	Folistels	Lithic, Glacic, Typic
	Glacistels	Hemic, Sapric, Typic
	Fibristels	Lithic, Terric, Fluvaquentic, Sphagnic, Typic
	Hemistels	Lithic, Terric, Fluvaquentic, Typic
	Sapristels	Lithic, Terric, Fluvaquentic, Typic
Turbels	Histoturbels	Lithic, Glacic, Ruptic, Typic
	Aquiturbels	Lithic, Glacic, Sulfuric, Ruptic–Histic, Psammentic, Typic
	Anhyturbels	Lithic, Glacic, Petrogypsic, Gypsic, Nitric, Salic, Calcic, Typic
	Molliturbels	Lithic, Glacic, Vertic, Andic, Vitrandic, Cumulic, Aquic, Typic
	Umbriturbels	Lithic, Glacic, Vertic, Andic, Vitrandic, Cumulic, Aquic, Typic
	Psammoturbels	Lithic, Glacic, Spodic, Typic
	Haploturbels	Lithic, Glacic, Aquic, Typic
Orthels	Historthels	Lithic, Glacic, Fluvaquentic, Fluventic, Ruptic, Typic
	Aquorthels	Lithic, Glacic, Sulfuric, Ruptic–Histic, Andic, Vitrandic, Salic, Psammentic, Fluvaquentic, Typic
	Anhyorthels	Lithic, Glacic, Petrogypsic, Gypsic, Nitric, Salic, Calcic, Typic
	Mollorthels	Lithic, Glacic, Vertic, Andic, Vitrandic, Cumulic, Aquic, Typic
	Umbrorthels	Lithic, Glacic, Vertic, Andic, Vitrandic, Cumulic, Aquic, Typic
	Argiorthels	Lithic, Glacic, Natric, Typic
	Psammorthels	Lithic, Glacic, Spodic, Typic
	Haplorthels	Lithic, Glacic, Fluvaquentic, Aquic, Fluventic, Typic

Source: Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS, Washington, DC.

usually accumulates on the surface of the permafrost, oriented rock fragments, and silt caps and silt-enriched subsoil horizons. Permafrost occurs within 200 cm of the soil surface. These soils are differentiated into seven great groups that link them with other orders not containing permafrost, including the Histoturbels (40%–80% organic materials by volume in the upper 50 cm), Aquiturbels (aquic conditions), Anhyturbels (anhydrous conditions) (anhydrous conditions refer to soils of cold deserts and other areas with permafrost [often dry permafrost] and low precipitation [usually <50 mm year⁻¹ water equivalent]. Anhydrous soil conditions are similar to the aridic and torric soil moisture regimes except that the soil temperature is <0°C [Soil Survey Staff, 2010]), Molliturbels (mollic epipedon), Umbriturbels (umbric epipedon), Psammoturbels (sandy texture), and Haploturbels (other Turbels).

33.6.3.3 Orthels

Orthels are mineral soils containing permafrost within the upper 100 cm, but they lack cryoturbation (Figure 33.32c). These soils comprise <3.6% of the Gelisols mapped in the circumarctic and occur primarily in areas with dry permafrost such as floodplains and the dry valleys of Antarctica. They are divided

(a)   (b)   (c)

FIGURE 33.32   Examples of the three suborders of Gelisols: (a) Histel, (b) Turbel, and (c) Orthel. (Photo courtesy by J. Bockheim.)

TABLE 33.27 Comparison of Soil Taxa among Soil Taxonomy, the Canadian System, and the Revised *World Reference Base for Soil Resources*

Soil Taxonomy (Soil Survey Staff, 2010)	World Reference Base for Soil Resources (2006)	Canadian System (Agric. Canada Expert Committee on Soil Survey, 1987)
Gelisol (order)	Cryosol (soil group)	Cryosol (order)
Histels (suborder)		Organic Cryosols (great group)
Folistels	Cryic Folic Histosols	No equivalent
Glacistels	Glacic Histosols	Glacic Organic Cryosols
Fibristels	Cryic Fibric Histosols	Fibric Organic Cryosols
Hemistels	Cryic Hemic Histosols	Mesic Organic Cryosols
Sapristels	Cryic Sapric Histosols	Humic Organic Cryosols
Terric Fibristels	Cryic Fibric Histosols (Terric)	Terric Fibric Organic Cryosols
Terric Hemistels	Cryic Hemic Histosols (Terric)	Terric Mesic Organic Cryosols
Terric Sapristels	Cryic Sapric Histosols (Terric)	Terric Humic Organic Cryosols
Turbels (suborder)		Turbic Cryosols (great group)
Histoturbels	Turbic Cryosols	Gleysolic Turbic Cryosols
Aquiturbels	Turbic Cryosols (Reductaquic)	Gleysolic Turbic Cryosols
Anhyturbels	Turbic Cryosols (Natric, Salic, Calcic)	No equivalent
Molliturbels	Turbic Crysols (Eutric)	Brunisoilc Turbic Cryosols
Umbriturbels	Turbic Cryosols (Dystric)	Brunisolic Turbic Cryosols
Psammoturbels	Turbic Cryosols	Regosolic Turbic Cryosols
Haploturbels	Turbic Cryosols	Orthic Turbic Cryosols
Glacic subgroups	Glacic Cryosols	No equivalent
Orthels (suborder)		Static Cryosols
Historthels	Histic Cryosols	Gleysolic Static Cryosols
Aquorthels	No equivalent	Gleysolic Static Cryosols
Anhyorthels	Salic, Calcic, Anhyorthels	No equivalent
Mollorthels	Mollic Cryosols	Brunisolic Static Cryosols
Umbrorthels	Umbric Cryosols	Brunisolic Static Cryosols
Argiorthels	No equivalent	No equivalent
Psammorthels	Arenic Cryosolos	Regosolic Static Cryosols
Haplorthels	Haplic, Cambic, Cryosols	Regosolic Static Cryosols
Glacic subgroups	Glacic Cryosols	No equivalent
No equivalent	Technic Cryosols	No equivalent
No equivalent	Hyperskeletic Cryosols	No equivalent
Lithic subgroups	Leptic Cryosols	No equivalent
Vitrandic subgroups	Vitric Cryosols	No equivalent
Spodic Psammorthels	Spodic Cryosols	No equivalent

into great groups that parallel the Turbels, that is, Historthels, Aquorthels, etc. Subgroups within each great group of Gelisols are listed in Table 33.27.

33.6.4 Cryopedogenic Processes

Cryopedogenic processes that lead to gelic materials are driven by the physical volume change of water to ice, moisture migration along (1) thermal, (2) hydrostatic pressure, (3) solute concentration, and (4) electrical potential gradients in the frozen (or unfrozen) system (Marion, 1995), or thermal contraction of the frozen material by continued rapid cooling. These processes include freezing and thawing, cryoturbation, frost heaving, cryogenic sorting, thermal cracking, and ice segregation (Tedrow, 1977; Washburn, 1980; Figure 33.33).

It should be emphasized that cryopedogenic processes are soil-forming processes characteristic of soils with permafrost and should not be viewed as operating against the other soil-forming processes in lower latitude soils; rather, they are distinctive processes producing horizons and properties that are uncommon to other soil orders (Bockheim et al., 2010). Processes common to the other soil orders operate in Gelisols but at a lesser magnitude because of the dominance of cryopedogenic processes.

33.6.5 The Pedon as a Basic Soil Unit

The pedon is the basic soil unit for sampling in *Soil Taxonomy* (Soil Survey Staff, 1999); it is an especially important concept for describing, classifying, and sampling Gelisols. The pedon is defined so as to encompass the full cycle of patterned ground

Cryopedogenic processes: Effects on soil morphology
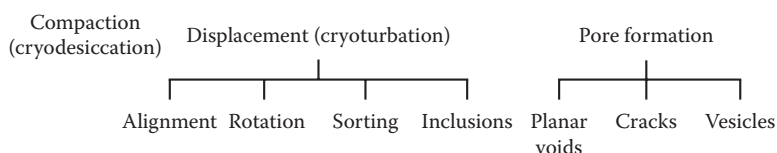
FIGURE 33.33 Conceptual framework showing the interrelationships of the effects of cryopedogenic processes on development of specific fabric types in Gelisols. (Revised from Fox, C.A. 1994. Micromorphology of permafrost-affected soils, p. 51–62. *In* J.M. Kimble and R.J. Ahrens (eds.) Proc. Meet. Classif. Correl. Manage. Permafrost-Affected Soils. USDA-SCS, National Soil Survey Center, Lincoln, NE. With permission.)

with a 1 or 2 m linear interval or a half cycle with a 2–7 m cycle (Figure 33.34a). This interval is suitable for most patterned ground features such as earth hummocks, circles, nets, and nonsorted polygons. In the case of large-scale (>7 m) ice-rich polygons, such as those that occur along the Alaskan Coastal Plain, two pedons are delineated: one within the center of the polygon and the other within the ice wedge (Figure 33.34b).

If no patterned ground exists, the pedon is arbitrarily selected but approximates about 1 m² in area.

Scaled sketches of a pedon showing soil horizons, including patches of cryoturbated material, should be drawn on graph paper in the field, or digital images should be taken and the horizons annotated directly on the image (Figure 33.29). Samples should be collected from each horizon across the pedon and composited for subsequent laboratory characterization. In the case of highly cryoturbated soils, the areal percentage of each horizon is reported in soil descriptions rather than depth intervals (Kimble et al., 1993).

33.6.6 Historical Approaches to Classification of Permafrost-Affected Soils

Tedrow (1977) provided a comprehensive review of early approaches to classification of soils in the cold regions, emphasizing the zonal systems. Of the soil taxonomic systems used today, only the approaches of Canada (Agriculture Canada Expert Committee on Soil Survey, 1987), the United States (Soil Survey Staff, 2010), and *World Reference Base (WRB) for Soil Resources* (2006) recognize soils with permafrost in a separate order (Table 33.27).

(a)

(b)
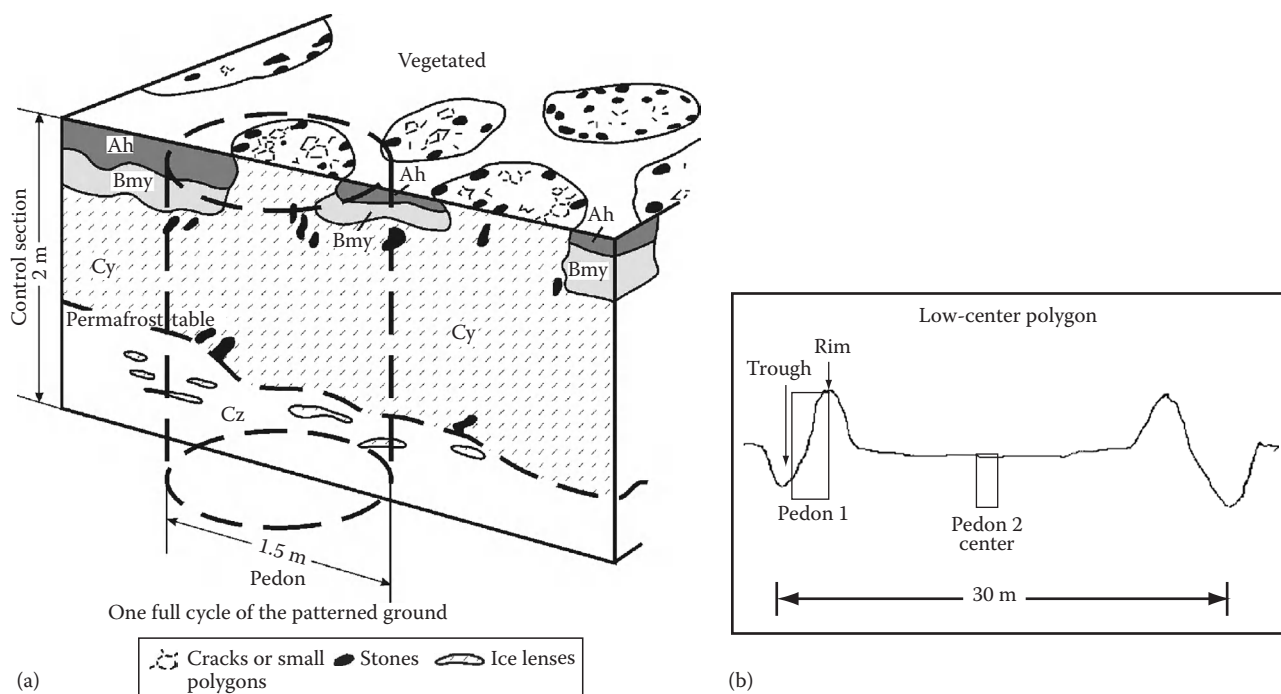
FIGURE 33.34 The pedon concept as it applies to Gelisols and patterned ground form, including (a) a nonsorted circle (From Tarnocai, C. and Smith, C.A.S., The formation and properties of soils in the permafrost regions of Canada, in: D.A. Gilichinsky (ed.), *Proceedings of the First International Conference on Cryopedology*, Russian Academy of Science, Pushchino, Russia, 1992. With permission.) and (b) a large-scale, low-centered polygon. (From Bockheim, J.G., L.R. Everett, K.M. Hinkel, F.E. Nelson, and J. Brown. 1999. Soil organic carbon storage and distribution in arctic tundra, Barrow, Alaska. Soil Sci. Soc. Am. J. 63:934–940. With permission.) The Canadian soil horizon nomenclature is used, where Ah = A, Bmy = Bwjj, Cy = Cjj, and Cz = Cf in *Soil Taxonomy*.

Whereas permafrost-affected soils are identified as Gelisols (from the Greek word, gelid, meaning very cold) in *Soil Taxonomy*, they are called Cryosols (from the Greek word, kraios, meaning cold or ice) in the other two systems. The Canadian and U.S. systems divide permafrost-affected soils into three categories: organic soils, cryoturbated mineral soils, and other mineral soils. The soils are further delineated on the basis of key properties that link them with soils of lower latitudes. In the WRB, the Cryosol soil group is divided into 16 soil units. Fifteen suffix qualifiers are used to link Cryosols with soils of lower latitudes.

33.6.7 Properties of Gelisols

Gelisols/Cryosols encompass a vast array of soils in terms of chemical and mineralogical properties. However, with the exception of the Histels, Gelisols are uniform in terms of physical and thermal properties.

33.6.7.1 Macromorphology

The most common macroscopic soil features are due to cryoturbation and include irregular or broken horizons and incorporation of organic matter in lower horizons, especially along the top of the permafrost. Oriented stones and displacement of soil materials are common in Gelisols. Freezing and thawing produce granular and platy structures in surface horizons and blocky, prismatic or massive structures in subsurface horizons. The massive structure is due to cryostatic pressure and desiccation that develop when the two freezing fronts, one from the surface and the other from the permafrost, merge during freeze back in the autumn. The perennially frozen layer commonly contains ground ice in the form of segregated ice crystals, vein ice, ice lenses and wedges, and thick ground ice.

The granular, platy, or blocky structures of the surface mineral horizons are also the result of cryopedogenic processes such as the freeze–thaw process and vein-ice formation (ice segregation process). The subsurface horizons often have massive structures and are associated with higher bulk densities, especially in fine-textured soils.

Almost all Gelisols contain ice in the form of crystals, lenses, layers (vein ice), wedges, or massive ground ice, often to a thickness of several meters. Soil texture is one of the factors controlling ice content in mineral soils with fine-textured Gelisols generally having higher ice contents than coarse-textured soils. Coarse-textured soils often have a relatively low ice content; however, they may contain ice wedges in the form of polygons. Histels have an ice content of 60%–90% on a volume basis (Zoltai and Tarnocai, 1975).

The active layer that is subject to annual thawing and refreezing lies above the permafrost and not only supports biological life, but also protects the underlying permafrost. The thickness of the active layer is controlled by soil texture and moisture, thickness of the surface organic layer, vegetation cover, aspect, and latitude.

Dilitancy, often confused with thixotropy, is common in soils with high silt content, greatly affects the trafficability of the soil, and is frequently present during the thaw period. Dilitancy is a property of granular masses expanding due to the increase of space between rigid particles upon displacement of the particles. When dilitant soils dry out, a characteristic vesicular porosity develops. Salt crusts, patches, and pans are common in polar desert soils of the high arctic and cold desert soils of Antarctica (Tedrow, 1977; Bockheim, 1997).

33.6.7.2 Micromorphology

When viewed in thin sections, Gelisols contain a variety of fabrics resulting from desiccation and displacement due to alignment, rotation, sorting, and inclusions (Figure 33.33). These features are accompanied by planar voids, cracks, and vesicles during pore formation. More specifically, the fabric of Gelisols varies from granular (granic and granoidic) in the surface horizons to mainly porphyroskelic with fragmic and fragmoidic components in subsurface mineral horizons (Pawluk and Brewer, 1975; Fox, 1985; Smith et al., 1991; Schaefer et al., 2008). The micromorphology of Gelisols can show evidence of matrix displacement and movement, with resultant reorganization of skeleton grains into circular or elliptical patterns, producing an orbicular fabric (Fox and Protz, 1981). Ice lensing and vein ice development lead to the formation of lenticular fabrics, while cryoturbation and cryodesiccation can lead to granic or granoidic fabrics (Smith et al., 1991). In addition, suscitic and conglomeric fabrics also are common in Gelisols (Fox and Protz, 1981).

33.6.7.3 Thermal Characteristics

Soil temperatures for a Gelisol located north of the arctic treeline in Canada are shown in Figure 33.35. The unique thermal signature that separates Gelisols from all other soils is the presence of a perennially frozen layer, usually below 50 cm. Because of this frozen layer, Gelisols have a steep vertical temperature gradient. If these soils are associated with certain types of patterned ground, the horizontal temperature gradient can also be large. For example, in the case of Gelisols associated with earth hummocks, the soil temperature at the center of the hummock can decrease from 12°C at the surface to 0°C at 50 cm during the summer months. Soil temperatures at comparable depths under the interhummock depression, <1 m away and at equivalent depths, can be 5°C–7°C lower (Tarnocai and Zoltai, 1978).

33.6.7.4 Physical and Chemical Characteristics

Histels contain weakly to well-decomposed moss, sedge, woody, and amorphous organic material. The pH of these soils ranges from 2.5 to 7.0 (Table 33.28). Histels derived from moss peats (Fibristels) tend to have lower pH values than other organic soils with permafrost. Turbels and Orthels developed from calcareous parent materials have a pH_{CaCl_2} of 7 and a high base saturation. On the other hand, where these soils are developed from acidic parent materials, $pH_{CaCl_2} < 5.5$.
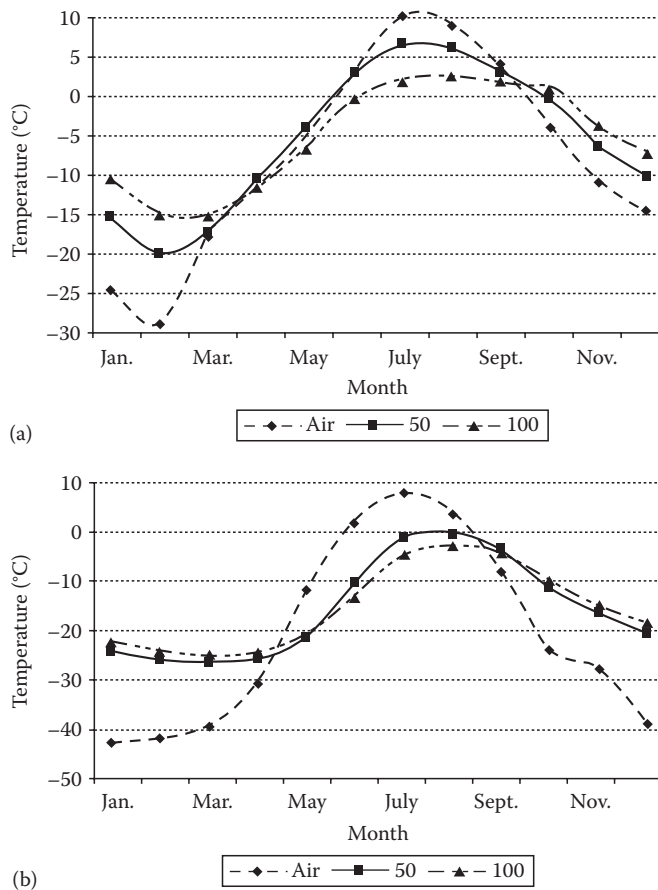
FIGURE 33.35 Soil temperature variation in Gelisols from (a) Auyuittuq National Park, Baffin Island, Canada (66°23′30″N, 65°29′20″W) and (b) Lake Hazen, Quttinirpaaq National Park, Ellesmere Island, Canada (81°49′15″N, 71°33′17″W) during 1999. (From Tarnocai, C. 2008. Arctic permafrost soils, p. 3–16. *In* R. Margesis (ed.) Permafrost soils. Soil Biology Series. Vol. 16. Springer-Verlag, Berlin, Heidelberg, Germany. With permission.)

Anhyturbels and Anhyorthels in polar and cold deserts often have high salt contents and electrical conductivities. Although these soils have many features common to Aridisols (Section 33.10), they have patterned ground, dry permafrost in the upper meter over ice-cemented permafrost, and show evidence of thermal cracking (Bockheim, 1997).

Gelisols, especially the Histels and Turbels, often contain large amounts of organic C in the upper m, with values commonly ranging from 3 kg C m⁻³ in Haplorthels to over 100 kg C m⁻³ in Sapristels and averaging about 50 kg C m⁻³ (Bockheim et al., 1999, 2004; Tarnocai et al., 2007). This is similar to Histosols of other regions (Section 33.2). Less than half of the organic C in Turbels is in the active layer, with the remainder in the upper 30–50 cm of permafrost (Bockheim et al., 1997). Much of this organic carbon exists below 1 m, the depth to which carbon stores are normally reported.

The particle size distribution of Gelisols varies from clayey to coarse gravelly sand. The composition of the fine-earth fraction is commonly dependent on the composition and age of the parent materials.

33.6.8 Special Problems in Managing Gelisols

Gelisols present special problems in terms of management, not only because of frost churning, heaving, sorting, and cracking, but also because melting of segregated ice following a disturbance to the thermal regime leads to subsidence, or thermokarst. To preserve the integrity of structures (buildings, roads, and pipelines) in permafrost soil, it is important to maintain the negative thermal balance of the soil. This is achieved by using special construction methods. For agricultural development, it is important to determine the ice content of the soil; otherwise, after clearing of the land or within a few years after cultivation begins, severe subsidence and thermokarst can develop (Péwé, 1982).

33.6.9 Global Warming and Trace Gas Emissions

General circulation models (GCMs) predict that with a projected twofold increase in atmospheric CO_2 by the year 2050, the mean air temperature of the Earth's surface could increase by 1.5°C–4.5°C (Maxwell and Barrie, 1989); however, warming at the high northern latitudes could be on the order of 4°C–5°C. Indeed, sea ice variations over the past several decades are compatible with a distinct warming of air temperatures in the arctic, especially during the winter and spring (Arctic Climate Impact Assessment, 2004). In addition, permafrost temperatures in northernmost Alaska have increased by 2°C–4°C during the last few decades (Lachenbruch and Marshall, 1986).

As mentioned previously, Gelisols are large C sinks. There is concern that warming in the circumpolar regions could increase the thickness of the active layer and enhance heterotrophic respiration, releasing additional CO_2 to the atmosphere (Oechel and Billings, 1992; Shaver et al., 1992; Waelbroeck et al., 1997), and Gelisols would become a C source. However, Bockheim (2007) has pointed out that continued warming in the arctic could accelerate cryoturbation and enable some soils to store more organic C at depth than at present, thereby mitigating some of the loss of CO_2 to the atmosphere from increased soil respiration.

33.6.10 Human-Caused Disturbances

Arctic regions contain vast energy reserves, including fossil fuels (coal, oil, and gas), biomass, and hydropower. The extraction of fossil fuels and minerals and deforestation may have dramatic long-term effects on arctic ecosystems as they have a low resistance and resilience to disturbance (Reynolds and Tenhunen, 1996). As the world's population continues to expand, there will be increased pressure for development at the high latitudes. For example, Siberia already contains several cities that were developed in permafrost and have in excess of 500,000 inhabitants.

The circumarctic contains numerous indigenous people who are dependent on the terrestrial and marine ecosystems for

TABLE 33.28 Analytical Characteristics of Selected Gelisols

Horizon	Depth (cm)	pH	Bulk Density (mg m⁻³)	C	N	Clay	Silt	Sand	Extr. P (mg L⁻¹)	Ca	Mg	Na	K	Exch. Acid	Exch. Al	CEC NH₄OAc	Base Sat. (%)
							(%)					(cmol_c kg⁻¹)					
colspan Typic Haploturbel; Dry Acidic Tundra (Empetrum–Betula–Dryas–Arctostaphylos); Till/Congelifractate; 68°37′N, 149°18′W																
Ajj	2–46	4.91	0.56	8.68	0.53	9.8	53.2	37.0	0.7	2.4	1.2	TR	0.1	39.4	11.5	36.1	10
Bwjj	4–40	4.83	0.87	2.3	0.13	18.4	46.2	35.4	1.2	0.7	0.8	TR	TR	25.2	9.1	20.5	8
BCjj	0–34	5.14	0.68	9.31	0.26	11.0	55.3	33.7	3.5	0.7	0.4	0.1	0.1	34.5	9.7	2.6	5
C	8–90	6.94	ND	0.58	0.05	5.2	44.4	50.4	5.4	7	1.2	TR	0.1	3.1	0	10.8	77
Typic Molliturbel; Dry Nonacidic Tundra (Carex–Dryas–Tomentypnum); Loess/Outwash; 70°11′N, 149°17′W																
A	1–21	6.93	ND	5.66	0.37	14.0	26.8	59.1	0.1	61.3	2.6	TR	0.1	11.2	0	69.3	92
Oa/Ckjj	21–42	7.55	ND	15	0.96	16.0	16.1	67.9	0.5	47.8	1.2	0.1	0.1	3.6	0	30.3	100
2Ck	42–61	7.88	ND	2.37	0.04	5.7	6.1	88.2	0.1	35.4	0.8	0	TR	0	0	3.3	100
Ruptic–Histic Aquiturbel; Moist Nonacidic Tundra (Dryas–Salix–Eriophorum); Loess/Outwash; 69°27′N, 148°40′W																
Oi	0–8	8.37	0.12	35.7	1.25				15.2	84.1	15.4	0.2	1.5	16	0.1	113	89
Oejj1	0–40	7.81	0.20	24.2	1.25				2.1	84.3	6.8	0.3	0.2	11.6	0.1	102	89
Bw	0–12	8.01	1.07	3.55	0.24	19.8	70.0	10.2	0.0	47.1	3.1	TR	0.1	0.8	0	22.1	100
Bg	0–88	7.92	1.19	3.8	0.29	23.5	66.9	9.6	0.5	36.3	2	TR	0.2	3	0	24.3	100
Oejj2	0–50	7.55	0.43	20.12	1.29				1.1	81.1	5.1	0.1	0.3	13.7	0.1	97.8	89
2Ajj	0–85	7.73	1.09	7.27	0.39	16.7	69.3	14.0	1.8	55.1	2.3	TR	0.2	3.7	0	29.5	100
2Cd/Oejjfm	65–85	7.40	ND	10.93	0.74	18.1	70.1	11.8	3.0	69.5	3.8	TR	0.2	7.4	0	50.8	100
Typic Umbrorthel; Moist Acidic Prostrate Shrub–Grass Tundra (Salix–Carex–Polytrichum); Residuum; 68°46′N, 149°35′W																
A	0–8	4.21	0.57	7.13	0.48	3.8	15.6	80.6	10.3	1.9	1.2	0	0.3	21.4	3.3	18.1	19
Bw	8–22	4.05	1.30	2.05	0.12	5.2	15.9	78.9	1.8	0.5	0	TR	TR	13.9	3.9	9.2	7
BC	22–31	4.64	1.38	0.44	0.05	1.2	11.9	86.9	7.5	0.2	0	0	TR	4.6	1.5	3	7
Typic Histoturbel; Moist Acidic Tussock Tundra (Eriophorum–Sphagnum–Betula); Colluvium; 68°37′N, 149°19′W																
Oi	0–9	4.46	0.04	44.34	0.85				34.0	11.5	6.3	0.8	3.1	110	2.5	136	16
Oe1	9–17	5.06	0.10	39.8	1.40				2.4	9.3	2.4	0.3	0.4	93.5		104	12
Oe2	17–23	5.37	0.37	24.3	1.34				1.2	3.9	1.6	0.1	0.3	63.2		59.9	10
Bg	23–40	5.17	1.46	3.04	0.11	20.0	41.9	38.1	0.4	0.5	0.4	TR	TR	14.7	4.2	15.8	6
BCg	40–48	5.70	1.09	3.46	0.14	20.5	39.8	39.7	0.6	0.5	0.4		TR	15	3.9	14.7	6
Cg/Oefm	48–80	6.39	ND	4.24	0.19	19.0	38.9	42.1	1.5	0.7	0.4	TR	TR	16	3.3	14	9
Typic Fibristel; Wet Acidic Tundra (Eriophorum–Sphagnum); Organic Basin Deposits; 68°37′N, 149°19′W																
Oi1	0–13	4.48	0.03	43.56	0.65				10.4	1.9	0.8	0.1	0.3	57.2	4.1	109	3
Oi2	13–33	4.52	0.19	44.82	2.58				4.0	8	2.6	0.1	0.2	107	4.8	77.1	14
Oe	33–44	4.66	0.35	25.2	1.17				0.6	3.4	0.7	TR	TR	75.2	6.9	51.3	8
Oa	44–62	4.99	0.52	18.5	1.02				0.5	3.6	0.7	TR	TR	58.2	9	46.6	9

Source: Soil Survey Staff. 2011. National Cooperative Soil Characterization Database. Available Online at http://ssldata.nrcs.usda.gov. Accessed July 8, 2011.

food, fuel, and shelter. Special efforts must be made to ensure that these ecosystems remain sustainable. The polar regions are less diverse biologically than other life zones. However, ancient (~2.5 million years old) permafrost contains viable microorganisms that may give clues to the evolution of microbes (Gilichinsky, 1993).

References

Agriculture Canada Expert Committee on Soil Survey. 1987. The Canadian system of soil classification. 2nd edn. Research Branch, Agriculture Canada Publication 1646, Ottawa, Canada.

Arctic Climate Impact Assessment. 2004. Impacts of a warming arctic—Arctic climate impact assessment. Cambridge University Press, Cambridge, U.K.

Bockheim, J.G. 1997. Properties and classification of cold desert soils from Antarctica. Soil Sci. Soc. Am. J. 61:224–231.

Bockheim, J.G. 2007. Importance of cryoturbation in redistributing organic carbon in permafrost-affected soils. Soil Sci. Soc. Am. J. 71:1335–1342.

Bockheim, J.G., L.R. Everett, K.M. Hinkel, F.E. Nelson, and J. Brown. 1999. Soil organic carbon storage and distribution in arctic tundra, Barrow, Alaska. Soil Sci. Soc. Am. J. 63:934–940.

Bockheim, J.G., and K.M. Hinkel. 2007. The importance of "deep" organic carbon in permafrost-affected soils of arctic Alaska. Soil Sci. Soc. Am. J. 71:1889–1892.

Bockheim, J.G., K.M. Hinkel, W.R. Eisner, and X.Y. Dai. 2004. Carbon pools and accumulation rates in an age-series of soils in drained thaw-lake basins, arctic Alaska. Soil Sci. Soc. Am. J. 68:697–704.

Bockheim, J.G., G. Mazhitova, J.M. Kimble, and C. Tarnocai. 2006. Controversies on the genesis and classification of permafrost-affected soils. Geoderma 137:33–39.

Bockheim, J.G., C.L. Ping, J.P. Moore, and J.M. Kimble. 1994. Gelisols: A new proposed order for permafrost affected soils, p. 26–44. *In* J.M. Kimble and R.J. Ahrens (eds.) Proc. Meet. Classif. Correl Manage. Permafrost-Affected Soils. USDA-SCS, National Soil Survey Center, Lincoln, NE.

Bockheim, J.G., D.A. Walker, and L.R. Everett. 1997. Soil carbon distribution in nonacidic and acidic tundra of artic Alaska, p. 143–155. *In* R. Lal et al. (eds.) Soil processes and the carbon cycle. CRC Press, Boca Raton, FL.

Brown, J., O.J. Ferrians, Jr., J.A. Heginbottom, and E.S. Melnikov. 1997. Circumarctic map of permafrost and ground ice conditions, 1:10 000 000 scale map. U.S. Geological Survey International Permafrost Association. U.S. Geological Survey, Washington, DC.

Douglas, L.A., and J.C.F. Tedrow. 1960. Tundra soils of arctic Alaska. Trans. 7th Int. Congr. Soil Sci. IV:291–304.

Everett, K.R. 1979. Evolution of the soil landscape in the sand region of the Arctic Coastal Plain as exemplified at Atkasook, Alaska. Arctic 32:207–223.

Food and Agricultural Organization of the United Nations. 2006. World reference base for soil resources: A framework for international classification, correlation and communication. World Soil Resources Report 103, p 145, FAO, Rome, Italy.

Fox, C.A. 1985. Micromorphology of an orthic turbic cryosol—A permafrost soil, p. 699–705. *In* P. Bullock and C.P. Murphy (eds.) Soil micromorphology, soil genesis. Vol. 2. Academic Publishers, Berkhamsted, U.K.

Fox, C.A. 1994. Micromorphology of permafrost-affected soils, p. 51–62. *In* J.M. Kimble and R.J. Ahrens (eds.) Proc. Meet. Classif. Correl. Manage. Permafrost-Affected Soils. USDA-SCS, National Soil Survey Center, Lincoln, NE.

Fox, C.A., and R. Protz. 1981. Definition of fabric distributions to characterize the rearrangement of soil particles in the turbic cryosols. Can. J. Soil Sci. 61:29–34.

Gilichinsky, D.A. 1993. Viable microorganisms in permafrost: The spectrum of possible applications to new investigations, p. 268–270. *In* D.A. Gilichinsky (ed.) Proc. 1st Int. Conf. Cryopedology. Russian Academy of Science, Pushchino, Russia.

Kimble, J.M., C. Tarnocai, C.L. Ping, R. Ahrens, C.A.S. Smith, J. Moore, and W. Lynn. 1993. Determination of the amount of carbon in highly cryoturbated soils, p. 277–291. *In* D.A. Gilichinsky (ed.) Post Seminar Proc. Joint Russian–American Seminar Cryopedo. Global Change. Russian Academy of Science, Pushchino, Russia.

Lachenbruch, A.H., and B.V. Marshall. 1986. Changing climate: Geothermal evidence from permafrost in the Alaskan Arctic. Science 234:689–696.

Marion, G.R. 1995. Freeze–thaw processes and soil chemistry. U.S. Army Corps of Engineers, Cold Regions Research and Engineering Laboratory Special Report 95-12, pp 29, Hanover, NH.

Maxwell, J.B., and L.A. Barrie. 1989. Atmospheric and climate change in the Arctic and Antarctic. Ambio 18:42–49.

Oechel, W.C., and W.D. Billings. 1992. Effects of global change on the carbon balance of arctic plants and ecosystems, p. 139–168. *In* F.S. Chapin, III, R.L. Jeffries, J.F. Reynolds, G.R. Shaver, and J. Svoboda (eds.) Arctic ecosystems in a changing climate. Academic Press, San Diego, CA.

Pawluk, S., and R. Brewer. 1975. Micromorphological and analytical characteristics of some soils from Devon and King Christian Islands, N.W.T. Can. J. Soil Sci. 55:349–361.

Pettapiece, W.W. 1975. Soils of the subarctic in the lower Mackenzie basin. Arctic 28:35–53.

Péwé, T.L. 1982. Geological hazards of the Fairbanks area, Alaska. Alaska Geological Survey. Special Report 15. Alaska Division of Geological and Geophysical Surveys, Fairbanks, AK.

Reynolds, J.F., and J.D. Tenhunen. 1996. Ecosystem response, resistance, resilience, and recovery in arctic landscapes: Progress and prospects, p. 419–428. *In* J.F. Reynolds and J.D. Tenhunen (eds.) Landscape function and disturbance in Arctic Tundra. Ecological studies. Vol. 120. Springer-Verlag, Berlin, Germany.

Schaefer, C.E.G.R., F.N.B. Simas, R.J. Gilkes, C. Mathison, L.M. de Costa, and M.A. Albuquerque. 2008. Micromorphology and microchemistry of selected Cryosols from maritime Antarctica. Geoderma 144:104–115.

Shaver, G.R., W.D. Billings, F.S. Chapin, A.E. Giblin, K.J. Nadelhoffer, W.C. Oechel, and E.B. Rastetter. 1992. Global change and the carbon balance of arctic ecosystems. Bioscience 42:433–441.

Simas, F.N.B., C.E.G.R. Schaefer, M.R. Filho Albuquerque, M.R. Francelino, E.I. Filho Fernandes, and L.M. de Costa. 2008. Genesis, properties and classification of Cryosols from Admiralty Bay, maritime Antarctica. Geoderma 144:116–122.

Smith, C.A.S., C.A. Fox, and A.E. Hargrave. 1991. Development of soil structure in some Turbic Cryosols in the Canadian low arctic. Can. J. Soil Sci. 71:11–29.

Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd edn. USDA-SCS, agriculture handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. Soil Survey Division, National Resources Conservation Service, Lincoln, NE.

Soil Survey Staff. 2011. National Cooperative Soil Characterization Database. Available Online at http://ssldata.nrcs.usda.gov. Accessed July 8, 2011.

Tarnocai, C. 1972. Some characteristics of cryic organic soils in northern Manitoba. Can. J. Soil Sci. 52:485–496.

Tarnocai, C. 1973. Soils of the Mackenzie River area. Report No. 73-26, Task Force on Northern Oil Development QS-1528-000-EE-A1. Information Canada, Winnipeg, Canada.

Tarnocai, C. 2008. Arctic permafrost soils, p. 3–16. *In* R. Margesis (ed.) Permafrost soils. Soil Biology Series. Vol. 16. Springer-Verlag, Berlin, Heidelberg, Germany.

Tarnocai, C., and C.A.S. Smith. 1992. The formation and properties of soils in the permafrost regions of Canada, p. 21–42. *In* D.A. Gilichinsky (ed.) Proc. 1st Int. Conf. Cryopedol. Russian Academy of Science, Pushchino, Russia.

Tarnocai, C., D. Swanson, J. Kimble, and G. Broll. 2007. Northern circumpolar carbon database. Digital Database, Research Branch, Agriculture and Agri-Food Canada, Ottawa, Canada.

Tarnocai, C., and S.C. Zoltai. 1978. Earth hummocks of the Canadian arctic and subarctic. Arct. Alp. Res. 10:581–594.

Tedrow, J.C.F. 1970. Soil investigations of Inglefield Land, Greenland. Meddelelser om Grönland 188:1–93.

Tedrow, J.C.F. 1977. Soils of the polar landscapes. Rutgers University Press, New Brunswick, NJ.

Tedrow, J.C.F., P.F. Bruggeman, and G.F. Walton. 1968. Soils of Prince Patrick Island. Research Paper 44. Arctic Institute of North America, Washington, DC.

Waelbroeck, C., P. Monfray, W.C. Oechel, S. Hastings, and G. Vourlitis. 1997. The impact of permafrost thawing on the carbon dynamics of tundra. Geophys. Res. Lett. 24:229–232.

Washburn, A.L. 1980. Geocryology. Wiley & Sons, Halstead Press, New York.

Zoltai, S.C., and C. Tarnocai. 1971. Properties of a wooded palsa in northern Manitoba. Arct. Alp. Res. 3:115–129.

Zoltai, S.C., and C. Tarnocai. 1974. Soils and vegetation of hummocky terrain. Report No. 74-5. Task Force on Northern Oil Development, QS-1552-000-EEE-A1. Information Canada, Winnipeg, Canada.

Zoltai, S.C., and C. Tarnocai. 1975. Perennially frozen peatlands in the western arctic and subarctic of Canada. Can. J. Earth Sci. 12:28–43.

33.7 Vertisols

Randal J. Southard

Steven G. Driese

Lee C. Nordt

33.7.1 Introduction

Vertisols are clayey soils that exhibit significant volume change due to shrink and swell processes as the soil dries and wets. Most Vertisols have smectitic mineralogy, but Vertisols with mixed or kaolinitic mineralogy also occur. The intrinsic shrink/swell behavior is the dominant soil-forming process of Vertisols and results in significant close-range spatial and temporal variability of soil properties, possibly more so than in any other soil order. As a result, Vertisols are often difficult to manage for agricultural, natural resource, and engineering applications. A substantial body of information about Vertisols exists in the literature, including summaries (e.g., Coulombe et al., 2000; Blokhuis, 2002), and more in-depth discussions about specific aspects of Vertisols (e.g., Ahmad, 1983; Probert et al., 1987; Wilding and Puentes, 1988; Kimble, 1991; Ahmad and Mermut, 1996; Coulombe et al., 1996b).

33.7.2 Distribution and Formation

Geographically, Vertisols occur in more than 100 countries and occupy 309 Mha (10^6 ha), which corresponds to about 2% of the ice-free land area (Table 33.1; Figure 33.36; USDA-NRCS, 2009). Globally, Vertisols are most extensive in India (80 Mha), Australia (70 Mha), and Sudan (50 Mha), followed by the United States (18 Mha), Ethiopia (13 Mha), and China (12 Mha) (Dudal and Eswaran, 1988; Isbell, 1991; Blokhuis, 2002).

In the United States, Vertisols are reported in 25 states and territories (Coulombe et al., 1996b, 2000), primarily west of the Mississippi River and in the Black Belt region of the southern states. Vertisols are most extensive in Texas (6.5 Mha), South Dakota (1.5 Mha), California (1 Mha), and Montana (0.6 Mha). The remaining states each have less than 0.25 Mha of Vertisols and vertic intergrades.

Vertisols are formed from a wide variety of parent materials of varying age and under a wide range of climatic conditions (Coulombe et al., 1996a, 1996b). A common attribute is a soil-forming environment that favors preservation of silica and basic cations and that produces episodic or periodic wetting and drying to allow cracks to form. Most Vertisols occur
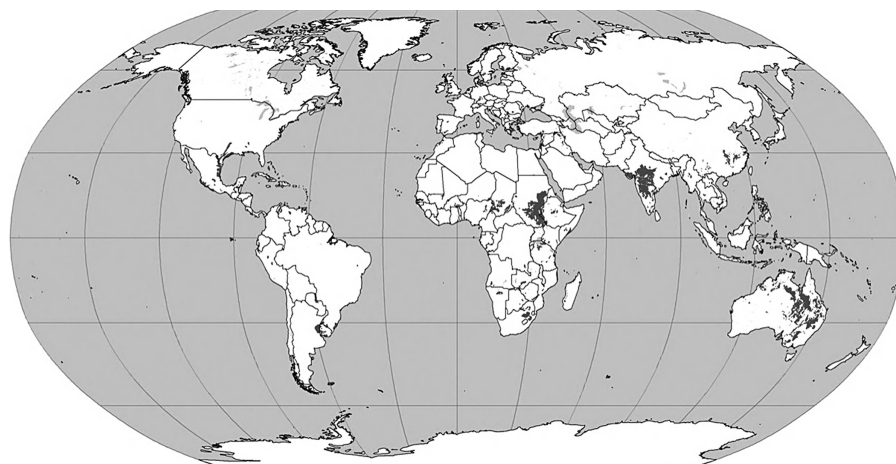
FIGURE 33.36 Global distribution of Vertisols. (Courtesy of USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC, 1998.)

on shallow slopes in uplands and on a variety of alluvial landforms. Natural vegetation of Vertisols is typically grassland and savanna, probably because most woody plant roots cannot survive shearing during wetting and drying, but mixed/deciduous forested and scrub/shrub can occur, depending on the ecological succession of the vegetation and human impact.

33.7.3 Morphology

33.7.3.1 Color

In general, Vertisols are dark in color, particularly in the upper horizons. Typically, they exhibit a moist Munsell value of 3 or less and chroma of 2 or less in the matrix, and have hues of 7.5YR or yellower. The dark colors are generally attributed to the close association of organic matter with clay minerals and to various compounds of Mn, Fe, and Ti (Ahmad and Mermut, 1996). Vertisols with black, gray, brown, or red pigmentation have been reported in various regions of the world. In the most recent *Keys to Soil Taxonomy* (Soil Survey Staff, 2010), dark Vertisols are considered the norm; soil color differentiates Chromic subgroups, which require a value of 4 or more (moist), or 6 or more (dry), or a chroma of 3 or more. The chroma requirement does not apply to

the subgroup of Aquerts. Lighter soil colors are usually associated with the accumulation of calcium carbonate, gypsum, or more soluble salts, especially in subsurface horizons. Reddish redox concentrations can occur in poorly drained Vertisols; reddish matrix colors and high chroma are often associated with Vertisols on better drained landscapes, and in less humid climates, and possibly in Vertisols formed from parent material with low Mn content.

33.7.3.2 Texture

Vertisols are fine-textured soils. In general, Vertisols must have at least 30% clay in the fine-earth fraction in the upper 50 cm or to a densic, lithic, or paralithic contact, a duripan, or a petrocalcic horizon if any of those features is shallower (Soil Survey Staff, 2010). Clay content of Vertisols can be as high as 90%, particularly for those derived from pyroclastic deposits. Soils that do not meet the clay content requirements, or that do not exhibit other required properties (slickensides or wedge-shaped peds, and cracks, described below), but that exhibit considerable shrink/swell activity as measured by the coefficient of linear extensibility (COLE), often are classified as vertic intergrades to other soil orders. Particle size distributions for selected Vertisols are shown in Table 33.29.

TABLE 33.29 Particle Size Distribution of Selected Vertisols Formed from a Wide Variety of Parent Materials

Regions	% Sand 2.00–0.50 mm	% Silt 0.50–0.002 mm	% Clay <0.002 mm	References
Australia	16.0 (2.5)	33.6 (3.7)	50.3 (5.5)	Boettinger (1992)
India	13.8 (3.7)	21.9 (2.6)	61.0 (5.2)	Hirekurubar et al. (1991)
Sudan	16.8 (8.4)	19.9 (10.1)	63.1 (9.7)	Blokhuis (1993)
Texas	9.6 (7.3)	36.3 (6.9)	53.5 (3.7)	Kunze et al. (1963), Yule and Ritchie (1980a), and Hallmark et al. (1986)
West Africa	18.4 (11.6)	27.7 (8.0)	54.0 (15.4)	Beavington (1978) and Yerima et al. (1988)
California	3.4 (0.4)	34.6 (1.2)	62.0 (1.2)	NSSL Pedon No. 94P0056
El Salvador	10.8 (4.0)	31.2 (7.3)	58.1 (8.5)	Yerima et al. (1985)
West Indies	17.5 (9.7)	10.6 (3.0)	74.9 (11.4)	Ahmad and Jones (1969)
Uruguay	26.3 (5.6)	27.0 (8.1)	46.7 (13.3)	Lugo-Lopez et al. (1985)

Data are reported as mean (sd).

33.7.3.3 Differentiating Physical Properties and Features

In addition to the clay content requirements, Vertisols must exhibit significant shrink/swell characteristics to develop diagnostic soil structure and features in the soil profile. Evidence of shrink/swell behavior includes slickensides or wedge-shaped aggregates with the long axis inclined at 10°–60° from the horizontal (Soil Survey Staff, 2010). Slickensides are more or less planar features that exhibit shiny and grooved surfaces at the interface of the peds (Figure 33.37) that indicate soil movement and shearing. The presence of slickensides in a soil horizon is typically indicated by the morphological horizon designation, Bss. Wedge-shaped aggregates are soil structural units that generally result from limited formation of slickensides bounding ped surfaces (Figure 33.38). The presence of cracks (Figure 33.39) that open and close in response to drying and wetting is also a requirement for Vertisols (Soil Survey Staff, 2010).

Gilgai and diapir are other features that may be present in Vertisols. Gilgai (Figure 33.40) is an Australian aboriginal term used to describe microtopography (Figures 33.41 and 33.42) that
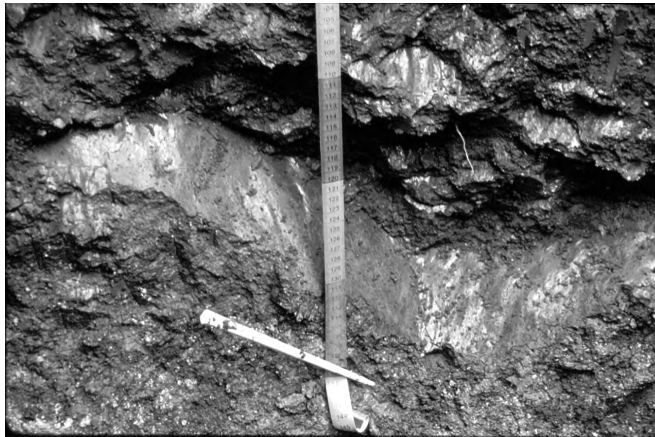
FIGURE 33.37 The glossy slickensides show evidence of subsoil shrink-swell and shearing. (Photograph from Eswaran et al., 1999).
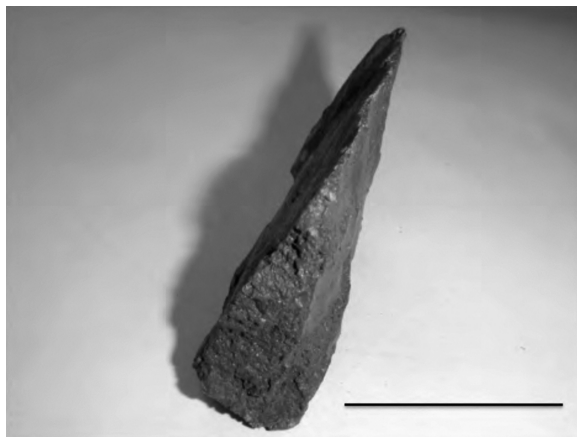
FIGURE 33.38 Wedge-shaped peds are a result of subsoil shearing in many Vertisols. The bar represents 5 cm. (Photograph courtesy of Susan B. Southard, USDA-NRCS).

FIGURE 33.39 Cracks that open and close periodically are a required morphological feature of Vertisols. These cracks range from about 1 to 6 cm wide. (Photograph from Eswaran et al., 1999).

FIGURE 33.40 Vertisol microtopography (gilgai) is especially evident during the wet season. (Photograph from Eswaran et al. 1999).

FIGURE 33.41 Cross-section of a Vertisol exposed in a trench. Darker-colored bowls, or microlows, are evident on the left and right sides of the exposure. A lighter-colored diapir, or microhigh, is visible between the bowls to the right of the tape measure. Strings have been used to delineate regions of distinctive morphology. The tape measure is 180 cm long (Photograph from Eswaran et al., 1999).
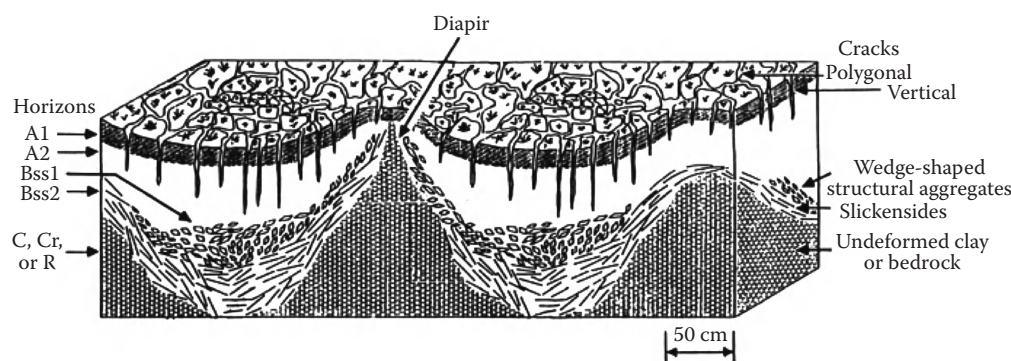
FIGURE 33.42 Schematic representation of Vertisol morphology and horizon arrangement. (From Coulombe, C.E., L.P. Wilding, and J.B. Dixon. 1996b. Overview of Vertisols: Characteristics and impacts on society. Adv. Agron. 57:289–375. With permission.)

consists of mounds (microhigh), shelves (intermediate), and depressions (microlow), either in circular, linear, or complex patterns. Prior to 1992, *Soil Taxonomy* used gilgai as a differentiating characteristic of Vertisols, but, that criterion was eliminated because not all Vertisol landscapes have gilgai. Diapir, also called mukkara in Australia, in this context was adopted from geology terminology, where the term is used to describe large-scale intrusions of ductile material (often mud or salt) into more rigid, surrounding rock. In Vertisols, a diapir is a smaller-scale intrusion of soil material that penetrates from the subsoil through the upper layers and can be identified by a contrast in color and/or texture. Diapirs may occur independent of gilgai; however, when gilgai are present, the tip of the diapir coincides with the mound or microhigh.

33.7.4 Genesis

The dominant processes that lead to the characteristic properties of Vertisols is shrinking and swelling induced by changes in soil water conditions. The shrink/swell action can be observed from macroscopic to submicroscopic scales of resolution (Figure 33.43). The shrink/swell phenomena are clearly related to the changes in the interlayer spacing of expansible phyllosilicate minerals, principally smectites, by sorption and desorption of interlayer water. However, changes in the interlayer spacing of smectites contribute only 10%–30% of the change in volume when saturated with Ca (Greene-Kelly, 1974; Tessier, 1984). Tessier (1984) reported that the type of microstructure, for example, crystallite, domains, or quasicrystals (Figure 33.43) and the surface area of the clay mineral phases govern the shrink/swell potential. The microstructure and the porosity accommodate the changes in water content and potential of the clay-water systems. Other factors affecting shrink/swell phenomenon include particle rigidity, stress history, predominance of cations present on the exchange sites (e.g., Na vs. Ca), the electrolyte concentration of the soil solution, and the amount and location of the charge in the clay mineral crystallographic structure, which may cause variation in the structural organization and the interlayer distance among smectites.

Tessier (1984), Quirk (1994), and Coulombe et al. (1996a, 1996b) extensively reviewed the models and factors involved in shrink/

swell processes. According to Coulombe et al. (1996b), the application of the Coulomb–Mohr theory of shear failure constitutes the best model to account for the formation and occurrence of features and structure in Vertisols. Shear failure is more likely to occur under moist or wet than dry conditions, since a lower normal stress is required to promote shear failure. Even though shrink/swell processes are a common phenomenon among Vertisols, other pedogenic processes also occur, such as accumulation of salts (e.g., salic, natric, calcic, and gypsic great groups), silica (duric great group), and depth of saturation by water (epi and endo great groups).

33.7.5 Classification

Vertisol is the widely accepted term to designate fine-textured soils with extensive shrink/swell behavior. The term originates from *vertere* (L.), meaning to churn or turn over and *sol* for soils (Dudal and Eswaran, 1988). Coulombe et al. (1996b) discussed the history and evolution of the Vertisol order in the U.S. soil classification systems. Vertisols were first introduced as a soil order in the *7th Approximation* (Soil Survey Staff, 1960), and differentiating criteria were subsequently modified in *Soil Taxonomy* (Soil Survey Staff, 1975, 1999). Prior to the *7th Approximation*, most Vertisols were identified as Grumosols, Regosols, Rendzina, or Alluvial soils (Soil Survey Staff, 1975). In 1960, Vertisols were the first order in the key and were differentiated on the basis of clay content, cation exchange capacity, cracks, and evidence of shrink–swell in the form of gilgai, slickensides, or wedge-shaped aggregates. Two suborders, Aquerts and Usterts, were identified, and an important property for identifying great groups was the presence of a self-mulching surface horizon (Figure 33.44; "Grumic" great groups, e.g., Grumaquerts) or of a platy or massive surface crust ("Mazic" great groups, e.g., Mazusterts). In the first edition of *Soil Taxonomy* (Soil Survey Staff, 1975), Vertisols were placed as the fourth order in the key, subgroups keyed in the sequence: Xererts, Torrerts, Uderts, and Usterts (no Aquerts), and great groups were identified on the basis of soil color, not grumic or mazic properties (e.g., Chromoxererts, with moist soil chroma ≥1.5 in the upper 30 cm, otherwise Pelloxererts). Revisions based on the efforts of ICOMERT (the International Committee on the Classification of Vertisols) added the Aquerts
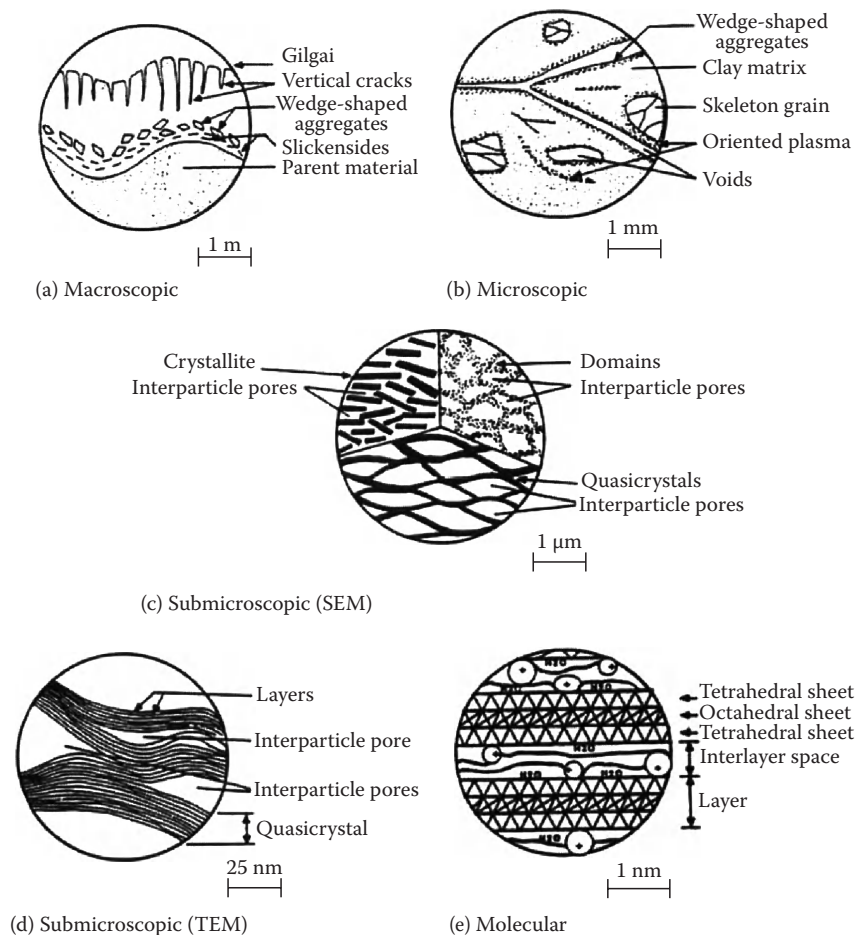
FIGURE 33.43 Shrink-swell features depicted at various scales of resolution. (From Coulombe, C.E., L.P. Wilding, and J.B. Dixon. 1996b. Overview of Vertisols: Characteristics and impacts on society. Adv. Agron. 57:289–375. With permission.)
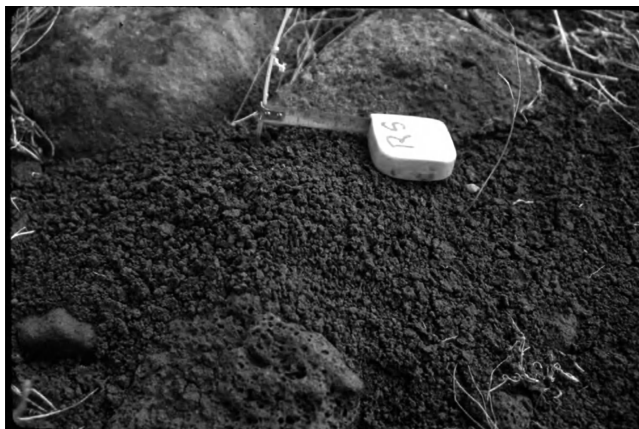
FIGURE 33.44 Vertisol surface mulch of very fine and fine angular blocks and granules.

and Cryerts suborders and included some of the subsurface diagnostic horizons as criteria for differentiation of great groups (Soil Survey Staff, 1999).

The Vertisol order currently keys out as the sixth order, based on the following (Soil Survey Staff, 2010):

Other soils that have:

1) A layer 25 cm or more thick, within 100 cm of the mineral soil surface, that has *either* slickensides *or* wedge-shaped peds that have their long axes tilted 10 to 60 degrees from the horizontal; *and* 2) A weighted average of 30 percent or more clay in the fine-earth fraction either between the mineral soil surface and a depth of 18 cm or in an Ap horizon, whichever is thicker, *and* 30 percent or more clay in the fine-earth fraction of all horizons between a depth of 18 cm and either a depth of 50 cm or a densic, lithic, or paralithic contact, a duripan, or a petrocalcic horizon if shallower; *and* 3) Cracks that open and close periodically.

The order is subdivided into 6 suborders, 24 great groups, and 158 subgroups (Soil Survey Staff, 2010; Table 33.30). The suborders are identified on the basis of aquic conditions, cryic soil temperature regime, and the frequency and duration of soil cracking. The cracking criterion allow inferences about the soil moisture regime (as implied by the suborder names), but soil moisture regimes in Vertisols are not well characterized due to cracking and slow permeability when wet, leading to heterogeneous wetting and drying of the soils and preferential bypass flow. Because

TABLE 33.30 Listing of Suborders, Great Groups, and Subgroups for the Vertisol Order

Suborders	Great Groups	Subgroups
Aquerts	Sulfaquerts	Salic, Sulfic, Typic
	Salaquerts	Aridic, Ustic, Leptic, Entic, Chromic, Typic
	Duraquerts	Aridic, Xeric, Ustic, Aeric, Chromic, Typic
	Natraquerts	Typic
	Calciaquerts	Aeric, Typic
	Dystraquerts	Sulfaqueptic, Aridic, Ustic, Aeric, Leptic, Entic, Chromic, Typic
	Epiaquerts	Halic, Sodic, Aridic, Xeric, Ustic, Aeric, Leptic, Entic, Chromic, Typic
	Endoaquerts	Halic, Sodic, Aridic, Xeric, Ustic, Aeric, Leptic, Entic, Chromic, Typic
Cryerts	Humicryerts	Sodic, Typic
	Haplocryerts	Sodic, Chromic, Typic
Xererts	Durixererts	Halic, Sodic, Aquic, Aridic, Udic, Haplic, Chromic, Typic
	Calcixererts	Lithic, Petrocalcic, Aridic, Leptic, Entic, Chromic, Typic
	Haploxererts	Lithic, Halic, Sodic, Aridic, Aquic, Udic, Leptic, Entic, Chromic, Typic
Torrerts	Salitorrerts	Aquic, Leptic, Entic, Chromic, Typic
	Gypsitorrerts	Chromic, Typic
	Calcitorrerts	Petrocalcic, Leptic, Entic, Chromic, Typic
	Haplotorrerts	Halic, Sodic, Leptic, Entic, Chromic, Typic
Usterts	Dystrusterts	Lithic, Aquic, Aridic, Udic, Leptic, Entic, Chromic, Typic
	Salusterts	Lithic, Sodic, Aquic, Aridic, Leptic, Entic, Chromic, Typic
	Gypsiusterts	Lithic, Halic, Sodic, Aridic, Udic, Leptic, Entic, Chromic, Typic
	Calciusterts	Lithic, Halic, Sodic, Petrocalcic, Aridic, Udic, Leptic, Entic, Chromic, Typic
	Haplusterts	Lithic, Halic, Sodic, Petrocalcic, Gypsic, Calcic, Aridic Leptic, Aridic, Leptic Udic, Entic Udic, Chromic Udic, Udic, Leptic, Entic, Chromic, Typic
Uderts	Dystruderts	Aquic, Oxyaquic, Leptic, Entic, Chromic, Typic
	Hapluderts	Lithic, Aquic, Oxyaquic, Leptic, Entic, Chromic, Typic

Source: Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS, National Soil Survey Center, Lincoln, NE.

of low permeability and water conductance via bypass flow along macrovoids, Vertisols rarely saturate in their entirety (Jacob et al., 1997; Nordt and Driese, 2009). The great group categories reflect additional dominant soil processes and resulting diagnostic horizons, and subgroups represent intragrade or intergrade concepts toward other soil orders or categories. Vertisols are not required to have a specific diagnostic horizon, although many have mollic epipedons and cambic horizons.

Vertisols are also recognized in other international soil classification systems such as FAO–UNESCO (1988, 1998) and the World Reference Base (WRB) for Soil Resources (IUSS Working Group WRB, 2007) and the French systems (CPCS, 1987; Baize and Girard, 1996). Some variations in the criteria used to differentiate Vertisols may occur from one classification system to another, but the *Soil Taxonomy* criteria for Vertisols (no vertic horizon) and the following WRB criteria (IUSS Working Group WRB, 2007) for a vertic horizon and for identification of Vertisols are essentially the same.

33.7.5.1 Vertic Horizon

The vertic horizon (from Latin vertere, to turn) is a clayey subsurface horizon that, as a result of shrinking and swelling, has slickensides and wedge-shaped structural aggregates. Diagnostic criteria (1) contains 30% or more clay throughout; (2) has wedge-shaped structural aggregates with a longitudinal axis tilted between 10° and 60° from the horizontal; (3) has slickensides; and (4) has a thickness of 25 cm or more.

Vertisols key out as the sixth Reference Soil Group as follows: Other soils having (1) a vertic horizon starting within 100 cm of the soil surface; and (2) after the upper 20 cm have been mixed, 30% or more clay between the soil surface and the vertic horizon throughout; and (3) cracks that open and close periodically.

33.7.6 Mineralogical Properties

Most Vertisols are dominated by smectitic mineralogy. However, Vertisols containing abundant kaolinite, illite, halloysite, hydroxy-interlayered smectite, or a combination of interstratified clay minerals are also found in various regions of the world (Coulombe et al., 1996a). Recognition of Vertisols with these mineralogical characteristics has caused geoscientists to reconsider the Vertisol paradigm and the mechanisms involved in the shrink/swell processes of soils that are not dominated by smectite. As discussed earlier, shrink/swell behavior includes expansion and collapse of interlayer space, but also involves changes in microstructure and porosity associated with changes in water content and potential (Tessier, 1984; Wilding and Tessier, 1988; Coulombe et al., 1996a).

The smectites that dominate most Vertisols tend to concentrate in the fine clay fraction and contribute to high particle surface area and shrink–swell potential. Characteristics of smectites in Vertisols, which may also apply to other clay minerals and other soil orders, include (1) relatively high thermodynamic stability, (2) high layer charge (>0.45 per half unit cell), which may be as high as vermiculite or illite and may impact the availability of some ions such as K and NH_4, and (3) a high Fe content in some crystallographic structures such as nontronite or beidellite, particularly when derived from basaltic parent materials containing abundant ferromagnesian minerals (Coulombe et al., 1996a). A CEC/clay ratio ≥0.60 is typical of soils in the smectitic mineralogy class at the family taxonomic level (a CEC/clay ratio ≥0.60 identifies the "superactive" CEC activity family class [Soil Survey Staff 2010], which suggests that the clay fraction is dominated by smectites).

33.7.7 Chemical Properties

33.7.7.1 pH

Vertisols are commonly neutral to alkaline in reaction (Table 33.31), given that the majority are derived from calcareous or base-rich parent materials. However, acidic Vertisols (Ahmad, 1985) do exist and are recognized as Dystric great groups of the Aquerts, Uderts, and Usterts (pH values of 4.5 or less in 0.01M $CaCl_2$ [5.0 or less in 1:1 water] and electrical conductivity less than 4.0 dS m^{-1}). Some Vertisols have a sulfuric horizon with a pH or 3.5 or less (Soil Survey Staff, 2010) produced by the oxidation of sulfides and resulting formation of sulfuric acid, or they contain sulfidic materials that can oxidize and acidify the soil (e.g., Sulfaquerts).

33.7.7.2 Cation Exchange Capacity and Exchangeable Cations

Cation exchange capacity (CEC) of Vertisols generally ranges from about 30 to about 60 cmol$_c$ kg^{-1} (soil) and depends on the amount and type of mineral phases and soil organic matter (SOM) present (Table 33.31). Calcium and Mg commonly are the

main exchangeable cations. In some cases, Na is an abundant exchangeable cation, and when exchangeable Na percentage (ESP) exceeds 15, particle dispersion and soil structure deterioration can occur. Particle dispersion can occur at ESP values less than 15 if soluble salts are absent and the resulting soil solution electrolyte concentration is low. The presence of soluble salts and gypsum can maintain solution electrolyte concentrations at high enough levels to prevent soil dispersion even at ESP values greater than 15. Under acidic conditions, Al and H may become more prevalent in the soil solution and replace exchangeable Ca and the other base cations. An increase in the concentration of exchangeable Al and H and a decrease in CEC indicate that chemical weathering has occurred, via dissolution of primary and secondary (clay) minerals (Ahmad, 1985).

33.7.8 Soil Biochemical Properties

Biochemical properties are most strongly influenced by the quantity and type of organic matter, and land use (Table 33.32). Generally, organic C content ranges from 3 to 60 g kg^{-1}, but most commonly from 3 to 35 g kg^{-1} (Coulombe et al., 1996a), and varies with (1) climatic conditions (higher under more humid and cooler regimes); (2) microtopography (higher in microlows than in microhighs; Kunze and Templin, 1956; Wilding and Tessier, 1988; Nordt et al., 2004); (3) depth (generally decreases with depth in the soil profile, which refutes the argument that complete self-mixing occurs in Vertisols, or at least indicates that the rate of mixing is slower than the rate of soil organic C accumulation and redistribution; Wilding and Tessier, 1988; Southard and Graham, 1992; Coulombe, 1997); and (4) tillage (higher in noncultivated than in cultivated areas; Kunze and Templin, 1956; Skjemstad and Dalal, 1987; Coulombe, 1997; Chevallier et al., 2000; Nordt and Wilding, 2009).

Soil organic matter has some influence on soil color, but black Vertisols may contain as little as 3 g kg^{-1} organic C. The contribution of soil organic matter to the dark color is the result of strong clay–organic complexes in which smectites are coated

TABLE 33.31 Chemical Characteristics^a of Selected Vertisols Formed from a Wide Variety of Parent Materials

Regions	pH (1:1 Water)	CEC (cmol$_c$ kg^{-1} Soil)	Extractable Bases (cmol$_c$ kg^{-1} Soil)				References
			Ca	Mg	K	Na	
Australia	7.0 (0.6)	33.7 (2.5)	22.4 (3.4)	15.7(3.4)	0.1 (0.1)	0.9 (0.4)	Boettinger (1992)
India	8.2 (0.2)	55.8 (4.7)	53.4 (5.6)	^{b}	0.9 (0.1)	1.4 (0.5)	Hirekurubar et al. (1991)
Sudan	7.3 (0.4)	40.6 (11.0)	n/a	n/a	n/a	2.95 (3.9)	Blokhuis (1993)
Texas	7.4 (0.9)	41.0 (5.4)	52.9 (23.0)	6.8 (3.5)	0.7 (0.3)	2.7 (3.4)	Kunze et al. (1963) and Hallmark et al. (1986)
West Africa	7.4 (0.8)	30.3 (8.0)	17.9 (7.9)	7.5 (3.1)	0.7 (0.4)	3.2 (3.3)	Beavington (1978) and Yerima (1986)
California	8.0 (0.6)	45.7 (0.9)	26.8 (2.5)	17.1 (1.3)	0.8 (0.3)	1.9 (1.2)	NSSL Pedon No. 94P0056
El Salvador	6.7 (0.6)	45.5 (3.4)	33.9 (10.3)	11.6 (2.5)	0.5 (0.2)	1.9 (1.5)	Yerima et al. (1985)
West Indies	7.3 (0.6)	48.0 (11.7)	51.8 (23.6)	8.1 (5.8)	0.4 (0.3)	2.4 (2.1)	Ahmad and Jones (1969)
Uruguay	6.5 (0.5)	33.9 (7.9)	23.1 (7.5)	8.0 (2.3)	0.3 (0.1)	0.3 (0.1)	Lugo-Lopez et al. (1985)

n/a, Not available.

^{a} Data are reported as mean (sd).

^{b} Extractable Mg was compiled with data for extractable calcium.

TABLE 33.32 Biochemical Characteristics of Surface Horizons of Selected Vertisols

Regions	Soil Series (and/or Parent Material)	Land Use	OC (g kg⁻¹)	Total N (g kg⁻¹)	C/N Ratio	Total P (mg kg⁻¹)	References
Australia	Langlands-Logie clays	Virgin	23.0	2.41	9.5	n/a	Skjemstad et al. (1986)
	(Alluvial clayey sediments)	20 year cultivation	11.4	1.12	10.2	n/a	Skjemstad et al. (1986)
		35 year cultivation	8.9	0.89	10.0	n/a	Skjemstad et al. (1986)
		45 year cultivation	7.8	0.77	10.1	n/a	Skjemstad et al. (1986)
	Black Earth	10 year pasture	14.2	0.9–2.6	5.8–12.2	n/a	Jocteur Monrozier et al. (1991)
Texas	Houston Black	Virgin	20.9–48.2	1.37–3.53	13.7–21.1	0.35–0.55	Coulombe (1997)
		20 year restoration	13.7–24.0	0.87–1.45	18.4–19.8	0.23–0.29	Coulombe (1997)
		Pasture	15.5–33.3	0.97–2.02	18.2–25.8	0.29–0.44	Coulombe (1997)
		Cereal rotation	12.2–15.6	0.83–1.03	14.7–20.3	0.34–0.43	Coulombe (1997)
		Row crop rotation	15.6–19.3	0.96–1.45	13.4–19.4	0.29–0.48	Coulombe (1997)
Nigeria	Lacustrine plain	Cereal	0.26–0.58	0.06–0.08	5.2–7.2	5–10	Beavington (1978)
Trinidad	Princes Town clay	>100 year sugar cane	31.0–38.0	2.7–3.5	11	29–58	Ahmad and Jones (1969)
Jamaica	Carron Hall clay	Unimproved pasture	42.0–60.0	4.5–5.7	9–11	13–15	Ahmad and Jones (1969)
Barbados	Black soil	>300 year sugar cane	12.0–15.0	0.8–1.0	15	250–500	Ahmad and Jones (1969)
Antigua	Fitches clay	Mixed shrub/grasses	32.0–38.0	2.2–2.3	15–17	13–17	Ahmad and Jones (1969)
Dominica	Black soil	Shrub	11.0–28.0	1.0–2.1	11–13	2–4	Ahmad and Jones (1969)
California	Capay	Grass rangeland	12.8	1.56	8.2	n/a	NSSL Pedon No. 94P0056

n/a, Not available.

with humic materials in the presence of Ca (Singh, 1954, 1956). The resistance to degradation is due to this smectite–Ca–organic complex, rather than due to the recalcitrant chemical composition of the organic material (Skjemstad et al., 1986; Skjemstad and Dalal, 1987).

The C/N ratio generally ranges from 10 to 25 (Table 33.32). Microbial biomass may represent up to 5% of total organic C (Jocteur Monrozier et al., 1991). Polysaccharides and fungal hyphae tend to be concentrated at interparticle positions in the domains and quasicrystals of the clay minerals (Chenu, 1989; Chenu and Jaunet, 1990). Total P ranges widely (Table 33.32), mostly as a function of parent material composition, but may also reflect organic forms of P that vary as a function of SOM content and soil management.

Loss of organic C can occur rapidly after a change in land use from native conditions to cultivation. After a few decades or more of continuous cultivation, loss of organic C can be significant (Skjemstad et al., 1986; Skjemstad and Dalal, 1987; Puentes, 1990; Coulombe, 1997; Nordt and Wilding, 2009), but the trend can be reversed, at least to some degree, when cultivated fields are allowed to return to native conditions (Chevallier et al., 2000).

33.7.9 Physical Properties

33.7.9.1 Moisture Retention and Shrink/ Swell Behavior

Soil moisture conditions are especially important in Vertisols because they dominate the physical properties and shrink/swell behavior. Due to the clay content, Vertisols exhibit high water retention characteristics at all water potentials (Table 33.33).

TABLE 33.33 Physical and Mechanical Characteristics[a] of Selected Vertisols from a Wide Array of Parent Materials

Regions	Water 0.033 MPa (Mass %)	Water 1.5 MPa (Mass %)	Bulk Density 0.033 MPa (Mg m⁻³)	Bulk Density Oven Dry (Mg m⁻³)	COLE (cm cm⁻¹)	Liquid Limit (%)	Plastic Limit (%)	Plasticity Index (%)	References
India	38.3 (2.9)	17.8 (0.9)	n/a	1.3 (0.1)	n/a	61.0 (2.4)	34.0 (1.2)	26.9 (1.3)	Hirekurubar et al. (1991)
Texas	42.2 (3.2)	23.4 (2.3)	n/a	1.8 (0.2)	n/a	67.1 (3.8)	22.7 (4.3)	44.3 (3.6)	Kunze et al. (1963) and Yule and Ritchie (1980a)
California	33.5 (1.0)	21.9 (0.6)	1.31 (0.02)	1.85 (0.03)	0.12 (0.01)	n/a	n/a	n/a	NSSL Pedon No. 94P0056
West Africa	31.4 (6.5)	n/a	1.4 (0.1)	1.4 (0.1)	0.10 (0.03)	n/a	n/a	n/a	Yerima (1986)
El Salvador	n/a	n/a	1.1 (0.1)	1.1 (0.1)	0.16 (0.03)	n/a	n/a	n/a	Yerima et al. (1985)

n/a, Data not available.

[a] Data are reported as mean (sd).

Soil water content varies considerably depending on the amount and kind of clay, climatic conditions, and land use.

Shrink/swell behavior places major constraints on the management of Vertisols. Under normal field conditions, most of the volume change occurs at water potentials >−1.5 MPa (Yule and Ritchie, 1980a; Wilding and Tessier, 1988), but it is not clear whether volume change is isotropic (equidimensional) or anisotropic. Current knowledge suggests that individual undisturbed peds shrink and swell isotropically (Yule and Ritchie, 1980a; Cabidoche and Voltz, 1995). Yule and Ritchie (1980b) reported that shrinkage in large (73 cm diameter, 140 cm long) and small (10 cm diameter) cores of a Texas Vertisol was similar. However, under field conditions, Cabidoche and Voltz (1995) reported anisotropic volume change in a Vertisol from Guadeloupe (French West Indies) with slightly larger movements vertically than horizontally (anisotropy ratio ~0.8). They attributed the anisotropy to movement of the peds along oblique slickensides, which highlights the importance of spatial arrangement or geometry of the peds in a soil profile on shrink/swell behavior.

33.7.9.2 Bulk Density and Coefficient of Linear Extensibility

Bulk density values for selected Vertisols are presented in Table 33.33. Factors that influence bulk density include (1) water content, (2) clay content, (3) particle density of the mineral fraction, and (4) the method used to determine bulk density (e.g., core vs. clods). Bulk density is routinely used to estimate total porosity and other volumetric percentages and to assess structural conditions of soils. Bulk density should be coupled with morphological characteristics and other properties for a thorough assessment of soil structural conditions.

The coefficient of linear extensibility (COLE) is an index of soil shrinkage (Table 33.33). The COLE value is typically calculated from soil bulk densities measured at −0.033 and −1.5 MPa, assuming that soil shrinkage is isotropic [COLE = (bulk density dry/bulk density moist)$^{1/3}$ − 1]. COLE is correlated with total and fine clay contents, surface area, water retention at field capacity, and ESP (Anderson et al., 1973; Yule and Ritchie, 1980a; Yerima et al., 1989).

33.7.9.3 Consistence and Atterberg Limits

Consistence refers to "the degree and kind of cohesion and adhesion that soil exhibits, and/or the resistance of soil to deformation or rupture under an applied stress" (Schoeneberger et al., 2002). The consistence of Vertisols ranges from hard to extremely hard when dry, from friable to very firm when moist, and from moderately sticky and moderately plastic to very sticky and very plastic when wet. Vertisols exhibit very narrow ranges of consistence and soil water content favorable for workability and trafficability.

The Atterberg limits (liquid limit, plastic limit, and plasticity index; Chapter 3) are empirical tests used by geoscientists and engineers to determine the mechanical characteristics of a soil and are approximated by the soil consistence properties described above. Values for Vertisols are presented in Table 33.33. In general, values of Atterberg limits increase as the clay content increases and decrease in the order: smectite > illite > kaolinite.

33.7.9.4 Structure/Porosity and Related Properties

Soil processes such as gas exchange, heat transfer, water flow, and movement of solutes and organics are dependent on structure and porosity. Vertisols tend to form large pores (macro- and mesopores, i.e., cracks) during soil drying and shrinkage. These cracks have a profound impact on several soil processes. They allow oxygen at atmospheric partial pressure to penetrate to the depth of cracking, they allow greater nitrification rates than in the rest of the soil mass (Kissel et al., 1974), and they promote movement of water and solutes through preferential or bypass flow (Kissel et al., 1973, 1974; Bouma, 1983; Bouma and Loveday, 1988). Lin et al. (1996) reported, for a Texas Vertisol, that macropores (>0.5 mm effective diameter) and mesopores (0.06 to <0.5 mm effective diameter) contributed 89% and 10% of the water flow, respectively. Several soil properties influence infiltration rates (Table 33.34). In clayey soils that shrink and swell, hydraulic conductivity (K) does not meet the assumption of uniform flow in a homogenous soil (Darcy's law), so infiltration rates are often preferred over K to characterize water movement (Lin, 1995; Lin and McInnes, 1995). Nonetheless, infiltration rates can be highly variable in space and time and remain dependent on a variety of factors including land use and the method used to measure infiltration rate.

33.7.10 Land Use and Management Considerations

33.7.10.1 Natural Resources

33.7.10.1.1 Agriculture

For many regions of the world, Vertisols are highly productive and sustainable resources for various crop production systems, including a wide array of food and fiber crops. Nonetheless, Vertisols in developed and developing countries are subject to degradation due to intensive and/or inappropriate management practices. Soil properties may partially recover from the effects of cultivation, but it is unlikely that native conditions can be rejuvenated even after a restoration period of 20 years or more (Puentes, 1990; Coulombe, 1997; Chevallier et al., 2000; Nordt and Wilding, 2009). One of the difficulties in successfully managing Vertisols involves some limitations on technology transfer. Despite the common properties of high clay content, shrink/swell behavior, and cracking, many Vertisols require site-specific management for crop production (Ahmad and Mermut, 1996) due to variations in climate, soil chemistry, and cropping systems.

33.7.10.1.2 Wetlands

Vertisols, especially those on the wetter end of the climatic spectrum (i.e., Aquerts and Uderts), can potentially be used for the creation or restoration of wetland habitats (Chapter 20

TABLE 33.34 Infiltration Rates under Saturated Conditions for Surface Horizons of Selected Vertisols with Reference to Land Use and Method

Regions	Soil Series (and/or Parent Material)	Land Use	Method (References)	Infiltration ($\mu m\ s^{-1}$)	References
India	(Chlorite schist)	n/a	Double ring infiltrometer (Marshall and Stirk, 1950)	3.3	Hirekurubar et al. (1991)
	(Shale)	n/a	Double ring infiltrometer (Marshall and Stirk, 1950)	5.6	Hirekurubar et al. (1991)
	(Granite gneiss)	n/a	Double ring infiltrometer (Marshall and Stirk, 1950)	0.8	Hirekurubar et al. (1991)
	(Deccan trap)	n/a	Double ring infiltrometer (Marshall and Stirk, 1950)	1.2	Hirekurubar et al. (1991)
	(Limestone)	n/a	Double ring infiltrometer (Marshall and Stirk, 1950)	11.1	Hirekurubar et al. (1991)
Texas	Heiden	15–25 year pasture	Double ring infiltrometer (Bouwer, 1986)	0.70–0.94	Puentes (1990)
	Heiden	15–25 year cultivation	Double ring infiltrometer (Bouwer, 1986)	0.14	Puentes (1990)
	Houston Black	Virgin	Column method (Bouma, 1982)	1.86	Coulombe (1997)
			Ponded infiltrometer (Prieksat et al., 1992)	1.77	Potter (pers. comm.)
			Disc permeameter (Perroux and White, 1988)	2.13	Lin (pers. comm.)
		20 year restoration	Column method (Bouma, 1982)	1.04	Coulombe (1997)
			Ponded infiltrometer (Prieksat et al., 1992)	1.77	Potter (pers. comm.)
			Disc permeameter (Perroux and White, 1988)	2.29	Lin (pers. comm.)
		Pasture	Column method (Bouma, 1982)	0.61	Coulombe (1997)
			Ponded infiltrometer (Prieksat et al., 1992)	1.26	Potter (pers. comm.)
			Disc permeameter (Perroux and White, 1988)	1.86	Lin (pers. comm.)
		Cereal rotation	Column method (Bouma, 1982)	0.39	Coulombe (1997)
			Ponded infiltrometer (Prieksat et al., 1992)	1.75	Potter (pers. comm.)
			Disc permeameter (Perroux and White, 1988)	1.40	Lin (pers. comm.)
		Row crop rotation	Column method (Bouma, 1982)	0.88	Coulombe (1997)
			Ponded infiltrometer (Prieksat et al., 1992)	1.65	Potter (pers. comm.)
			Disc permeameter (Perroux and White, 1988)	1.37	Lin (pers. comm.)

Source: From Coulombe, C.E., L.P. Wilding, and J.B. Dixon. 2000. Vertisols, p. E-269–E-286. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.; Based on 1997 estimates by the USDA-NRCS.

n/a, Not available.

of *Handbook of Soil Sciences: Resource Management and Environmental Impacts*). Besides the presence of hydric conditions, a requirement for successful establishment of desired wetland functions includes development of wetland hydrology and the presence of hydrophytic vegetation. The swelling of Vertisols when wet drastically reduces permeability and provides a mechanism to more easily manage surface waters to maintain hydric conditions, establish wetland hydrology, and provide an environment that favors hydrophytic vegetation.

33.7.10.2 Engineering

33.7.10.2.1 Environmental Engineering

Due to their extensive shrink/swell behavior, Vertisols are generally not suitable for solid waste and wastewater disposal. Water generally percolates too rapidly when the soils are dry and too slowly when the soils are wet. In either case, these conditions could contribute to contamination of surface and groundwater and result in potential environmental health hazards. For similar reasons, Vertisols that exhibit a permanent (endoaquic) or seasonally perched (epiaquic) water table would not be a suitable filter for wastewater disposal. Further, disposal facilities built on Vertisols are prone to damage by shrink/swell processes that can crack foundations or rupture pipes.

On the other hand, Vertisols can be used as constructed treatment wetlands for final disposal of tertiary effluent provided that wetland hydrology is maintained at saturation to enhance low soil permeability and to reduce groundwater recharge. Treatment wetlands are effective technologies for water quality improvement of various wastewaters as well as being beneficial for wildlife habitats (Kadlec and Knight, 1995).

33.7.10.2.2 Civil Engineering

The shrink/swell nature of Vertisols severely limits civil engineering applications. Vertisols must be stabilized for construction of buildings, roads, pipelines, and utility corridors. For example, several major cities in Texas (Dallas, Houston, San Antonio, and Corpus Christi) have been built on Vertisols. The swell/shrink properties of Vertisols make effective construction site stabilization very difficult. The most economic choice may be to select stable soil (not Vertisols) for construction sites. If construction on a Vertisol is necessary, on-site investigation is generally required. Morphological properties of these soils often indicate movement at depths between 2 and 4 m due to natural moisture variability (Coulombe et al., 2000). These depths of movement suggest that soil management is the best way to minimize damage to structures on Vertisols and other expansive soils. Homeowners can maintain moist soils around the

foundations of their dwellings that are built on steel reinforced concrete slabs to minimize movement. Application of gypsum to reduce ESP minimizes excessive swelling of smectites.

Prior to construction of roads, streets, and many structures that are to be built on Vertisols, lime [Ca(OH)$_2$] stabilization of the soil is required to a depth of about 0.3 m or less (Eades and Grim, 1960; National Lime Association, 1991). Although this is beneficial, it obviously is only a partial treatment and complete stabilization would be extremely expensive. Larger multistory buildings sited on vertic soils and deep smectitic clays typically are supported by concrete piers with large bell-shaped feet at about 4 m depth to avoid contact with expansive soil.

Many other commercial methods to stabilize Vertisols are being promoted, some of which may be as successful as lime stabilization. One such treatment with a sulfonated naphthalene was tested under laboratory conditions on three Vertisols with some success (Marquart, 1995), but the results were inconsistent, depending on soil conditions and level of treatment.

It is imperative for construction workers to seriously consider the use of shoring materials and other safety precautions when a Vertisol is excavated. In Texas, from 1980 to 1985, 50 construction workers were killed due to destabilization of Vertisols that occurred during excavation (USDA-SCS, 1986).

33.7.11 Geological Studies of Lithified Paleosols Interpreted as Vertisols

33.7.11.1 Introduction

Vertisol-like paleosols here and elsewhere (hereafter referred to as "paleovertisols") are clay-rich paleosols that show abundant evidence for shrink–swell processes including slickensides, cracks, and strongly developed sepic–plasmic microfabrics. Vertisols are one of the few soil orders for which *Soil Taxonomy* can be easily applied to corresponding paleosols in the geological record (Mack et al., 1993). Paleovertisols have provided a wealth of information on paleoenvironment conditions, including Paleozoic (540 to 225 million years before present) levels of atmospheric CO$_2$ and changes in soil morphology accompanying the evolution and diversification of terrestrial land plants (Mora et al., 1996; Mora and Driese, 1999; Driese and Mora, 2001; Driese et al., 2005; Driese, 2009). Much of the interest in paleosols is driven by their potential to serve as climatic proxies in ancient environments.

Vertisols are characterized by macro- and micromorphologic characteristics formed in response to shrink–swell phenomena associated with either seasonal precipitation or seasonal soil moisture deficits. Many of these features preserve well in the ancient rock record in Paleovertisols and can be easily recognized in the field (in outcrop exposures and in cores) and in thin section; thus, Paleovertisols are increasingly reported in the geologic record, ranging in age from the Proterozoic (Rye and Holland, 2000; Driese, 2004) to the Cenozoic. Paleovertisols formed over a wide range of paleolatitudes, ranging from equatorial–low latitude (Appalachian basin; references cited below) to as high as 78° (Triassic of Antarctica; Retallack and Alonso-Zarza, 1998). Virtually every Paleozoic red bed formation in

the Appalachian Basin stratigraphic succession has common occurrences of Paleovertisols or vertic intergrades of other soil orders (Mora and Driese, 1999; Driese et al., 2000a; Driese and Mora, 2001; Stiles et al., 2001). The wide spatial and temporal distribution of Paleovertisols reflects the potential for their formation under wide ranges of moisture regimes (from semiarid to humid); the only strict climatic requirement is that the moisture is seasonally distributed. The widespread distribution also reflects the predominantly physical controls on Vertisol formation; these soils are not limited to specific environments or ecosystems. Retallack (1994) compiled a nonlinear relationship that predicts annual rainfall based upon the depth from the soil surface to pedogenic carbonate horizons. This relationship, based on measurements in carbonate-bearing soils representing a variety of soil orders, was better constrained specific for Vertisols by Nordt et al. (2006). Stable carbon and oxygen isotope analysis of pedogenic carbonate in Vertisols helps to elucidate the soil ecosystem (dominance of C$_3$ vs. C$_4$ flora) and changes in soil water compositions that may indicate climate change (Kovda et al., 2003, 2006). Similar information may be preserved in the isotopic compositions of Paleosols, as well as constraints on paleoatmospheric pCO$_2$ using Cerling's (1991) carbon isotope paleobarometer (Ekart et al., 1999).

33.7.11.2 Distribution of Fossil Vertisols in Geologic Record

Modern Vertisols today only comprise about 2% of world soils, far less abundant than most other soil orders recognized in USDA *Soil Taxonomy*, yet they appear to comprise the majority of paleosols preserved in the geologic record. Possible hypotheses that might explain this apparent overrepresentation of Paleovertisols in the geological record include the following:

1. There a strong recognitional bias for Paleovertisols in outcrop exposures and in cores because of ease of identification (discussed previously).
2. There are significant preservational processes biased for Vertisols. Because Vertisols are known to form very rapidly (within tens to hundreds of years, in some cases) and mostly in low-lying floodplains and lacustrine floodbasins, not much time is necessary for subaerial exposure and pedogenesis, and the resultant formation of a paleosol with high preservation potential. In addition, Vertisols can form over wide ranges of latitudes (e.g., from North Dakota to south Texas in the United States), and under wide ranges of moisture regimes. As such, they may have been more widely distributed in the geologic past than they are now. Because the physical processes of clay shrink–swell dominate soil-forming processes, plants may be unnecessary in Vertisol genesis (Mora and Driese, 1999). A wider temporal range (i.e., pre-Silurian) is therefore possible. This contrasts for example, with the distribution of soils with mollic epipedons (mostly Mollisols), the first appearance of which in the Tertiary is related to the appearance of dryland grasses (Retallack, 1997).

TABLE 33.35 Diagenetic Alteration of Paleovertisols (High Clay-Content Paleosols with Extensive Shrink–Swell Features Such as Slickensides) Compared with a Surface Soil Analog (Houston Black Series, TX)

Name (Geologic Age)	Burial Depth (km)	Burial Temperature (°C)	Clay Mineral Assemblage	wt% K_2O
Houston Black (modern)	Surface	22° Mean annual temperature	Na–smectite	1.29
Pennington (325 Ma)	2–3	60–90	Illite (+ kaolinite + chlorite)	4.61
Hekpoort (2.25 Ga)	12–15	350	Sericite–muscovite–chlorite	9.50

Sources: Driese, S.G., C.I. Mora, and J.M. Elick. 2000a. The paleosol record of increasing plant diversity and depth of rooting and changes in atmospheric pCO_2 in the Siluro-Devonian, p. 47–61. *In* R.A. Gastaldo and W.A. DiMichele (eds.) Phanerozoic terrestrial ecosystems, a short course. The Paleontological Society Papers 6, New Haven, CT; Driese, S.G., C.I. Mora, C.A. Stiles, R.M. Joeckel, and L.C. Nordt. 2000b. Mass-balance reconstruction of a modern Vertisol: Implications for interpretations of geochemistry and burial alteration of paleovertisols. Geoderma 95:179–204; Rye, R., and H.D. Holland. 2000. Geology and geochemistry of paleosols developed on the Hekpoort Basalt, Pretoria Group, South Africa. Am. J. Sci. 300:85–141. Driese, S.G. 2009. Paleosols, pre-Quaternary, p. 748–751. *In* V. Gornitz (ed.) Encyclopedia of paleoclimatology and ancient environments. Kluwer Academic Publishers, New York.

Ma, millions of years; Ga, billions of years. Note potassium enrichment occurring over time and with increasing burial depths and temperatures.

33.7.11.3 Paleoclimate Reconstructions Based on Fossil Vertisols

Paleoclimate reconstructions based on fossil Vertisols primarily emphasize estimation of mean annual precipitation (MAP) using a variety of proxy measures, including (1) depth to pedogenic carbonate (Bk) horizon (Caudill et al., 1996; Nordt et al., 2006); (2) Fe content of pedogenic Fe–Mn nodules and concretions (Stiles et al., 2001); (3) chemical indices of weathering used to estimate paleoprecipitation based on bulk geochemistry (e.g., Chemical Index of Alteration minus Potash or CIA − K, Sheldon et al., 2002 [note: CIA − K = molar ratio of $Al_2O_3/(Al_2O_3 + Na_2O + CaO) \times 100$, and the linear regression has an $r^2 = 0.73$], but a new paleoprecipitation proxy specific for Paleovertisols termed CALMAG has been developed where CALMAG = molar ratio of $Al_2O_3/(Al_2O_3 + CaO + MgO) \times 100$, and the linear regression obtained has an $r^2 = 0.90$; Nordt and Driese, 2010a); (4) total element mass-flux and mass-balance calculations (Driese et al., 2000b; Stiles et al., 2003); and (5) transfer functions for reconstructing colloidal properties of Paleovertisols from bulk geochemistry Nordt and Driese (2010b). Paleolandscape reconstructions in which topography or hydrology is important variable (interpretation of paleocatenas) are commonly conducted at both local and more regional scales. Reconstructions of Phanerozoic pCO_2 employ the CO_2-carbonate paleobarometer of Cerling (1991) to interpret Paleozoic (Mora et al., 1996; Mora and Driese, 1999) and Mesozoic paleoatmospheres from fossil Vertisols (Nordt et al., 2002, 2003). This technique utilizes the $\delta^{13}C$ values measured from pedogenic carbonates, measurements of $\delta^{13}C$ values of paleosol organic matter (or an estimate thereof based on marine proxy records), and assumptions of soil productivity (soil-CO_2 production rates) to estimate paleoatmospheric pCO_2. These results are in good agreement with pCO_2 estimates based on long-term mass-balance carbon models (Berner, 2006). Other research directions include systematic studies of root diameter, depth, and density in Paleosols using root traces, and relating these changes to development of soil morphology (Driese and Mora, 2001), widespread deposition of black shales, and carbon sequestration.

Problems in interpreting paleoclimates from pre-Quaternary fossil Vertisols include extensive burial diagenetic alteration and some compaction (Sheldon and Retallack, 2001). Commonly cited diagenetic alteration processes include physical compaction and consequent modifications of original soil thickness and morphology, oxidation of soil organic matter, burial gleization (Fe loss, largely through microbial reduction), color modifications (intensification) related to dehydration and recrystallization of hydrous mineral phases (such as FeOOH), recrystallization of soil smectites to illites (or even to metamorphic mineral assemblages; Table 33.35), and exchange of oxygen isotopes between burial fluids and pedogenic carbonates and clays (Mora et al., 1998; Mora and Driese, 1999). Evolutionary changes in terrestrial plant and animal communities over time also create difficulties in paleosol interpretations because of attendant changes in morphological features such as root traces, which are (1) not present in Ordovician paleosols; (2) rhizomatous and fine in Silurian paleosols; and (3) larger, deeper, and more "modern" in Devonian and post-Devonian paleosols (Driese et al., 2000a, 2000b; Driese and Mora, 2001).

Acknowledgment

References

Ahmad, N. 1983. Vertisols, p. 91–124. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy. Elsevier Scientific Publishing Co., New York.

Ahmad, N. 1985. Acid Vertisols of Trinidad, p. 141–151. *In* Proc. 5th Int. Soil Classification Workshop, Sudan, 2–11 November 1982. Soil Survey Administration, Khartoum, Sudan.

Ahmad, N., and R.L. Jones. 1969. Genesis, chemical properties and mineralogy of Caribbean Grumosols. Soil Sci. 107:166–174.

Ahmad, N., and A.R. Mermut. 1996. Vertisols and technologies for their management. Elsevier Scientific Publishers, Amsterdam, the Netherlands.

Anderson, J.U., K.E. Fadul, and G.A. O'Connor. 1973. Factors affecting the coefficient of linear extensibility in Vertisols. Soil Sci. Soc. Am. Proc. 37:296–299.

Baize, D., and M.C. Girard. 1996. Referentiel Pedologique 1995, Principaux sols d´Europe. INRA editions. Versailles, France.

Beavington, F. 1978. Studies of some cracking clay soils in the Lake Chad basin of north east Nigeria. J. Soil Sci. 29:575–583.

Berner, R.A. 2006. Geocarbsulf: A combined model for Phanerozoic atmospheric O_2 and CO_2. Geochim. Cosmochim. Acta 70:5653–5664.

Blokhuis, W.A. 1993. Vertisols of the central clay plain of the Sudan. Ph.D. Thesis. Wageningen Agricultural University. Wageningen, the Netherlands.

Blokhuis, W.A. 2002. Vertisols, p. 1370–1379. *In* R. Lal (ed.) Encyclopedia of soil science. Marcel Dekker, New York.

Boettinger, J.L. 1992. Genesis, mineralogy and geochemistry of a red-black (Alfisol–Vertisol) complex, northeastern Queensland, Australia. Ph.D. Dissertation. University of California. Davis, CA.

Bouma, J. 1982. Measuring the hydraulic conductivity of soil horizons with continuous macropores. Soil Sci. Soc. Am. J. 46:438–441.

Bouma, J. 1983. Hydrology and soil genesis of soils with aquic moisture regimes, p. 253–281. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy: I. Concepts and interactions. Elsevier Scientific Publishers, Amsterdam, the Netherlands.

Bouma, J., and J. Loveday. 1988. Characterizing soil water regimes in swelling clay soils, p. 83–96. *In* L.P. Wilding and R. Puentes (eds.) Vertisols: Their distribution, properties, classification and management. Texas A&M Printing Center, College Station, TX.

Bouwer, H. 1986. Intake rate: Cylinder infiltrometer, p. 825–844. *In* A. Klute (ed.) Methods of soil analysis, Part. I. Physical and mineralogical methods. 2nd edn. SSSA, Madison, WI.

Cabidoche, Y.M., and M. Voltz. 1995. Non-uniform volume and water content changes in swelling clay soil: II. A field study on a Vertisol. Eur. J. Soil Sci. 46:345–355.

Caudill, M.R., S.G. Driese, and C.I. Mora. 1996. Preservation of a paleo-Vertisol and an estimate of Late Mississippian palaeo-precipitation. J. Sediment. Res. 66:58–70.

Cerling, T.E. 1991. Carbon dioxide in the atmosphere: Evidence from Cenozoic and Mesozoic paleosols. Am. J. Sci. 291:377–400.

Chenu, C. 1989. Influence of a fungal polysaccharide, scleroglucan, on clay microstructures. Soil Biol. Biochem. 21:299–305.

Chenu, C., and A.M. Jaunet. 1990. Modifications de l'organisation texturale d'une montmorillonite calcique liees a l'adsorption d'un polysaccharide. C.R. Acad. Sci. Paris 310:975–980.

Chevallier, T., M. Voltz, E. Blanchart, J.L. Chotte, V. Eschenbrenner, M. Mahieu, and A. Albrecht. 2000. Spatial and temporal changes of soil C after establishment of a pasture on a long-term cultivated Vertisol (Martinique). Geoderma 93:43–58.

Coulombe, C.E. 1997. Surface properties of Vertisols of Texas and Mexico under different land use systems. Ph.D. Dissertation. Texas A&M University. College Station, TX.

Coulombe, C.E., J.B. Dixon, and L.P. Wilding. 1996a. Mineralogy and chemistry of Vertisols, p. 115–200. *In* N. Ahmad and A.R. Mermut (eds.) Vertisols and technologies for their management. Elsevier Scientific Publishers, Amsterdam, the Netherlands.

Coulombe, C.E., L.P. Wilding, and J.B. Dixon. 1996b. Overview of Vertisols: Characteristics and impacts on society. Adv. Agron. 57:289–375.

Coulombe, C.E., L.P. Wilding, and J.B. Dixon. 2000. Vertisols, p. E-269–E-286. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

CPCS. 1987. Classification des sols. E.N.S.A., Grignon, France.

Driese, S.G. 2004. Pedogenic translocation of Fe in modern and ancient Vertisols and implications for interpretations of the Hekpoort paleosol (2.25 Ga). J. Geol. 112:543–560.

Driese, S.G. 2009. Paleosols, pre-Quaternary, p. 748–751. *In* V. Gornitz (ed.) Encyclopedia of paleoclimatology and ancient environments. Kluwer Academic Publishers, New York.

Driese, S.G., and C.I. Mora. 2001. Diversification of Siluro-Devonian plant traces in paleosols and influence on estimates of paleoatmospheric CO_2 levels, p. 237–253. *In* P.G. Gensel and D. Edwards (eds.) Plants invade the land: Evolutionary and environmental perspectives. Columbia University Press, New York.

Driese, S.G., C.I. Mora, and J.M. Elick. 2000a. The paleosol record of increasing plant diversity and depth of rooting and changes in atmospheric pCO_2 in the Siluro-Devonian, p. 47–61. *In* R.A. Gastaldo and W.A. DiMichele (eds.) Phanerozoic terrestrial ecosystems, a short course. The Paleontological Society Papers 6, New Haven, CT.

Driese, S.G., C.I. Mora, C.A. Stiles, R.M. Joeckel, and L.C. Nordt. 2000b. Mass-balance reconstruction of a modern Vertisol: Implications for interpretations of geochemistry and burial alteration of paleovertisols. Geoderma 95:179–204.

Driese, S.G., L.C. Nordt, W.C. Lynn, C.A. Stiles, C.I. Mora, and L.P. Wilding. 2005. Distinguishing climate in the soil record using chemical trends in a Vertisol climosequence from the Texas Coastal Prairie, and application to interpreting Paleozoic paleosols in the Appalachian basin. J. Sediment. Res. 75:339–349.

Dudal, R., and H. Eswaran. 1988. Distribution, properties and classification of Vertisols, p. 1–22. *In* L.P. Wilding and R. Puentes (eds.) Vertisols: Their distribution, properties, classification and management. Texas A&M Printing Center, College Station, TX.

Eades, J.L., and R.E. Grim. 1960. The reaction of hydrated lime with pure clay minerals in soil stabilization. Highway Research Board Bulletin 262:51–63 (accessed on 11–15 January 1960).

Ekart, D.D., T.E. Cerling, I.P. Montañez, and N.J. Tabor. 1999. A 400 million year carbon isotope record of pedogenic carbonate: Implications for paleoatmospheric carbon dioxide. Am. J. Sci. 299:805–827.

Eswaran, H., F.H. Beinroth, P.R. Reich, and L.A. Quandt. 1999. Guy D. Smith memorial slide collection. Vertisols: Their propertied, classification, distribution and management. USDA-NRCS, Washington, DC.

FAO–UNESCO. 1988. Soil map of the world: Revised legend. World Soil Resources Report 60. FAO, Rome, Italy.

FAO–UNESCO. 1998. World reference base for soil resources, by ISSS–ISRIC–FAO. World Soil Resources Report 84. FAO, Rome, Italy.

Greene-Kelly, R. 1974. Shrinkage of clay soils: A statistical correlation with other soil properties. Geoderma 11:243–257.

Hallmark, C.T., L.T. West, L.P. Wilding, and L.R. Drees. 1986. Characterization data for selected Texas soils. TAES, USDA-SCS, College Station, TX.

Hirekurubar, B.M., V.S. Doddamani, and T. Satyanarayana. 1991. Some physical properties of Vertisols derived from different parent materials. J. Indian Soc. Soil Sci. 39:242–245.

Isbell, R.F. 1991. Australian Vertisols, p. 73–80. *In* J.M. Kimble (ed.) Characterization, classification, and utilization of cold Aridisols and Vertisols, Proc. VI ISCOM. USDA-SCS, National Soil Survey Center, Lincoln, NE.

IUSS Working Group WRB. 2007. World reference base for soil resources 2006, first update 2007. World Soil Resources Report No. 103. FAO, Rome, Italy. Available online at: http://www.fao.org/fileadmin/templates/nr/images/resources/pdf_documents/wrb2007_red.pdf (accessed on 21 June 2011).

Jacob, J.S., R.W. Griffin, W.L. Miller, and L.P. Wilding. 1997. Aquerts and aquertic soils: A querulous proposition, p. 61–77. *In* M.J. Vepraskas and S.W. Sprecher (eds.) Aquic conditions and hydric soils: The problem soils. SSSA Special Publication No. 50. SSSA, Madison, WI.

Jocteur Monrozier, L., J.N. Ladd, R.W. Fitzpatrick, R.C. Foster, and M. Raupach. 1991. Components and microbial biomass content of size fractions in soils of contrasting aggregation. Geoderma 50:37–62.

Kadlec, R.H., and R.L. Knight. 1995. Treatment wetlands. CRC Press, Boca Raton, FL.

Kimble, J.M. 1991. Characterization, classification, and utilization of cold Aridisols and Vertisols. *In* Proc. VI ISCOM. USDA-SCS, National Soil Survey Center, Lincoln, NE.

Kissel, D.E., J.T. Ritchie, and E. Burnett. 1973. Chloride movement in undisturbed swelling clay soil. Soil Sci. Soc. Am. Proc. 37:21–24.

Kissel, D.E., J.T. Ritchie, and E. Burnett. 1974. Nitrate and chloride leaching in a swelling clay soil. J. Environ. Qual. 3:401–404.

Kovda, I., C.I. Mora, and L.P. Wilding. 2006. Stable isotope compositions of pedogenic carbonates and soil organic matter in a temperate climate Vertisol with gilgai, southern Russia. Geoderma 136:423–435.

Kovda, I.V., L.P. Wilding, and L.R. Drees. 2003. Micromorphology, submicroscopy and microprobe study of carbonate pedofeatures in a Vertisol gilgai soil complex, South Russia. Catena 54:457–476.

Kunze, G.W., H. Oakes, and M.E. Bloodworth. 1963. Grumusols of the Coast Prairie of Texas. Soil Sci. Soc. Am. Proc. 27:412–421.

Kunze, G.W., and E.H. Templin. 1956. Houston black clay, the type Grumusol: II. Mineralogical and chemical characterization. Soil Sci. Soc. Am. Proc. 27:412–421.

Lin, H.S. 1995. Hydraulic properties and macropore flow of water in relation to soil morphology. Ph.D. Dissertation. Texas A&M University. College Station, TX.

Lin, H.S., and K.J. McInnes. 1995. Water flow in clay soil beneath a tension infiltrometer. Soil Sci. 159:375–382.

Lin, H.S., K.J. McInnes, L.P. Wilding, and C.T. Hallmark. 1996. Effective porosity and flow rate with infiltration at low tensions into a well-structured subsoil. Trans. ASAE 39:131–135.

Lugo-Lopez, M.A., J.P. Carnelli, and G. Acevedo. 1985. Morphological, physical and chemical properties of major soils from Calagua in northwestern Uruguay. Soil Sci. Soc. Am. J. 49:108–113.

Mack, G.H., W.C. James, and H.C. Monger. 1993. Classification of paleosols. Geol. Soc. Am. Bull. 105:129–136.

Marquart, D.K. 1995. Chemical stabilization of three Texas Vertisols with sulfonated naphthalene. M.S. Thesis. Texas A&M University. College Station, TX.

Marshall, J.T., and G.B. Stirk. 1950. The effect of lateral movement of water in soil on infiltration measurement. Aust. J. Agric. Res. 1:253–257.

Mora, C.I., and S.G. Driese. 1999. Palaeoclimatic significance and stable carbon isotopes of Palaeozoic red bed paleosols, Appalachian Basin, USA and Canada, p. 61–84. *In* M. Thiry and R. Simon-Coinçon (eds.) Palaeoweathering, palaeosurfaces and related continental deposits. International Association of Sedimentologists. Special Publication No. 27. Wiley, Hoboken, NJ.

Mora, C.I., S.G. Driese, and L.A. Colarusso. 1996. Middle to Late Paleozoic atmospheric CO_2 from soil carbonate and organic matter. Science 271:1105–1107.

Mora, C.I., B.T. Sheldon, W.C. Elliott, and S.G. Driese. 1998. An oxygen isotope study of illite and calcite in three Appalachian Paleozoic vertic paleosols. J. Sediment. Res. 68:456–464.

National Lime Association. 1991. Lime stabilization construction manual. Bulletin 326. National Lime Association, Arlington, VA.

Nordt, L., S. Atchley, and S.I. Dworkin. 2002. Paleosol barometer indicates extreme fluctuations in atmospheric CO_2 across the Cretaceous–Tertiary boundary. Geology 30:703–706.

Nordt, L., S. Atchley, and S.I. Dworkin. 2003. Terrestrial evidence for two greenhouse events in the latest Cretaceous. GSA Today 13:4–9.

Nordt, L.C., and S.G. Driese. 2009. Hydropedological assessment of a Vertisol climosequence on the Gulf Coast Prairie land resource area of Texas. Hydrol. Earth Syst. Sci. Discuss. 6:3637–3668.

Nordt, L.C., and S.G. Driese. 2010a. New weathering index improves paleorainfall estimates from Vertisols. Geology 38:407–410.

Nordt, L.C., and S.G. Driese. 2010b. A modern soil characterization approach to reconstructing physical and chemical properties of paleo-Vertisols. Am. J. Sci. 310:37–64.

Nordt, L., M. Orosz, S. Driese, and J. Tubbs. 2006. Vertisol carbonate properties in relation to mean annual precipitation: Implications for paleoprecipitation estimates. J. Geol. 114:501–510.

Nordt, L.C., and L.P. Wilding. 2009. Organic carbon stocks and sequestration potential of Vertisols in the Coast Prairie land resource area of Texas, p. 159–168. *In* R. Lal and R. Follett (eds.) Soil carbon sequestration and the greenhouse effect. SSSA Special Publication No. 57. SSSA, Madison, WI.

Nordt, L., L. Wilding, W. Lynn, and C. Crawford. 2004. Vertisol genesis in a humid climate in the coastal plain of Texas. Geoderma 122:83–102.

Perroux, K.M., and I. White. 1988. Design for disc permeameters. Soil Sci. Soc. Am. J. 52:1205–1215.

Prieksat, M.A., M.D. Ankeny, and T.C. Kaspar. 1992. Design for an automated, self-regulating, single ring infiltrometer. Soil Sci. Soc. Am. Proc. 36:874–879.

Probert, M.E., I.F. Fergus, B.J. Bridge, D. McGarry, C.H. Thompson, and J.S. Russel. 1987. The properties and management of Vertisols. CAB International, Wallingford, U.K.

Puentes, R. 1990. Soil structure restoration in Vertisols under pastures in Texas. Ph.D. Dissertation. Texas A&M University. College Station, TX.

Quirk, J.P. 1994. Interparticle forces: A basis for the interpretation of soil physical behavior. Adv. Agron. 53:121–183.

Retallack, G.J. 1994. The environmental factor approach to the interpretation of paleosols, p. 31–64. *In* R. Amundson, J. Harden, and M. Singer (eds.) Factors in soil formation—A fiftieth anniversary perspective. SSSA Special Publication No. 33. SSSA, Madison, WI.

Retallack, G.J. 1997. Neogene expansion of the North American prairie. Palaios 12:380–390.

Retallack, G.J., and A.M. Alonso-Zarza. 1998. Middle Triassic paleosols and paleoclimate of Antarctica. J. Sediment. Res. 68:169–184.

Rye, R., and H.D. Holland. 2000. Geology and geochemistry of paleosols developed on the Hekpoort Basalt, Pretoria Group, South Africa. Am. J. Sci. 300:85–141.

Schoeneberger, P.J., D.A. Wysocki, E.C. Benham, and W.D. Broderson (eds.). 2002. Field book for describing and sampling soils, Version 2.0. USDA-NRCS, National Soil Survey Center, Lincoln, NE.

Sheldon, N.D., and G.J. Retallack. 2001. Equation for compaction of paleosols due to burial. Geology 29:247–250.

Sheldon, N.D., G.J. Retallack, and S. Tanaka. 2002. Geochemical climofunctions from North American soils and application to paleosols across the Eocene–Oligocene boundary in Oregon. J. Geol. 110:687–696.

Singh, S. 1954. A study of the black cotton soils with special reference to their coloration. J. Soil Sci. 5:289–299.

Singh, S. 1956. The formation of dark coloured clay–organic complexes in black soils. J. Soil Sci. 7:43–58.

Skjemstad, J.O., and R.C. Dalal. 1987. Spectroscopic and chemical differences in organic matter of two Vertisols subjected to long periods of cultivation. Aust. J. Soil Res. 25:323–335.

Skjemstad, J.O., R.C. Dalal, and P.F. Barron. 1986. Spectroscopic investigations of cultivation effects on organic matter of Vertisols. Soil Sci. Soc. Am. J. 50:354–359.

Soil Survey Staff. 1960. Soil classification, a comprehensive system. 7th Approximation. USDA-SCS, U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1975. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. USDA-SCS, agriculture handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd edn. USDA-NRCS, agriculture handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS, National Soil Survey Center, Lincoln, NE.

Southard, R.J., and R.C. Graham. 1992. Cesium-137 distribution in a California Pelloxerert: Evidence of pedoturbation. Soil Sci. Soc. Am. J. 56:202–207.

Stiles, C.A., C.I. Mora, and S.G. Driese. 2001. Pedogenic iron-manganese nodules in Vertisols: A new proxy for paleoprecipitation? Geology 29:943–946.

Stiles, C.A., C.I. Mora, S.G. Driese, and A.C. Robinson. 2003. Distinguishing climate and time in the soil record: Mass-balance trends in Vertisols from the Texas Gulf Coastal Prairie. Geology 31:331–334.

Tessier, D. 1984. Etude experimentale de l'organisation des materiaux argileux; Hydratation, gonflement et structuration au cours de la dessication et de la rehumectation. Ph.D. Thesis. University of Paris. Paris, France.

USDA-NRCS. 2009. The twelve orders of soil taxonomy. Available online at: http://soils.usda.gov/technical/soil_orders/ (accessed on 21 June 2011).

USDA-SCS. 1986. Soil cave-in: A fatal slip. Fact sheet. USDA, Washington, DC.

Wilding, L.P., and R. Puentes. 1988. Vertisols: Their distribution, properties, classification, and management. Texas A&M Printing Center, College Station, TX.

Wilding, L.P., and D. Tessier. 1988. Genesis of Vertisols: Shrink/swell phenomena, p. 55–81. *In* L.P. Wilding and R. Puentes (eds.) Vertisols: Their distribution, properties, classification and management. Texas A&M Printing Center, College Station, TX.

Yerima, B.P.K. 1986. Soil genesis, phosphorus and micronutrients of selected Vertisols and associated Alfisols of northern Cameroon. Ph.D. Dissertation. Texas A&M University. College Station, TX.

Yerima, B.P.K., F.G. Calhoun, A.L. Senkayi, and J.B. Dixon. 1985. Occurrence of interstratified kaolinite-smectite in El Salvador Vertisols. Soil Sci. Soc. Am. J. 49:462–466.

Yerima, B.P.K., L.R. Hossner, L.P. Wilding, and F.G. Calhoun. 1988. Forms of phosphorus and phosphorus sorption in northern Cameroon Vertisols and associated Alfisols, p. 147–164. *In* L.P. Wilding and R. Puentes (eds.) Vertisols: Their distribution, properties, classification and management. Texas A&M Printing Center, College Station, TX.

Yerima, B.P.K., L.P. Wilding, C.T. Hallmark, and F.G. Calhoun. 1989. Statistical relationships among selected properties of northern Cameroon Vertisols and associated Alfisols. Soil Sci. Soc. Am. J. 53:1758–1763.

Yule, D.F., and J.T. Ritchie. 1980a. Soil shrinkage relationships of Texas Vertisols: I. Small cores. Soil Sci. Soc. Am. J. 44:1285–1291.

Yule, D.F., and J.T. Ritchie. 1980b. Soil shrinkage relationships of Texas Vertisols: II. Large cores. Soil Sci. Soc. Am. J. 44:1291–1295.

33.8 Mollisols

P.A. McDaniel

R.G. Darmody

J.C. Bell

D.G. McGahan

33.8.1 Introduction

Mollisols are generally characterized as soils with thick, dark surface horizons (mollic epipedons) resulting from organic C incorporation. The terms mollic and Mollisol are derived from the Latin *mollis*, that is, soft. While the soil taxon, Mollisols, is defined by an exact set of soil property criteria, these criteria arose from conceptual ideas and empirical observations of soil landscapes with thick, dark, and often fertile surface horizons. Mollisols can form under multiple environmental conditions that facilitate accumulation of organic C in the upper soil profile. Although Mollisols correspond to the Chernozem soils of the Russian and older U.S. classification systems, the Mollisol soil order includes soils that are beyond the central concept of Chernozems (Fanning and Fanning, 1989).

Mollisols are among the most important soils for food and fiber production due to relatively high levels of native fertility coupled with climatic conditions conducive to plant growth. For this reason, the characteristics of Mollisols have been studied extensively, with early efforts focusing on the Chernozems of the Russian Steppes. During the late nineteenth century, droughts and crop failures afflicted the Russian Chernozem region resulting in widespread famine. In an effort to determine if the droughts could be prevented or moderated, Dokuchaev (1883) initiated an extensive study of the climate, soils, vegetation, and topography of the Chernozem Steppes. Based on the physical evidence that he and his colleagues collected, Dokuchaev theorized that the Chernozems of the Russian Steppes formed under specific conditions of climate, vegetation, and topography. These ideas contradicted the in vogue theories that Chernozems originated under either aquatic or marsh conditions. Dokuchaev's classic studies not only provided insight into the potential processes responsible for the formation of Chernozems, and subsequently Mollisols, but also articulated a conceptual framework for soil genesis that is still one of the foundations of modern pedology.

33.8.2 Geography and General Characteristics

33.8.2.1 Geographic Distribution

Extensive areas of Mollisols are distributed throughout the mid-latitudes of the world, predominantly on subhumid steppes and prairies (Figure 33.45). Mollisols occur under a wide range of climatic conditions including xeric, ustic, udic, and aquic soil moisture regimes and cryic, frigid, mesic, thermic, and hyperthermic soil temperature regimes. Spatially, Mollisols are the 8th (of 12) most common soil order and are estimated to cover
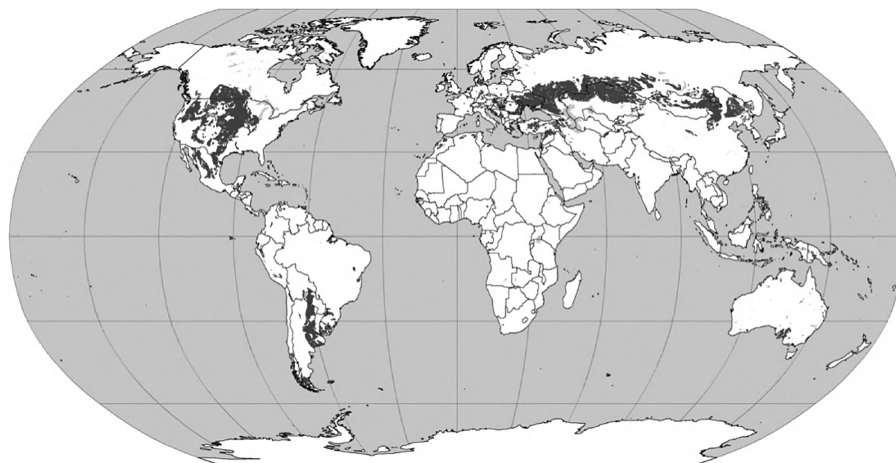
FIGURE 33.45 Global distribution of Mollisols. (Courtesy of USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC, 1998.)

TABLE 33.36   Estimated Occurrence of Mollisol Suborders Worldwide by General Climatic Condition

Climatic Conditions	Major Geographic Locations	Mollisol Suborder	Approximate Land Area (km²)
Semiarid	Great Plains of United States and Canada, Southern Russia, Pampas Region of Argentina, Yucatan Peninsula of Mexico	Ustolls	5,244,636
Humid	Black Sea Region of Russia, North-Central United States, Western Europe	Udolls and Rendolls	1,526,878
Cold	Mountainous regions of the Western United States and Canada, higher latitudes across North America, Kazakhstan, Ukraine, and Russia	Gelolls and Cryolls	1,163,797
Mediterranean	Turkey, Argentina, Palouse Region of United States, California	Xerolls	924,394
Seasonally saturated	Red River Valley of North, riparian zones, Gulf coastal plain, glaciated regions of the United States	Aquolls and Albolls	145,728

Source: Based on 1997 Estimates by the USDA-NRCS; NRCS Soil Survey Staff: http://soils.usda.gov/survey/
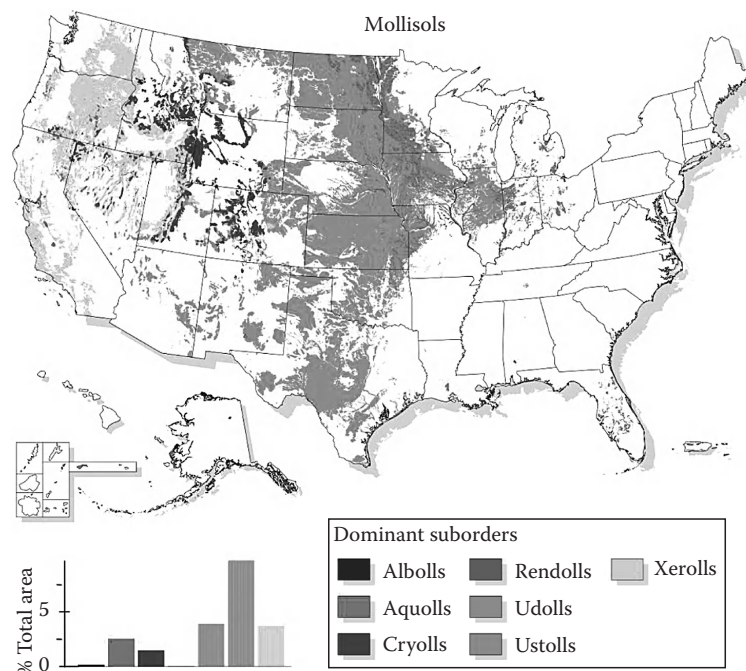
FIGURE 33.46   Distribution of Mollisols in the United States. (From http://soils.usda.gov/technical/classifcation/orders/mollisols_map.html).

9,005,433 km² or approximately 7% of the ice-free land area on Earth (Table 33.1). The distribution of Mollisol suborders can be distinguished by differences in climate, with the exception of wet Mollisols (Aquolls and Albolls), which typically occur in localized areas subject to saturated and reduced soil conditions (Table 33.36). The distribution of Mollisol suborders in the United States is shown in Figure 33.46. The close relationship of prairie ecosystems and Mollisols is apparent in the U.S. distribution where Mollisols are more common in the Western United States, west of the forested eastern half of the country. Noticeably, the "Prairie Peninsula" shows up in Figure 33.46 as the eastern expansion of Mollisols into Illinois and Indiana (Geis and Boggess, 1968).

33.8.2.2 General Characteristics of Mollisols

The general concept of Mollisols is that of dark colored soils of semiarid to subhumid grassland ecosystems. The dark color reflects soil organic matter (SOM) enrichment in the upper portion of the profile (Figure 33.47). The formation of dark surface horizons is termed melanization, which is actually a combination of several processes involving the addition of organic matter to the soil in the form of plant residues and its subsequent transformation into humus (Buol et al., 2003). Distinguishing features of Mollisols include the presence of a mollic epipedon and high base status (Soil Survey Staff, 2010). A mollic surface layer by itself is not diagnostic for Mollisols, as it can be a feature of soils belonging to other orders. Mollisols are differentiated in these cases by the presence of high base status (>50% base saturation) horizons that underlie the mollic epipedon, or by the lack of features associated with high shrink/swell clays, typical in Vertisols (Graham and Southard, 1983). A wide variety of subsurface horizons can occur beneath the mollic epipedon of Mollisols, including albic, cambic, argillic, calcic, petrocalcic, natric, duripans, and gleyed horizons. However, fragipans are unknown in Mollisols (Franzmeier et al., 1989). The nature of
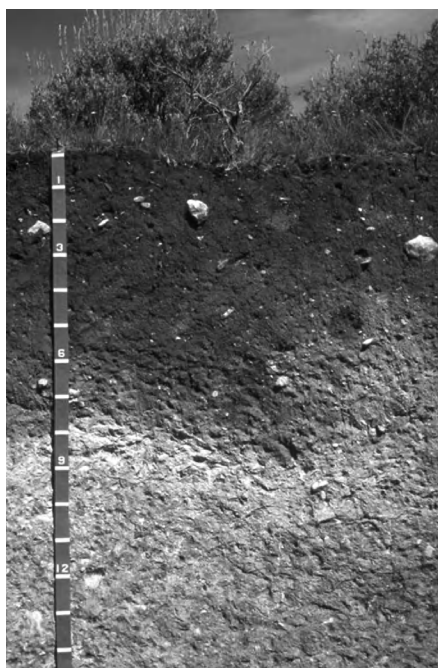
FIGURE 33.47 Soil profile of a Mollisol showing thick, dark mollic epipedon. The mollic epipedon extends to a depth of almost 60 cm; scale is in decimeters.

any diagnostic horizon is usually expressed at the great group level in classification with the exception of Albolls (albic) and Aquolls (gleyed).

A mollic epipedon is defined in *Soil Taxonomy* by several quantitative criteria including horizon thickness, organic C content, color, consistence, structure, and base status (Soil Survey Staff, 1999, 2010). Although a mollic epipedon must contain at least 1% SOM (0.6% organic C) on a weight basis, many Mollisols have higher SOM contents. A mollic epipedon must generally be a minimum of 18 cm thick for shallow and 25 cm thick for deep soils. Soils, which are very shallow, require only a 10 cm thick mollic epipedon (Soil Survey Staff, 2010). Munsell color requirements include a value (a measure of lightness/darkness) that is 3 or less for a moist sample and 5 or less for a dry sample. In both cases, the chroma (intensity or purity of color) is 3 or less. Thus, the colors of a mollic epipedon range from black or very dark brown when moist and are somewhat lighter (higher value) when dry. The color criteria were established as a field criterion to separate high organic matter soils formed under grasslands from their lower organic matter counterparts under forest. Mollic epipedons often have well-developed granular structure that is quite friable. By definition, a mollic epipedon cannot be both hard and massive when dry. A minimum of 50% base saturation is required for a mollic epipedon indicating a dominance of basic cations, primarily Ca and Mg, occupying cation exchange sites.

The high base status of Mollisols generally translates into a high level of native fertility. Both calcium and Mg, typically the dominant exchangeable cations, are required in fairly large quantities for plant growth. The pH values for Mollisols may be quite variable ranging from strongly acid (5.1–5.5) to strongly alkaline (8.5–9.0). However, many Mollisols have pH values somewhere in the middle of this range and are generally considered to be favorable for plant growth without widespread use of liming agents. Their high native fertility is one of the main reasons for their extensive use in agriculture throughout the world.

Mollisols generally have not undergone intensive weathering, and therefore, their mineralogy is often dominated by minerals inherited from their parent materials. Many Mollisols have formed in recently deposited parent materials such as glacial till and loess, which has not allowed sufficient time for significant mineral weathering to occur. Mineral weathering is also limited by the cooler temperatures and/or lack of moisture that are associated with many Mollisols. Clay mineralogy of Mollisols is typically dominated by 2:1 layer silicates, including clay-sized mica (illite), smectite, and vermiculite (Allen and Fanning, 1983). Generally, kaolinite, a 1:1 layer silicate, is commonly found in Mollisols although in small quantities (Allen and Hajek, 1989). However, soils dominated by kaolinite, such as the Ewa soil series from Hawaii (a fine, kaolinitic, isohyperthermic Aridic Haplustoll), are sometimes found (USDA-NRCS, 2009).

Many of the physical characteristics of Mollisols are influenced by SOM and the associated biological processes. The relatively high SOM content of Mollisols is conducive to the formation of water-stable aggregates, which typically have crumb or granular structure, and is formed and maintained by the interaction of plant roots, microbial structures, and biological molecules (Tate, 1992). This type of stable structure is important for good water infiltration and reduced susceptibility of soils to erosion. Organic matter plays a considerable role in enhancing the water-holding capacity of Mollisols, as it can hold up to 20 times its weight of water (Stevenson, 1994). The dark colors associated with Mollisols are attributable to the humic fractions of SOM, which are black and mask the lighter colors associated with fulvic fractions (Schultze et al., 1993).

33.8.3 Classification

Soils classified as Mollisols in *Soil Taxonomy* can be divided into 8 suborders and 37 great groups. While all these soils have a mollic epipedon, they are separated into the various suborders largely on the basis of their climatic regimes. Mollisols include soils that were formerly classified as Chernozems, Brunizems, Chestnut, and Rendzina soils and some soils formerly classified as Brown, Brown Forest, Humic Gley, and Planosols (Fenton, 1983). Table 33.37 lists the suborders, great groups, and subgroups in the Mollisol order. Tables 33.38 and 33.39 list the estimated areal extent of the Mollisol suborders and great groups in the NRCS database for the United States.

33.8.3.1 Albolls

Albolls are the sixth most commonly mapped Mollisol suborder (Table 33.38). They have a well-developed, light-colored horizon (albic) and a fluctuating water table. The albic horizon forms as pigmenting agents such as SOM, clays, and Fe oxides are

TABLE 33.37 Listing of Suborders, Great Groups, and Subgroups in the Mollisols Order

Suborder	Great Group	Subgroups
Albolls	Natralbolls	Leptic, Typic
	Argialbolls	Xerertic, Vertic, Argiaquic Xeric, Argiaquic, Xeric, Aquandic, Typic
Aquolls	Cryaquolls	Vertic, Histic, Thapto-Histic, Aquandic, Argic, Calcic, Cumulic, Typic
	Duraquolls	Natric, Vertic, Argic, Typic
	Natraquolls	Vertic, Glossic, Typic
	Calciaquolls	Petrocalcic, Aeric, Typic
	Argiaquolls	Arenic, Grossarenic, Vertic, Abruptic, Typic
	Epiaquolls	Cumulic Vertic, Fluvaquentic Vertic, Vertic, Histic, Thapto-Histic, Aquandic, Duric, Cumulic, Fluvaquentic, Typic
	Endoaquolls	Lithic, Cumulic Vertic, Fluvaquentic Vertic, Vertic, Histic, Thapto-Histic, Aquandic, Duric, Cumulic, Fluvaquentic, Typic
Rendolls	Cryrendolls	Lithic, Typic
	Haprendolls	Lithic, Vertic, Inceptic, Entic, Typic
Gelolls	Haplogelolls	Lithic, Andic, Aquic, Cumulic, Typic
Cryolls	Duricryolls	Argic, Calcic, Typic
	Natricryolls	Typic
	Palecryolls	Aquic, Oxyaquic, Abruptic, Pachic, Ustic, Xeric, Typic
	Argicryolls	Lithic, Vertic, Andic, Vitrandic, Abruptic, Aquic, Oxyaquic, Calcic Pachic, Pachic, Calcic, Alfic, Ustic, Xeric, Typic
	Calcicryolls	Lithic, Vitrandic, Petrocalcic, Pachic, Ustic, Xeric, Typic
	Haplocryolls	Lithic, Vertic, Andic, Vitrandic, Aquic Cumulic, Cumulic, Fluvaquentic, Aquic, Oxyaquic, Calcic Pachic, Pachic, Fluventic, Calcic, Ustic, Xeric, Typic
Xerolls	Durixerolls	Vertic, Vitritorrandic, Vitrandic, Aquic, Paleargidic, Abruptic Argiduridic, Cambidic, Haploduridic, Argidic, Argiduridic, Haplic Palexerollic, Palexerollic, Haplic Haploxerollic, Haploxerollic, Haplic, Typic
	Natrixerolls	Vertic, Aquic Duric, Aquic, Aridic, Duric, Typic
	Palexerolls	Vertic, Vitrandic, Aquic, Pachic, Petrocalcidic, Duric, Aridic, Petrocalcic, Ultic, Haplic, Typic
	Calcixerolls	Lithic, Vertic, Aquic, Oxyaquic, Pachic, Vitrandic, Aridic, Vermic, Typic
	Argixerolls	Lithic Ultic, Lithic, Torrertic, Vertic, Andic, Vitritorrandic, Vitrandic, Aquultic, Aquic, Oxyaquic, Alfic, Calcic Pachic, Pachic Ultic, Pachic, Argiduridic, Duric, Calciargidic, Aridic, Calcic, Ultic, Typic
	Haploxerolls	Lithic Ultic, Lithic, Torrertic, Vertic, Andic, Vitritorrandic, Vitrandic, Aquic Cumulic, Cumulic Ultic, Cumulic, Fluvaquentic, Aquic Duric, Aquultic, Aquic, Oxyaquic, Calcic Pachic, Pachic Ultic, Pachic, Torrifluventic, Duridic, Calcidic, Torripsammentic, Torriorthentic, Aridic, Duric, Psammentic, Fluventic, Vermic, Calcic, Entic Ultic, Ultic, Entic, Typic
Ustolls	Durustolls	Natric, Haploduridic, Argiduridic, Entic, Haplic, Typic
	Natrustolls	Leptic Torrertic, Torrertic, Leptic Vertic, Glossic Vertic, Vertic, Aridic Leptic, Leptic, Aquic, Aridic, Duric, Glossic, Typic
	Calciustolls	Salidic, Lithic Petrocalcic, Lithic, Torrertic, Udertic, Vertic, Petrocalcic, Gypsic, Aquic, Oxyaquic, Pachic, Aridic, Udic, Typic
	Paleustolls	Torrertic, Udertic, Vertic, Aquic, Pachic, Petrocalcic, Calcidic, Aridic, Udic, Calcic, Entic, Typic
	Argiustolls	Aridic Lithic, Alfic Lithic, Lithic, Aquertic, Torrertic, Pachic Udertic, Udertic, Pachic Vertic, Vertic, Andic, Vitritorrandic, Vitrandic, Aquic, Oxyaquic, Pachic, Alfic, Calcidic, Aridic, Udic, Duric, Typic
	Vermustolls	Lithic, Aquic, Pachic, Entic, Typic
	Haplustolls	Salidic, Ruptic-Lithic, Aridic Lithic, Lithic, Aquertic, Torrertic, Pachic Udertic, Udertic, Pachic Vertic, Vertic, Torroxic, Oxic, Andic, Vitritorrandic, Vitrandic, Aquic Cumulic, Cumulic, Anthraquic, Fluvaquentic, Aquic, Pachic, Oxyaquic, Torrifluventic, Torriorthentic, Aridic, Fluventic, Duric, Udorthentic, Udic, Entic, Typic
Udolls	Natrudolls	Petrocalcic, Leptic Vertic, Glossic Vertic, Vertic, Leptic, Glossic, Calcic, Typic
	Calciudolls	Lithic, Vertic, Aquic, Fluventic, Typic
	Paleudolls	Vertic, Petrocalcic, Aquic Pachic, Pachic, Aquic, Oxyaquic, Calcic, Typic
	Argiudolls	Lithic, Aquertic, Oxyaquic Vertic, Pachic Vertic, Alfic Vertic, Vertic, Andic, Vitrandic, Aquic Pachic, Pachic, Aquic, Oxyaquic, Lamellic, Psammentic, Arenic, Abruptic, Alfic, Oxic, Calcic, Typic
	Vermudolls	Lithic, Haplic, Typic
	Hapludolls	Lithic, Aquertic, Pachic Vertic, Vertic, Andic, Vitrandic, Aquic Cumulic, Cumulic, Fluvaquentic, Fluventic, Aquic Pachic, Pachic, Aquic, Oxyaquic, Vermic, Calcic, Entic, Typic

Source: Soil Survey Staff. 2006. Keys to soil taxonomy. 10th edn. USDA-NRCS., U.S. Government Printing Office, Washington, DC. NRCS Soil Survey Staff: http://soils.usda.gov/survey/

TABLE 33.38 Area of Mollisol Suborders as Mapped in the United States

Suborder	Acres	km²	%
Ustolls	397,898,253	1,610,239	45
Udolls	188,190,505	761,581	21
Xerolls	134,345,386	543,677	15
Aquolls	112,247,357	454,249	13
Cryolls	43,078,506	174,333	5
Albolls	7,750,437	31,365	1
Rendolls	1,400,059	5,666	0.2
Gelolls	30	0.1	0.0
Grand total	884,910,533	3,581,109	100

Source: NRCS Soil Survey Staff: http://soils.usda.gov/survey/

TABLE 33.39 Area of Mollisol Great Groups as Mapped in the United States

Great Group	Acres	km²	%
Argiustolls	169,731,080	686,878	19
Haplustolls	114,527,789	463,478	13
Hapludolls	94,569,463	382,709	11
Argiudolls	81,990,684	331,805	9
Calciustolls	64,759,491	262,073	7
Argixerolls	62,415,968	252,589	7
Endoaquolls	61,711,978	249,740	7
Haploxerolls	56,061,358	226,872	6
Paleustolls	38,498,924	155,800	4
Argiaquolls	24,051,906	97,335	3
Haplocryolls	22,774,652	92,166	3
Calciaquolls	20,982,190	84,912	2
Argicryolls	17,353,940	70,229	2
Natrustolls	10,145,698	41,058	1
Argialbolls	7,750,437	31,365	1
Durixerolls	6,806,457	27,545	1
Calciudolls	6,106,031	24,710	1
Palexerolls	4,687,101	18,968	1
Calcixerolls	4,000,851	16,191	0.4
Natrudolls	3,803,116	15,391	0.4
Epiaquolls	2,290,433	9,269	0.3
Natraquolls	1,890,157	7,649	0.2
Calcicryolls	1,632,530	6,607	0.2
Paleudolls	1,601,323	6,480	0.2
Palecryolls	1,281,435	5,186	0.1
Cryaquolls	1,219,124	4,934	0.1
Haprendolls	772,135	3,125	0.1
Cryrendolls	627,924	2,541	0.1
Natrixerolls	373,651	1,512	0.04
Durustolls	235,271	952	0.03
Vermudolls	119,888	485	0.01
Duraquolls	101,569	411	0.01
Duricryolls	35,949	145	0.00
Haplogelolls	30	0.1	0.00
Grand total	884,910,533	3,581,109	100

Source: NRCS Soil Survey Staff: http://soils.usda.gov/survey/

removed from mineral grains (primarily quartz) exposing their light color. In Albolls, the albic horizon typically occurs just below the mollic surface layer and directly above a less permeable horizon of clay accumulation, an argillic horizon. Albolls often have a water table at or near the surface for several months during the winter and spring. In the United States, Albolls are most extensive in loess deposits of the Midwest (Figure 33.46).

33.8.3.2 Aquolls

Aquolls are the sixth most commonly mapped Mollisol suborder (Table 33.38). They are the wettest Mollisols and are characterized by the presence of redoximorphic features below a thick and very dark brown or black epipedon. Depending on the duration of wet conditions and Fe reduction, expression of Fe-depleted zones can range from entire horizons with predominantly dull gray colors to horizons that have a strongly contrasting mottled gray color pattern. They typically occupy the lower lying areas of the landscape where water accumulates and a high water table exists. The most extensive areas of Aquolls in the United States occur in the Red River Valley of the North, along floodplains and terraces of major rivers in the central part, and along the Gulf coastal plain (Figure 33.46). Most uses of Aquolls, whether urban or agricultural, are limited by wetness. Many Aquolls are suitable for agriculture if artificially drained. Aquolls typically have black surface horizons underlain by gray colors indicating reduction and/or depletion of Fe (Table 33.40). Organic C accumulates in the surface horizon due to reduced rates of organic matter decomposition resulting from seasonal anaerobic conditions. The presence or absence of carbonates in Aquoll profiles is quite variable and can be used as an indicator of wetland hydrology. Carbonates tend to be removed from recharge wetlands by leaching, whereas discharge wetlands are enriched often to the soil surface. Depressional basins in subhumid climates often have recharge hydrology resulting in leaching of carbonates and formation of argillic horizons (Mausbach and Richardson, 1994).

33.8.3.3 Rendolls

Rendolls form in highly calcareous parent materials of humid or cold regions, often under forest vegetation. These soils typically consist of a mollic surface layer overlying mineral material that contains >40% $CaCO_3$ by weight (Soil Survey Staff, 2010). Rendolls are associated with parent materials such as chalk, limestone, highly calcareous glacial till, and shell deposits. These soils are not extensive in the United States where only a few soils, <1% of all Mollisols, have been recognized (Table 33.38). However, Rendolls are extensive in Western Europe (Smith, 1986).

33.8.3.4 Gelolls

Recent changes in *Soil Taxonomy* have made the presence of permafrost in the soil profile the first criterion for classification at the order level (Soil Survey Staff, 2010). Soils with permafrost within 100 cm of the surface, or within 200 cm with gelic materials, classify as Gelisols (see Section 33.6). Mollisols with a gelic soil temperature regime, that is, mean annual soil temperature (MAST) ≤0°C, are classified as Gelolls; these were formerly classified as

TABLE 33.40   Selected Morphological, Physical, and Chemical Properties of an Aquoll from Southern Minnesota

Horizon	Depth (cm)	Texture	Moist Color	Clay (%)	pH	Organic C (%)	Cation Exchange Capacity (cmol$_c$ kg^{-1})	CaCO$_3$ (%)
Ap	0–28	Clay loam	Black	37.5	7.5	4.5	45.5	1
A	28–41	Clay loam	Black	35.9	7.2	1.3	34.9	Trace
AB	41–55	Clay loam	Very dark gray	32.8	7.4	0.5	28.9	Trace
Bg	55–71	Clay loam	Olive gray	28.6	7.5	0.4	23.0	7
Cg1	71–105	Loam	Olive gray	26.3	7.5	0.2	20.5	11
Cg2	105–125	Loam	Gray	25.3	7.6	0.2	18.8	12

Location: Waseca County, Minnesota; mean annual air temperature: 6.2°C; mean annual precipitation: 836 mm; vegetation: Reed canary grass (*Phalaris arundinacea* L.), Cattails (*Typha* sp.); parent material: Wisconsinan-age glacial till; landscape position: depression, 0%–1% slope; classification: fine-loamy, mixed, mesic, Typic Endoaquoll.

Pergelic Cryoborolls (see the discussion on the fate of Borolls in Section 33.8.3.5). Gelolls are required to have a mean summer soil temperature (MSST) of ≤8°C, if no O horizon, or ≤5°C, if an O horizon is present; otherwise, they classify as Cryolls. Gelolls are rare and essentially not useful for agriculture. They are exclusively found at high elevations and latitudes. There is only one recognized soil series in the Geloll suborder; it is the Kanauguk series, a loamy-skeletal, mixed, mesic, superactive, subgelic Lithic Haplogeloll, found in Alaska. Consequently, only 30 ac (~0.1 km²) of Gelolls are represented in the NRCS database (Tables 33.38 and 33.39).

33.8.3.5 Cryolls

Cryolls occur in cold climates having a mean annual soil temperature <8°C with cool summers, but do not have permafrost. Cryolls are a replacement for the Cryoboroll great group of Borolls that was discontinued when the Boroll suborder was removed from *Soil Taxonomy* in 1998 (Soil Survey Staff, 1998). The fate of other Great Groups of Borolls is not as clear because not all soils that were classified formerly as Borolls have been reclassified and entered into the NRCS database. However, in addition to the placement of Cryoborolls (except Pergelic Cryoborolls, which are now Gelolls) into Cryolls, the convention for placement of former Borolls into the revised taxonomy is somewhat complicated. Borolls were defined as having a frigid, cryic, or pergelic temperature regime. As a result Borolls that had a pergelic regime and permafrost are now Gelisols, and those with a pergelic temperature regime, but no permafrost, are now Gelolls as discussed in Section 33.8.3.4. Those with a cryic temperature regime would have gone to Cryolls as discussed above. Those with a frigid temperature regime would have gone to Udolls and Ustolls but not to Xerolls because Xerolls preceded Borolls in the seventh edition key to *Soil Taxonomy* (Soil Survey Staff, 1975).

Cryolls are relatively rare, ~5% of mapped Mollisols in the United States (Table 33.38) and generally are associated with a short growing season that limits their use for agricultural production. They can, however, be productive for some cold-tolerant crops such as potatoes and for small grains and hay production where practicable. Large quantities of organic C are typically associated with such soils giving a strong black color to the mollic epipedon. Cryolls are generally found at higher latitudes or elevations. In fact, the most northerly Mollisol described in the literature is a loamy-skeletal, micaceous, isofrigid, Typic Haplocryoll found under tundra vegetation in Arctic Sweden at 68°23.32′N, 18°20.40′E, and at 775 m elevation on a 78% slope (Darmody et al., 2000). Its mean annual soil temperature of 1.2°C barely keeps it out of the Gelolls.

33.8.3.6 Xerolls

Xerolls form in areas that receive the majority of precipitation during the winter and spring months when temperatures are cool (Mediterranean climates). Little precipitation is received during the growing season and, as a result, Xerolls frequently become dry at some point during the summer. This type of moisture regime is very effective for leaching CaCO$_3$ and other soil constituents in higher precipitation zones. Conversely, Xerolls of the drier areas are subject to long periods of soil moisture deficits during the growing season. In the United States, Xerolls comprise ~15% of the land area occupied by Mollisols and are restricted to the western states (Table 33.38, Figure 33.46). An example of a Xeroll profile from the Palouse region of northern Idaho is shown in Figure 33.48, with selected data in Table 33.41. Important features of this soil include the lack of CaCO$_3$ in the profile, acidic conditions, and dense subsoil horizons (bulk density >1.7 g cm^{-3} at 94–152+ cm). Effective leaching by winter precipitation is responsible for the lack of CaCO$_3$ and the low pH. The dense subsoil horizons restrict downward water movement and result in the formation of perched water tables during the winter and spring months. Xerolls are used primarily for production of wheat, hay, and some timber.

33.8.3.7 Ustolls

Ustolls occur in subhumid or semiarid climates receiving significant amounts of spring and summer rain (350–900 mm annual precipitation), which is characteristic of the ustic soil moisture regime (Figure 33.46). As a result, the potential productivity of these soils for agriculture varies considerably. In the higher rainfall areas, drought-sensitive crops, such as corn and soybeans, can be produced, while in the drier regions, summer fallow, which allows recharge of soil moisture in a fallowed field (typically every other year), is often necessary for crop production (Soil Survey Staff, 1999). Ustolls are extensive throughout the central Great Plains and are the dominant suborder (~45% of the total Mollisol area) in the United States (Table 33.38). Data from
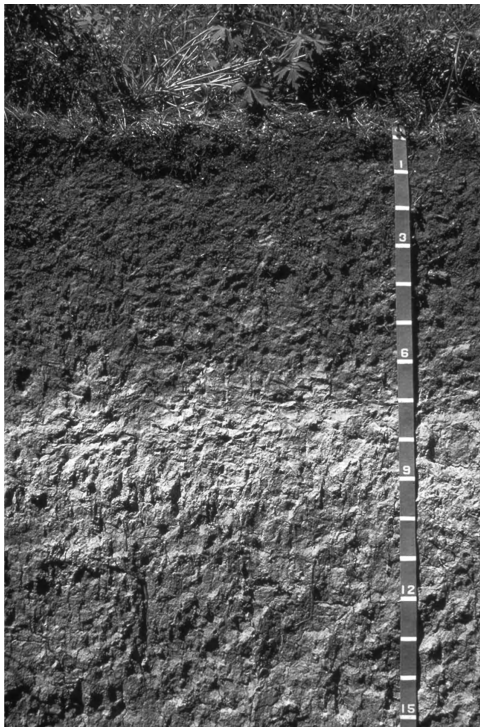
FIGURE 33.48 Soil profile of a Xeroll from northern Idaho. Accompanying data are in Table 33.41; scale is in decimeters.

a representative Argiustoll profile from Montana are presented in Table 33.42. Limited leaching in this profile has resulted in the removal of CaCO$_3$ from the upper 61 cm and the development of acid pH values near the surface. Base saturation increases with depth and is generally ~100% in lower horizons where CaCO$_3$ has accumulated. This particular pedon is from a frigid soil climate, and in this region, these soils are mainly used for winter wheat and hay production. However, often it is the dryness of Ustolls that limits agricultural productivity.

33.8.3.8 Udolls

Udolls occur in humid regions having typically formed in late Pleistocene or Holocene deposits under tall grass prairie

(Figure 33.46). They are the second most commonly mapped suborder, accounting for about 21% of the Mollisols in the NRCS database (Table 33.38). Udolls are characterized by having moderate precipitation, well-distributed throughout the year (Soil Survey Staff, 1999). Because extended soil moisture deficits are not experienced, and their high natural soil fertility, most Udolls have been placed into agricultural production. Udolls are extensive in the Midwestern United States and comprise one of the most productive grain-producing regions of the world (Soil Survey Staff, 1999). The famous Morrow Plots on the University of Illinois campus, where corn has grown continuously for well over 100 years, are developed in Udolls (Darmody and Peck, 1996). Corn and soybeans, which are among the principal crops produced on Udolls, are generally not produced or produced with higher risks, on other Mollisols because of climatic limitations. Data for a Udoll from Iowa are presented in Table 33.43. Although the parent material is calcareous loess, the relatively high annual precipitation has completely removed CaCO$_3$ from the upper 175 cm. Calcium is still the dominant soil cation and, with Mg, occupies most of the CEC.

33.8.4 Pedogenic Processes

Soils classified as Mollisols develop under a variety of environmental conditions through several distinct genetic pathways. Soil genesis typically occurs over time frames where environmental conditions, and hence the theoretical equilibrium state, are variable. The morphological imprints left by grassland vegetation are ephemeral in pedogenic time frames, and hence, are probably indicative of the soil climate over the past few hundreds to thousands of years. The genesis of Mollisols will be discussed from two different perspectives (factors and processes of soil formation) in an attempt to elucidate the development of these unique soils.

33.8.4.1 Factors of Soil Formation

The factor approach applied by Jenny (1941, 1980) to the soil-forming factor model is predicated on the hypothesis that soil properties result from the interaction of at least five site and flux factors. This model proposes that soil bodies result from

TABLE 33.41 Selected Morphological, Physical, and Chemical Properties of a Xeroll (Southwick Series, 93-ID-29151) from Northern Idaho

Horizon	Depth (cm)	Moist Color	Sand (%)	Silt (%)	Clay (%)	Bulk Density (g cm^{-3})	pH	Organic C (%)
A	0–18	Very dark brown	6.9	72.2	21.0	—[a]	6.5	3.09
AB	18–38	Very dark grayish brown	7.9	72.7	19.3	1.25	5.9	1.69
Bw	36–71	Dark brown	7.7	70.0	22.3	1.30	5.8	1.20
BE	71–58	Brown	7.3	74.0	18.7	1.37	5.6	0.46
E	58–94	Grayish brown	8.7	78.3	13.0	1.48	5.5	0.46
Btb1	94–114	Brown	5.5	63.2	31.3	1.76	5.2	0.42
Btb2	114–152+	Brown	6.6	63.8	29.6	1.75	5.6	0.17

Location: 46° 44′N Lat., 116° 50′W Long., Latah County, Idaho; 13 km east of Moscow; elevation: 825 m; mean annual air temperature: 7°C; mean annual precipitation: 635 mm; vegetation: ponderosa pine, Idaho fescue, snowberry; parent material: late Pleistocene and Holocene loess; landscape position: near summit of gently sloping (4%) spur ridge; classification: fine-silty, mixed, superactive, mesic Oxyaquic argixeroll.

[a] Not sampled.

TABLE 33.42   Selected Morphological, Physical, and Chemical Properties of a Pedon Mapped as Shawmut, a loamy-skeletal, mixed, superactive, frigid Typic Argiustoll (S71 MT 31-1) from Montana

Horizon	Depth (cm)	Moist Color	Clay (%)	pH	Organic C (%)	Cation Exchange Capacity (cmol$_c$ kg^{-1})	CaCO$_3$ (%)	Base Saturation (%)
A1	0–6	Black	28.1	5.5	9.88	43.7	nd[a]	70
A2	6–18	Black	28.6	5.8	5.27	36.8	nd	80
Bt1	18–29	Dark brown	32.8	6.1	1.88	32.1	nd	84
Bt2	29–42	Brown	28.9	6.4	1.11	32.4	nd	100
Bt3	42–61	Brown	27.1	7.0	0.87	34.0	nd	100
Bk1	61–86	Brown	25.0	7.7	0.55	27.2	11	100
Bk2	86–130	Brown	26.2	8.0	0.28	26.8	8	100
2 C	130–175	Brown	26.3	8.2	0.21	25.6	2	100

Sources: Data Courtesy of Dr. G.A. Nielsen, Montana State University, Emeritus; From Soil Survey Staff, *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 2nd edn., USDA Agriculture Handbook No. 436, U.S. Government Printing Office, Washington, DC, 1999.

Location: Gallatin County, Montana; elevation: 1585 m; mean annual air temperature: 5°C; mean annual precipitation: 510 mm; vegetation: bluebunch wheatgrass, Ldaho fescue, prairie junegrass, sagebrush; parent material: mixed alluvium/colluvium; landscape position: alluvial/colluvial fan, 4% slope.

   [a] Not detected.

TABLE 33.43   Selected Morphological, Physical, and Chemical Properties of a Udoll (S63 Iowa-83-2) from Iowa

Horizon	Depth (cm)	Texture	Moist Color	Clay (%)	pH	Organic C (%)	Cation Exchange Capacity (cmol$_c$ kg^{-1})	Exch. Ca (cmol$_c$ kg^{-1})
Ap	0–18	Silty clay loam	Very dark brown	30.4	5.6	2.20	22.0	13.9
A	18–33	Silty clay loam	Very dark brown	33.5	5.7	1.87	22.9	14.7
AB	33–46	Silty clay loam	Very dark grayish brown	32.8	5.8	1.11	21.6	14.8
Bw1	46–69	Silty clay loam	Dark brown/brown	30.4	5.8	0.58	20.0	14.8
Bw2	69–86	Silty clay loam	Dark brown/brown	28.2	5.9	0.33	20.7	14.7
BC1	86–110	Silt loam	Yellowish brown	26.9	5.9	0.21	20.4	14.6
BC2	110–125	Silt loam/silty clay loam	Yellowish brown/olive gray	28.0	6.0	0.17	20.7	15.2
C1	125–145	Silt loam	Brown/olive gray	26.9	6.0	0.11	20.8	14.6
C2	145–175	Silt loam	Yellowish brown/olive gray	25.7	6.2	0.10	19.7	13.8

Sources: Data from Soil Survey Staff, *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, USDA Agriculture Handbook No. 436, U.S. Government Printing Office, Washington, DC, 1975, Appendix 4; Data courtesy of USDA-Natural Resources Conservation Service, Soil Survey Division, World Soil Resources, 2009.

Location: Shelby County, Iowa; 135 km West of Des Moines; mean annual air temperature: 9°C; mean annual precipitation: 720 mm; vegetation: cropland; parent material: late Pleistocene loess; landscape position: axis of short interfluve, 3% slope; classification: fine-silty, mixed, mesic Typic Hapludoll.

the action of climate, organisms, and relief acting upon parent material over time and provides a useful conceptual framework by which soil differences can be attributed to state or site factors.

33.8.4.1.1 Organisms

To a large degree, Mollisols are distinguished from other soils on the basis of organisms, or more specifically, the overriding effects of grassland vegetation. The reason why grassland vegetation imparts distinctive soil morphologies is discussed later. While some Mollisols have developed under consistent vegetative conditions, abundant evidence suggests that many were forested during glacial and postglacial periods (Walker, 1966; Ruhe, 1970; Wright, 1970). Similar evidence also suggests that the climate has changed sufficiently during the Holocene to allow fluctuations between forest and prairie vegetation, especially along humid to subhumid ecotones (Walker, 1966; Geis and Boggess, 1968). Most soils classified as Mollisols (with the possible exception of some Rendolls, Aquolls, and Albolls) have

probably had grassland vegetation at some time during their development. Because soil characteristics reflect the integration of the soil-forming factors over long time periods (centuries to millennia), the contemporary environment may or may not accurately reflect the conditions under which a soil has developed. Mollisols are common in transitional areas between humid and subhumid environments such as the north central United States. Changes in climate during the Holocene have probably resulted in the formation and subsequent degradation of Mollisols as forest encroached on areas that were previously prairies (Fenton, 1983).

The mixing and incorporation of organic matter within the mollic epipedon have also been attributed to soil faunal activity. Of particular importance are burrowing organisms, such as gophers, prairie dogs, worms, and ants (Thorp, 1948; Munn, 1993). The net effect of burrowing activities is to move organic matter lower in the soil profile and to mix soil materials, resulting in a rather homogeneous epipedon.

33.8.4.1.2 Climate

As previously suggested, climate is also a key factor in the formation of Mollisols, primarily through its effect on vegetation. Grassland ecosystems are associated with broad ranges in mean annual precipitation and temperature. Consequently, Mollisols occur under a wide range of soil temperature and moisture regimes. In the United Sates, these soils are found from as far south as Texas to the Canadian border and into Alaska. Mollisols are most common in the zone between the arid deserts of the Western states and the forests of the Midwestern states, a range of approximately 300 mm to more than 900 mm in annual precipitation (Figure 33.46).

The effects of climate on Mollisol formation can be demonstrated by comparing soil morphology along climatic gradients. Munn et al. (1978) examined a warm/dry-to-cool/moist sequence of grassland soils in Montana. They observed an increase in annual aboveground biomass production with increasing precipitation and decreasing temperature (Figure 33.49). Similarly, SOM content, which increased as well, is expressed by increasing thickness of the mollic epipedon. This suggests that the dark, organically enriched mollic epipedon provides a record of the balance between annual site production and microbial decomposition. As production increases with increasing precipitation, decreasing temperature favors the accumulation of SOM by decreasing microbial decomposition rates. Leaching of $CaCO_3$ in the soils was greater with increasing annual precipitation, indicating that depth to carbonate serves as an indicator of rainfall within a geographic region.

33.8.4.1.3 Relief

The local topography or hillslope position affects hydrologic processes, especially erosion and sedimentation and the timing, duration, and depth of seasonal soil saturation. Many glacial till landscapes of the eastern portion of the central United States
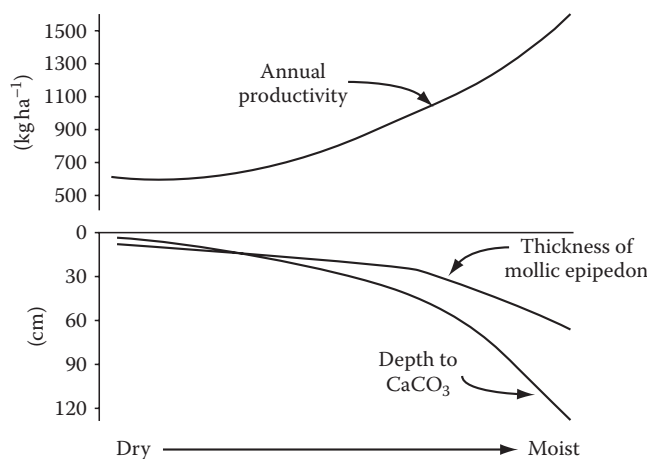
FIGURE 33.50 Glacial till landscape in central Indiana (the easternmost portion of the "prairie peninsula") showing patterns of Alfisols (light tones) and Mollisols (dark tones). Wet Mollisols (Aquolls) occupy lower lying, wetter portions of the landscape. Alfisols are found in the higher, better drained positions. (Photo used with permission of the Soil Science Society of America.)

contain both Mollisols and Alfisols, the "Prairie Peninsula" (Geis and Boggess, 1968). Aquolls and Albolls occur in the wetter, lower lying portions of the landscape (darker soil surface; Figure 33.50). Conversely, Alfisols occupy the adjacent higher positions having better drainage (Allen and Fanning, 1983).

Richardson et al. (1992) proposed a framework to view wetland hydrology based on differences in how water flows through depressional basins (recharge, flowthrough, or discharge). Differences in basin hydrology result in distinctive differences in soil morphology along hillslopes (Thompson and Bell, 1996; Bell and Richardson, 1997). Examination of a toposequence of Mollisols from south central Minnesota illustrates typical catena relationships (Figure 33.51). Along this hillslope continuum, the following changes occur from upper to lower landscape positions: (1) increased darkness (organic C) and thickness of the surface

FIGURE 33.49 Generalized relationship between aboveground productivity in rangelands in Western Montana, depth to $CaCO_3$, and thickness of mollic epipedon. (Adapted from Munn, L.C., G.A. Nielsen, and W.F. Mueggler. 1978. Relationships of soils to mountain and foothill range habitat types and production in western Montana. Soil Sci. Soc. Am. J. 42:135–139. With permission of the Soil Science Society of America.)

FIGURE 33.51 Toposequence of soil profiles along a hillslope in glacial tills near Waseca, Minnesota.
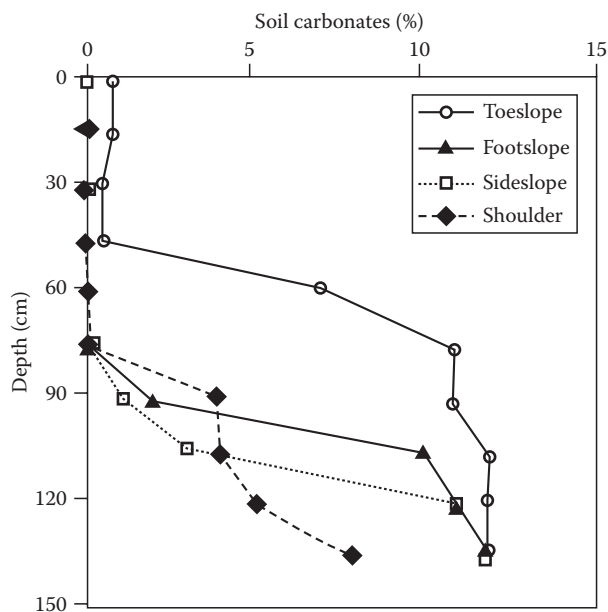
FIGURE 33.52 Depth distributions of soil carbonate concentrations at four hillslope positions in glacial till near Waseca, Minnesota.

FIGURE 33.53 Depth distributions of soil organic carbon at four hillslope positions in glacial till near Waseca, Minnesota.

horizon, (2) decreased depth to redoximorphic features, and (3) decreased chroma of the subsurface horizons. These morphological differences can often be traced back to differences in soil processes, such as erosion, deposition, and reduction that are affected by hydrology and the shape and position on the hillslope. Numerous studies have documented increased duration of saturation in lower landscape positions for similar landscapes (Khan and Fenton, 1994; Bell et al., 1995; Thompson and Bell, 1996). For the example from Waseca, Minnesota (Figure 33.52), higher carbonate concentrations between 45 and 90 cm at the toeslope are probably indicative of discharge hydrology and subsequent enrichment of the soil profile from carbonate laden groundwaters. The distribution of organic C with depth along the hillslope clearly indicates a trend of increasing organic C in the surface horizons from upper to lower hillslope positions (Figure 33.53). Organic C concentrations tend to converge at a depth of approximately 75 cm for all landscape positions where concentrations fall below 0.5%. These same spatial relationships cannot be assumed for all hillslopes as differences in soil stratigraphy, regional hydrology, and climate can cause different spatial patterns to develop.

33.8.4.1.4 Parent Material

Mollisols occur on a variety of substrate materials, but are most commonly associated with unconsolidated sediments from coastal, riverine, or glacial depositional environments, including loess (Soil Survey Staff, 1999). Loess is a common parent material of Mollisols in the Midwestern United States and in southern and central Russia. Duchaufour (1982) has suggested that the permeability and $CaCO_3$ content of loess is especially conducive to the formation of Mollisols. Many Mollisols have formed in calcareous parent materials resulting in appreciable quantities of $CaCO_3$ within the soil profile, particularly those Mollisols in
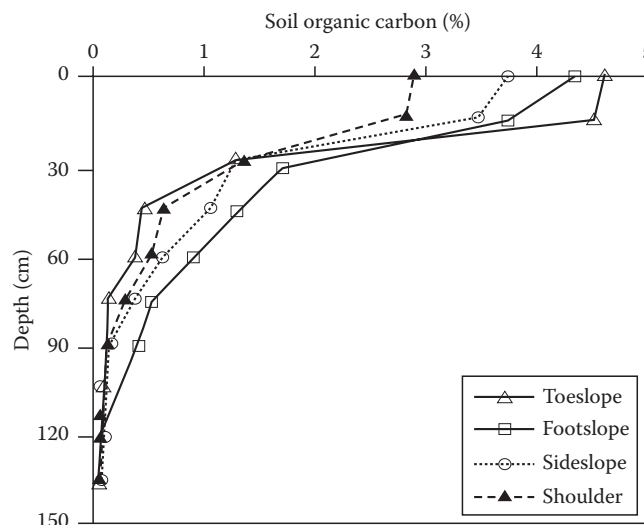
lower precipitation areas. Rendolls form exclusively in highly calcareous parent materials such as chalk, limestone, and shell deposits.

33.8.4.1.5 Time

Most Mollisols in the United States and Russia appear to be relatively young and have formed on late Pleistocene and Holocene deposits and surfaces (Soil Survey Staff, 1999). However, the recognition of Mollisols in intertropical areas may warrant reexamination of this concept. It has been shown that accumulation of SOM occurs relatively fast. Schafer et al. (1980) compared organic C levels in newly created mine soils with those of adjacent undisturbed grassland soils of eastern Montana and found that only 30 years was required for C to build up in the top 10 cm to the level in undisturbed soils. In Iowa, the rate of formation of a Udoll surface horizon was estimated at 0.08 cm year^{-1} (Buol et al., 2003). Foss et al. (1985) suggested that mollic epipedons developed in <900 years in the Red River Valley of the North in Minnesota and North Dakota. From these and other studies, it is clear that the formation of a mollic surface horizon can occur in relatively short time spans.

33.8.4.2 Genetic Pathways for Mollisol Formation (Soil Processes)

The accumulation of organic C in the upper soil profile is the salient morphological feature distinguishing Mollisols from other soil orders. The C content is high when the long-term rate of addition and retention exceeds that of decomposition. Rates of annual root production and decomposition in grasslands result in high rates of organic C turnover in the soil to depths often approaching 50–100 cm (Dahlman and Kucera, 1965; Dormaar and Sauerbeck, 1983). Thorp (1948) estimated that 113–409 kg ha^{-1} of raw OM (dry weight) were added annually for short grass and 136–500 kg ha^{-1} for tall grass prairies. Alternatively, soil organic C accumulates when rates of microbial decomposition are low due to either anaerobic (wet) conditions or cool
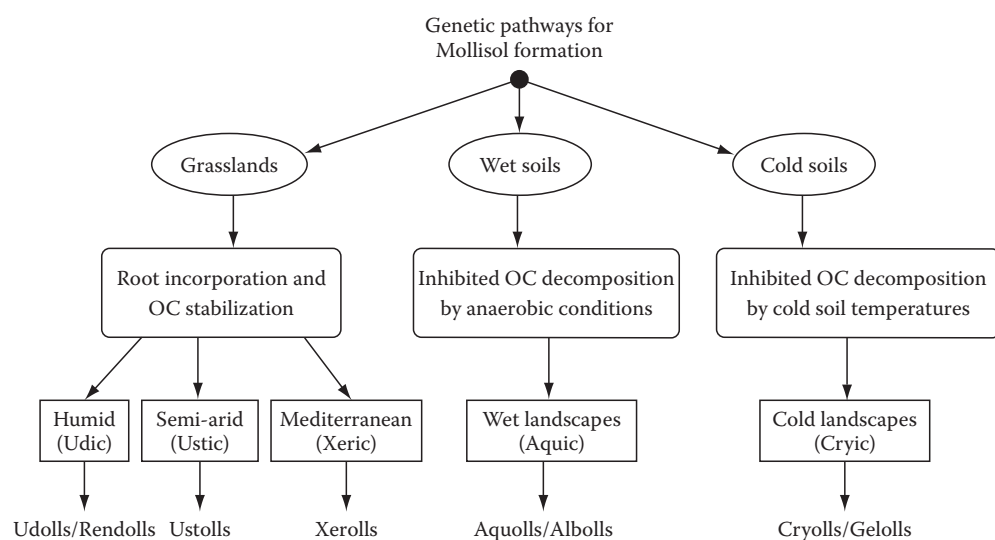
FIGURE 33.54 Theoretical pathways for Mollisol genesis.

soil temperatures (Jenny, 1930). Three primary pathways for Mollisol genesis based on the mode of organic C accumulation can be identified: (1) high rates of accumulation in grasslands, (2) low rates of decomposition under anaerobic (wet) conditions, and (3) low rates of decomposition in cold climates (Figure 33.54). These pathways are not necessarily mutually exclusive. For example, many wet soils may have also developed under grassland, but wet grassland soils frequently have higher organic C contents than drier soils in the same landscape (Figures 33.50 and 33.53).

33.8.4.2.1 Grassland Mollisols (Udolls, Ustolls, and Xerolls)

A comparison of the chemical composition and growth forms of forests and grasslands explains why soils under these covers have distinctive morphologies. In forests, organic C tends to accumulate on the soil surface from annual decomposition cycles of forest floor litter. Despite the fact that the forested soils (Alfisols) receive more annual precipitation, they contain less organic C throughout the mineral soil and lack the dark soil colors associated with Mollisols. Much of the organic matter associated with the forest soils is contained in litter layers lying on top of the mineral soil. Leachate from the decomposition of forest floor litter is often composed of carbonic, fulvic, and/or tannic acids facilitating translocation (eluviation) of soil materials deeper within the soil profile resulting in morphological features usually associated with Alfisols and Spodosols. These conditions and processes inhibit the accumulation and stabilization of organic C in the upper soil layers.

By contrast, photosynthesis and other metabolic processes in grassland vegetation quickly transport organic C to dense and fibrous root systems. Estimates indicate that grasslands have the highest annual additions of C to soil of any of the terrestrial ecosystems, including tropical forests (Bolin et al., 1979). Related research has demonstrated that the greatest biomass production in grassland ecosystems is below ground (Caldwell, 1975; Lauenroth and Whitman, 1977; Jenny, 1980). As such, Mollisols
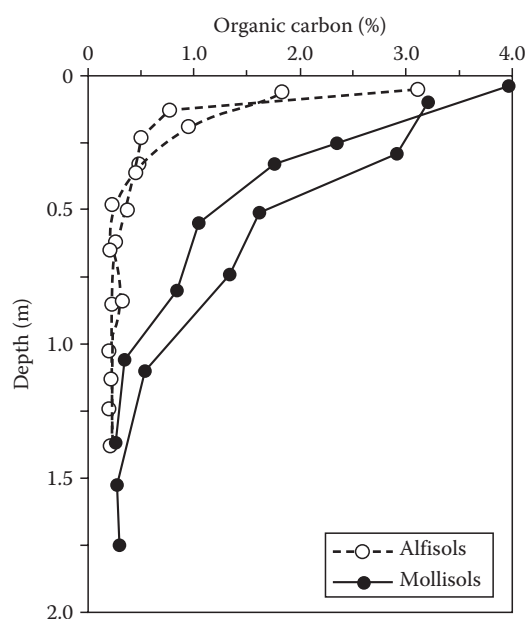
FIGURE 33.55 Depth distribution of soil organic carbon in Mollisols and Alfisols of the Palouse region of Northern Idaho. The Mollisols have formed under Idaho fescue grassland with approximately 580 mm of annual precipitation; the Alfisols have formed under grand fir forest with approximately 830 mm of annual precipitation. (Data courtesy of the Soil Characterization Laboratory, University of Idaho, Moscow, ID.)

have distinctive organic matter profiles with depth that differ markedly from Alfisols, where most of the biomass is produced aboveground (Figure 33.55). The annual proliferation of roots and their subsequent decomposition is responsible for the deep accumulation of SOM leading to the formation of mollic epipedons whose thickness is largely determined by the depth and amount of grass roots (Hole and Nielsen, 1968; Cannon and Nielsen, 1984). Incorporation of organic C is also facilitated by the mixing of near-surface soil horizons by ants (Formicidae),

earthworms (*Lumbricus*), and other soil fauna (Curtis, 1959; Baxter and Hole, 1967). Schlesinger (1977, 1991) have estimated that temperate grassland soils have a mean organic matter content of 19.2 kg C m^{-2}, ranking them behind only soils of wetlands and tundra/alpine ecosystems on a global scale.

Grasses have another effect that differs from forest vegetation on soil genesis. Grasses are generally better base cyclers. That, in part, accounts for the higher fertility of Mollisols when compared with forest soils such as Alfisols. In addition, that difference also accounts for the absence of fragipans in Mollisols (Franzmeier et al., 1989).

Once added to the soil, organic materials undergo further decomposition by complex, microbially mediated processes that results in the formation of a relatively stable organic fraction. The large annual additions of C to Mollisols and subsequent cycling result in the formation of an active (bioavailable) as well as a very stable organic fraction. The humus contained in Mollisols appears to be more stable than that found in other soils (Stevenson, 1994), which is possibly related to the chemical composition of grasses and soil parent material. Grasses have a higher ratio of humic to fulvic acids (Glazovskaya, 1985). Novak and Smeck (1991) found higher concentrations of humic substances in the surface horizons of Mollisols compared with Alfisols in southwestern Ohio where these soils are adjacent in the same landscape. Minimal differences were found in fulvic acid content. The combination of humic substance in the presence of calcareous soil parent material results in the formation of Ca humates that are thought to bind silicate grains to organic C, which becomes stabilized (Evans and Russell, 1959; Stevenson, 1994). Formation of these complexes increases the resistance of SOM to physical disintegration, chemical extraction, and further biological change through microbial decomposition (Oades, 1989; Stevenson, 1994). Estimates as to the extent of organic matter–mineral associations in grassland soils vary. As much as 80% of the organic C in a mollic epipedon may be so closely associated with the mineral fraction that it cannot be separated by physical means (Greenland, 1965; McKeague et al., 1986). Researchers have been able to determine mean residence time (MRT) of organic fractions of Mollisols in several different environments. The MRT represents the average length of time that an organic fraction has been present in the soil. Oldest MRT values measured in soils are associated with Mollisols and Histosols (Oades, 1989). For example, MRT values between 1255 and 2973 years have been measured for Mollisols in Canada and the United States (Anderson and Paul, 1984; Hsieh, 1992).

The stabilization of organic C in mollic epipedons is also facilitated by soil clays. Organic C associated with fine clays is protected from further rapid decomposition lengthening the turnover time of otherwise labile humic substances from days to months, years, or even decades (Anderson and Paul, 1984). Anderson (1979) found that as much as 50% of the total humus in some grasslands is associated with clay. From these and other studies, it appears that texture is an important factor in determining the stable level of organic C in Mollisols as well as in other soils. Texture influences the water-holding capacity of a soil and the quantity of clays available to form
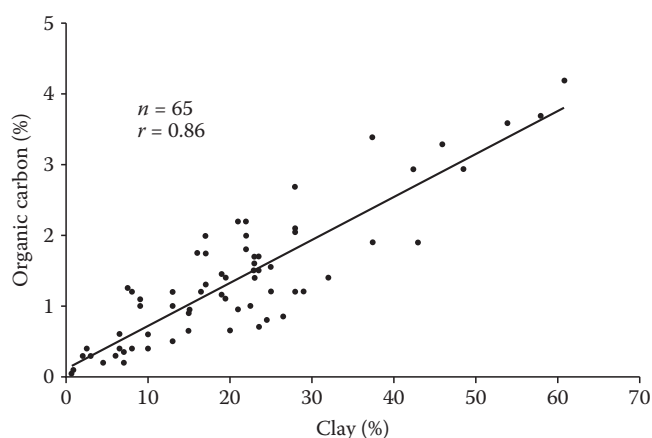
FIGURE 33.56 Relationship between clay and organic matter content in warm grassland soils of the southern United States. (Adapted from Nichols, J.D. 1984. Relation of organic carbon to soil properties and climate in the southern Great Plains. Soil Sci. Soc. Am. J. 48:1382–1384. With permission of the Soil Science Society of America.)

complexes with organic matter. Nichols (1984) observed a good correlation ($r = 0.86$) between soil organic C and clay content in 65 Mollisols and associated soils of Texas, Oklahoma, and New Mexico where mean annual temperatures exceed ~15°C (Figure 33.56). These data indicate that water-holding capacity and the protection afforded to humus through formation of mineral–organic complexes may control the equilibrium organic C contents in Mollisols of warmer regions. In contrast, clay content does not appear to exert such strong control of organic C contents in Mollisols of cooler regions. McDaniel and Munn (1985) found little correlation between SOM and clay in 137 Mollisols and associated soils in Montana and Wyoming having mean annual soil temperatures <8°C (Figure 33.57). Comparison of these data with those of Nichols (1984) suggests that it is temperature rather than moisture that becomes the determining factor in establishing an equilibrium SOM content. Furthermore, organic matter may be able to
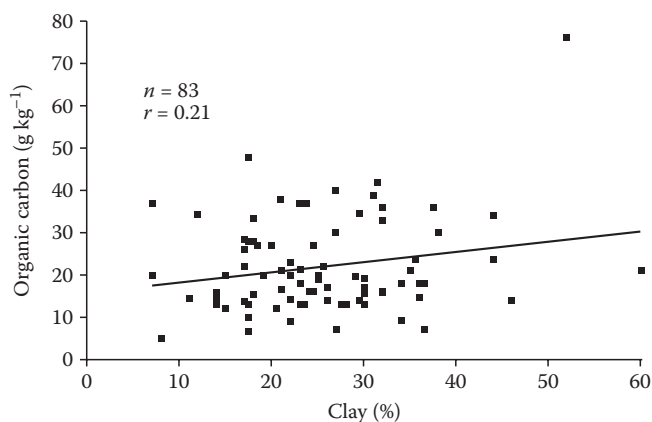
FIGURE 33.57 Relationship between clay and organic matter content in cool grassland soils of the northern United States. (Adapted from McDaniel, P.A., and L.C. Munn. 1985. Effect of temperature on organic carbon–texture relationships in Mollisols and Aridisols. Soil Sci. Soc. Am. J. 49:1486–1489. With permission of the Soil Science Society of America.)

persist in cold soils without the stabilizing influence afforded by complexation with clays (McDaniel and Munn, 1985).

33.8.4.2.2 Wet Mollisols (Albolls and Aquolls)

An alternate genetic pathway for the development of Mollisols involves organic C accumulation under anaerobic soil conditions due to reduced rates of organic matter decomposition (Ponnamperuma, 1972; Gambrell and Patrick, 1978; Tate, 1980). These Mollisols have an aquic soil moisture regime and are classified as Aquolls or Albolls. The decomposition of organic matter is less efficient under anaerobic conditions, and a net accumulation of SOM will occur if rates of biomass production are sufficiently high. If this rate of organic matter accumulation is high enough, organic soils (Histosols) will develop. In many landscapes, Mollisols are found at the transition between organic and mineral soils.

In general, Aquolls have black (N 2/0) surface horizons with predominant gleyed or depleted horizons directly beneath the surface horizon (Table 33.40). Increases in the duration of soil saturation and anaerobic conditions are usually associated with higher soil organic C contents (Khan and Fenton, 1994; Bell et al., 1995). Topographically, organic C increases downslope as evidenced by the darker soil surfaces in lower landscape positions (Figure 33.51) associated with higher surface concentrations of organic C in depressional areas (Figure 33.53). Thompson and Bell (1996) found good agreement between a profile darkness index based on the thickness and darkness of the surface horizons and the duration of saturation for soil hydrosequences at several locations in the United States. They also found that the profile darkness index was highly correlated with soil organic C for the soils that were derived from glacial deposits.

Soil saturation, the development of anaerobic conditions and subsequent reduction of Fe(III) to Fe(II), causes the formation of distinctive color patterns (redoximorphic features) in the horizons beneath the mollic epipedon. Biochemical reduction results in the translocation of Fe compounds, which are the primary determinants of soil color in low organic matter horizons. The processes involved in the development of redoximorphic features are discussed in Chapter 15. If the soil water is stagnant, Fe(II), which has a distinctive bluish or greenish color, may be present, thus creating a reduced (or gleyed) soil matrix. More commonly, the soil water is moving, and Fe(III) coatings on mineral grains will be reduced and removed as Fe^{2+} ions, leaving the mineral grains and revealing the dull, gray color of the uncoated minerals, thus forming redoximorphic depletions. Under conditions of fluctuating water tables, Fe(II) may be reoxidized creating discrete concentrations of orange or reddish Fe(III) (Vepraskas and Sprecher, 1997). Bell and Richardson (1997) discuss specific soil process and morphologies associated with Aquolls and Albolls.

33.8.4.2.3 Cold Mollisols (Gelolls and Cryolls)

The formation of Mollisols under cold conditions (i.e., cryic or gelic soil temperature regimes) is somewhat analogous to the formation of wet Mollisols. Low soil temperatures (similar to lack of soil O_2) reduce microbial activity and facilitate organic matter

accumulation if the long-term rate of biomass production exceeds that of decomposition. Again, as previously covered, where permafrost is within 2 m of the soil surface, the soils are classified as Gelisols, and further classified as Histels if they are dominantly composed of organic soil materials (see Section 33.2). Where permafrost is lacking, or deeper than 2 m, and where rates of organic matter accumulation are high, the soils are dominated by organic matter and are in Cryo great groups of Histosols (Cryofibrists, Cryofolists, Cryohemists, and Cryosaprists). Cold Mollisols, which are classified as Cryolls or Gelolls, develop in areas where organic matter accumulation rates are somewhat lower than are necessary for organic soil materials to accumulate. Gelolls and Cryolls, along with their wet but warmer analogs, Aquolls and Albolls, often develop under grasslands, or tundra-like vegetation in the case of the cold Mollisols, but can develop under other types of vegetation where conditions are favorable.

33.8.4.3 Associated Pedogenic Processes

The following processes are also important in the formation of some Mollisols.

33.8.4.3.1 Carbonate Translocation

The dissolution of $CaCO_3$ and its subsequent precipitation as secondary carbonates lower in the soil profile is common in many calcareous parent materials where sufficient leaching takes place. Carbonates in layers of maximum accumulation are more finely divided than those associated with the parent material supporting secondary carbonate formation (Redmond and McClelland, 1959). Total precipitation is important in determining the depth at which $CaCO_3$ is deposited (Jenny, 1941). In many Xerolls, where precipitation is low but exceeds evapotranspiration especially in winter, $CaCO_3$ is readily mobilized and reprecipitated at depth. In the Palouse region of eastern Washington and northern Idaho, $CaCO_3$ has been removed from the upper 1.5 m of the profile in Xerolls receiving more than 530 mm of rainfall (Barker, 1981). In contrast, Ustolls of the Great Plains receiving comparable amounts of total rainfall typically have well-developed zones of $CaCO_3$ accumulation within 1 m of the soil surface. The Ustolls receive much of their precipitation during the growing season when evapotranspirational demand is high, resulting in less moisture being available for mobilization of $CaCO_3$. As previously discussed, spatial patterns of carbonate removal and accumulation along hillslopes can often be used to interpret soil hydrology.

33.8.4.3.2 Clay Translocation

The movement and accumulation of clay are common soil-forming processes that occur in soils occupying relatively stable landscape positions and depression-focused recharge depressional basins. Clay translocation causes differences in soil texture that can affect soil water movement and the subsequent genesis of certain soil morphological features. This process is thought to be important in the formation of Albolls. Appreciable clay movement does not occur until carbonates have been removed from the upper portions of the soil profile (Duchaufour, 1982; Fanning

and Fanning, 1989). Carbonate removal releases clay previously cemented, which may be subject to further chemical alteration prior to movement (Fenton, 1983). In calcareous parent materials, subsoil clay accumulation commonly occurs immediately above the zone of $CaCO_3$ accumulation. In Mollisols, subsurface clay accumulation is highly variable and is not always observed.

In a study of Mollisols and associated soils in the forest steppe–dry steppe transition in Eastern Europe and the forest–prairie transition in the United States Great Plains, Bronger (1991) was unable to find micromorphological evidence of clay illuviation in Mollisols commonly described as having argillic horizons. While Nettleton et al. (1969) proposed that the lack of illuvial clay skins in some fine-textured soils was due to disruption by shrink/swell clays, Bronger (1991) suggested that increased clay contents in Mollisol subsoils are either lithologic discontinuities or are polygenetic in origin and, as such, are relict features that formed under past, moister climates. Sobecki and Wilding (1983) investigated Texas Coast Prairie Mollisols and found that argillic horizons were confined to microtopographic lows, suggesting depression-focused recharge. Weakly developed Bt horizons on microtopographic highs were determined to be relict features based on micromorphological analysis. They suggested that processes leading to both carbonate accumulation and clay illuviation in the same horizon were incompatible under the current humid climate.

33.8.4.3.3 *Erosion and Sedimentation*

Erosional processes result in the redistribution of surface soil material from upper to lower hillslope positions. This process often results in thickening of the mollic epipedons in concave slope positions. Soils with these overthickened surface horizons are often classified as cumulic at the subgroup level. Anthropogenically accelerated erosion causes significant reduction in the thickness of upland A horizons, which no longer meet the surface horizon thickness criteria for Mollisols. These eroded soils will typically be classified as Inceptisols or Alfisols (if an argillic horizon is present). Erosional losses as a genetic pathway for Alfisols or Inceptisols, which is contrary to the conceptual models for the genesis of these soils, has been a long-standing taxonomic dilemma in the classification of Mollisols (Fenton, 1983). The choice presented to soil mappers is between recognizing eroded phases of Mollisols, or classifying the soils based on the thinner epipedon resulting from anthropogenically accelerated soil erosion.

33.8.5 Land Use

Human impact on the use of Mollisols has increased dramatically over the past century. Early cultures had little influence on grassland soils other than the use of fire as a management tool. In the plains of North America, burning was used to stimulate the subsequent year's production and attract buffalo (Warkentin, 1969; Anderson, 1987). With the advent of the moldboard plow, Mollisols became widely used as arable soils. During the early 1900s, artificial drainage (tile drains and surface ditching) and

flood protection measures lowered water tables sufficiently for the conversion of extensive areas of Aquolls and other marginally wet Mollisols to productive cropland. In the United States, where Mollisols make up 22% of the land area, much of the agriculturally based, westward settlement was a direct result of the ease with which productive Mollisols could be cleared by fire. The high level of native fertility and favorable physical properties meant that most areas of Mollisols were converted to row cropping. Virgin areas of Mollisols and associated prairies are now quite rare in the United States.

Mollisols are among the most agriculturally productive soils in the world. Although Mollisols make up only ~7% of the Earth's ice-free land area, a large part of the world's wheat and other small grains are grown on Ustolls and Xerolls (Troeh and Thompson, 1993). Because Mollisols are commonly found in subhumid climates, lack of sufficient soil moisture can limit the production of traditional agricultural crops. Soil moisture conservation strategies, such as fallowing and residue management, must be implemented for sustainable dryland farming. Unfortunately, the use of unsustainable farming practices not designed to protect the soil from erosion or to conserve soil organic C has led to a decline in soil quality and subsequent productivity in some regions of the world. Long-term cultivation has reduced the organic C content of Mollisols, and reductions of ~35% in C contents in 60–70 years of cultivation have been documented (Tiessen et al., 1982). These reductions are often associated with degradation of soil structure and increased susceptibility to erosion. More recent research into the impacts of agriculture on the nearly level Aquolls and Udolls in Illinois indicated that these soils are quite resilient. After an initial loss of SOM when the native prairie is cultivated, they retain much of their organic C even after a century of row crop production (David et al., 2009). However, in most cases, the conversion of prairie to cropland further increases the erosion potential and has resulted in devastating wind erosion during cyclic droughts in portions of Asia and North America, most famously in the serious wind erosion of Ustolls during the Dust Bowl in the 1930s. These processes of soil degradation result in loss of organic matter and organically bound nutrients and subsequent declines in soil productivity. The use of appropriate soil conservation practices can greatly reduce soil degradation and maintain soil productivity.

As population pressure has increased, conversion of Mollisols from arable to urban uses increased. Many of the same characteristics that make Mollisols preferable for cropland also favor many nonagricultural uses. The conversion of these prime agricultural soils, and the associated decline in the resource base for food and fiber production, has generated much debate in the United States and stimulated farmland preservation efforts.

Acknowledgments

References

Allen, B.L., and D.S. Fanning. 1983. Composition and soil genesis, p. 141–192. *In* L.P. Wilding et al. (eds.) Pedogenesis and soil taxonomy. I. Concepts and interactions. Elsevier, Amsterdam, the Netherlands.

Allen, B.L., and B.F. Hajek. 1989. Mineral occurrence in soil environments, p. 199–278. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Anderson, D.W. 1979. Processes of humus formation and transformations in soils of the Canadian Great Plains. J. Soil Sci. 30:79–84.

Anderson, D.W. 1987. Pedogenesis of grassland and adjacent forests of the Great Plains. Adv. Soil Sci. 7:53–93.

Anderson, D.W., and E.A. Paul. 1984. Organo-mineral complexes and their study by radiocarbon dating. Soil Sci. Soc. Am. J. 48:298–301.

Barker, R.J. 1981. Soil survey of Latah County area, Idaho. USDA-SCS, U.S. Government Printing Office, Washington, DC.

Baxter, F.P., and F.D. Hole. 1967. Ant (*Formica cinerea*) pedoturbation in a prairie soil. Soil Sci. Soc. Am. Proc. 31:425–428.

Bell, J.C., and J.L. Richardson. 1997. Aquic conditions and hydric soil indicators for Aquolls and Albolls, p. 23–40. *In* M.J. Vepraskas and S. Sprecher (eds.) Aquic conditions and hydric soils: The problem soils. SSSA, Madison, WI.

Bell, J.C., J.A. Thompson, and C.A. Butler. 1995. Morphological indicators of seasonally saturated soils for a hydrosequence in southeastern Minnesota. J. Minn. Acad. Sci. 59:25–34.

Bolin, B.E., T. Degens, P. Duvigneaud, and S. Kempe. 1979. The global biogeochemical carbon cycle, p. 1–56. *In* B. Bolin et al. (eds.) The global carbon cycle. John Wiley & Sons, Ltd., Chichester, U.K.

Bronger, A. 1991. Argillic horizons in modern loess soils in an ustic soil moisture regime: Comparative studies in forest-steppe and steppe areas from eastern Europe and the United States. Adv. Soil Sci. 15:41–90.

Buol, S.W., R.J. Southard, R.C. Graham, and P.A. McDaniel. 2003. Soil genesis and classification. 5th edn. Blackwell Publishing Co., Ames, IA.

Caldwell, M. 1975. Primary production of grazing lands, p. 41–73. *In* J.P. Cooper (ed.) Photosynthesis and productivity in different environments. Cambridge University Press, New York.

Cannon, M.E., and G.A. Nielsen. 1984. Estimating production of range vegetation from easily measured soil characteristics. Soil Sci. Soc. Am. J. 48:1393–1397.

Curtis, J.T. 1959. The vegetation of Wisconsin: An ordination of plant communities. University of Wisconsin Press, Madison, WI.

Darmody, R.G., and T.R. Peck. 1996. Soil organic carbon changes through time at the University of Illinois Morrow Plots, p. 161–170. *In* E.A. Paul, K. Paustian, E.T. Elliott, and C.V. Cole (eds.) Soil organic matter in temperate agroecosystems: Long-term experiments in North America. CRC Press, Boca Raton, FL.

Darmody, R.G., C.E. Thorn, J.C. Dixon, and P. Schlyter. 2000. Soils and landscapes of Kärkevagge, Swedish Lapland. Soil Sci. Soc. Am. J. 64:1455–1466.

Dahlman, R.C., and C.L. Kucera. 1965. Root productivity and turnover in native prairie. Ecology 46:84–89.

David, M.B., G.F. McIsaac, R.G. Darmody, and R. Omonode. 2009. Long-term changes in mollisol organic carbon and nitrogen. J. Environ. Qual. 38:200–211.

Dokuchaev, V.V. 1883. Russian Chernozems (Russkii Chernozem). Israel Prog. Sci. Trans., Jerusalem, 1967. Translated from Russian by N. Kaner. U.S. Department of Commerce, Springfield, VA.

Dormaar, J.F., and D.R. Sauerbeck. 1983. Seasonal effects on photoassimilated carbon-14 in the root system of blue grama and associated soil organic matter. Soil Biol. Biochem. 15:475–479.

Duchaufour, P. 1982. Pedology: Pedogenesis and classification. George Allen and Unwin, London, U.K.

Evans, L.T., and E.W. Russell. 1959. The adsorption of humic and fulvic acids by clays. J. Soil Sci. 10:119–132.

Fanning, D.S., and M.C.B. Fanning. 1989. Soil morphology, genesis, and classification. John Wiley & Sons, New York.

Fenton, T.E. 1983. Mollisols, p. 125–163. *In* L.P. Wilding, N.E. Smeck, and G.G. Hall (eds.) Pedogenesis and soil taxonomy. II. The soil orders. Elsevier, Amsterdam, the Netherlands.

Foss, J.E., M.G. Michlovic, J.L. Richardson, J.L. Arndt, and M.E. Timpson. 1985. Pedologic study of archaeological sites along the Red River. N.D. Acad. Sci. Proc. 39:51–69.

Franzmeier, D.P., L.D. Norton, and G.C. Steinhardt. 1989. Fragipan formation in loess of the Midwestern United States, p. 69–98. *In* N.E. Smeck and E.J. Ciolkosz (eds.) Fragipans: Their occurrence, classification, and genesis. SSSA Special Publication No. 24. SSSA, Madison, WI.

Gambrell, R.P., and W.H. Patrick, Jr. 1978. Chemical and microbiological properties of anaerobic soils and sediments, p. 375–423. *In* D.D. Hook and R.M.M. Crawford (eds.) Plant life in anaerobic environments. Ann Arbor Science Publishers Inc., Ann Arbor, MI.

Geis, J.W., and W.R. Boggess. 1968. The Prairie Peninsula: Its origin and significance in the vegetational history of Central Illinois, p. 89–95. *In* R.E. Bergstrom (ed.) The Quaternary of Illinois. University of Illinois, College of Agriculture, Special Publication No. 14. University of Illinois, Urbana, IL.

Glazovskaya, M.A. 1985. Soils of the world. Vol. I. A.A. Balkema Publishers, Rotterdam, the Netherlands.

Graham, R.C., and A.R. Southard. 1983. Genesis of a Vertisol and associated Mollisol in Northern Utah. Soil Sci. Soc. Am. J. 47:552–559.

Greenland, D.J. 1965. Interaction between clays and organic compounds in soils. Part 2. Soils Fertil. 28:521–532.

Hole, F.D., and G.A. Nielsen. 1968. Some processes of soil genesis under prairie, p. 28–34. *In* P. Schramm (ed.) Proc. Symp. of Prairie and Prairie Restoration. Special Publication No. 3. Know College Field Station, Galesburg, IL.

Hsieh, Y.P. 1992. Pool size and mean age of stable soil organic carbon in cropland. Soil Sci. Soc. Am. J. 56:460–464.

Jenny, H. 1930. A study on the influence of climate upon the nitrogen and organic matter content of soils. University of Missouri Agricultural Experiment Station Research Bulletin No. 52.

Jenny, H. 1941. Factors of soil formation: A system of quantitative pedology. McGraw-Hill, New York.

Jenny, H. 1980. The soil resource: Origin and behaviour. Springer-Verlag, New York.

Khan, F.A., and T.E. Fenton. 1994. Saturated zones and soil morphology in a Mollisol catena of central Iowa. Soil Sci. Soc. Am. J. 58:1457–1464.

Lauenroth, W.K., and W.C. Whitman. 1977. Dynamics of dry matter production in a mixed-grass prairie in western North Dakota. Oecologia 27:339–351.

Mausbach, M.J., and J.L. Richardson. 1994. Biogeochemical processes in hydric soil formation. Curr. Top. Wetland Biogeochem. 1:68–127.

McDaniel, P.A., and L.C. Munn. 1985. Effect of temperature on organic carbon–texture relationships in Mollisols and Aridisols. Soil Sci. Soc. Am. J. 49:1486–1489.

McKeague, J.A., M.V. Cheshire, F. Andreux, and J. Berthelin. 1986. Organo-mineral complexes in relation to pedogenesis, p. 549–592. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and microbes. SSSA, Madison, WI.

Munn, L.C. 1993. Effect of prairie dogs on physical and chemical properties of soils. *In* J.L. Oldemeyer, D.E. Biggins, B.J. Miller, and R. Crete (eds.) Management of prairie dog complexes for reintroduction of the black-footed ferret. USDA—Fish and Wildlife Service, Washington, DC.

Munn, L.C., G.A. Nielsen, and W.F. Mueggler. 1978. Relationships of soils to mountain and foothill range habitat types and production in western Montana. Soil Sci. Soc. Am. J. 42:135–139.

Nettleton, W.D., K.W. Flatch, and B.R. Brasher. 1969. Argillic horizons without clay skins. Soil Sci. Soc. Am. Proc. 33:121–125.

Nichols, J.D. 1984. Relation of organic carbon to soil properties and climate in the southern Great Plains. Soil Sci. Soc. Am. J. 48:1382–1384.

Novak, J.M., and N.E. Smeck. 1991. Comparison of humic substances extracted from contiguous Alfisols and Mollisols in southwestern Ohio. Soil Sci. Soc. Am. J. 55:96–102.

Oades, J.M. 1989. An introduction to organic matter in mineral soils, p. 89–159. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Ponnamperuma, F.N. 1972. The chemistry of submerged soils. Adv. Agron. 24:29–96.

Redmond, C.E., and J.E. McClelland. 1959. Occurrence and distribution of lime in calcium carbonate Solonchak and associated soils of eastern North Dakota. Soil Sci. Soc. Am. Proc. 23:61–65.

Richardson, J.L., L.P. Wilding, and R.B. Daniels. 1992. Recharge and discharge of groundwater in aquic conditions illustrated with flownet analysis. Geoderma 53:65–78.

Ruhe, R.V. 1970. Soils, paleosols, and environment, p. 37–52. *In* W. Dort, Jr. and J.K. Jones, Jr. (eds.) Pleistocene and recent environments of the central Great Plains. University Press of Kansas, Lawrence, KS.

Schafer, W.M., G.A. Nielsen, and W.D. Nettleton. 1980. Minesoil genesis and morphology in a spoil chronosequence in Montana. Soil Sci. Soc. Am. J. 44:802–807.

Schlesinger, W.H. 1977. Carbon balance in terrestrial detritus. Ann. Rev. Ecol. Syst. 8:51–81.

Schlesinger, W.H. 1991. Biogeochemistry: An analysis of global change. Academic Press, San Diego, CA.

Schultze, D.G., J.L. Nagel, G.E. van Scoyoc, T.L. Henderson, M.F. Baumgarder, and D.E. Stott. 1993. Significance of organic matter in determining soil color, p. 71–90. *In* J.M. Bigahm and E.J. Ciolkosz (eds.) Soil color. SSSA, Madison, WI.

Smith, G. 1986. The Guy Smith interviews: Rationale for concepts in soil taxonomy, p. 259. *In* T.R. Forbes (ed.) SMSS Technical Monograph No. 11. USDA-SCS, Washington, DC.

Sobecki, T.M., and L.P. Wilding. 1983. Formation of calcic and argillic horizons in selected soils of the Texas coast prairie. Soil Sci. Soc. Am. J. 47:707–715.

Soil Survey Staff. 1975. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. USDA agriculture handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1998. Keys to soil taxonomy. 8th edn. USDA-NRCS., U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd edn. USDA agriculture handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2006. Keys to soil taxonomy. 10th edn. USDA-NRCS., U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS., U.S. Government Printing Office, Washington, DC.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Official soil series Descriptions. Available online at http://soils.usda.gov/technical/classification/osd/index.html.Accessed 7/8/2011.

Stevenson, F.J. 1994. Humus chemistry: Genesis, composition, reactions. 2nd edn. John Wiley & Sons, New York.

Tate, R.L., III. 1980. Microbial oxidation of organic matter of histosols. Adv. Microb. Ecol. 4:169–201.

Tate, R.L., III. 1992. Soil organic matter: Biological and ecological effects. Krieger Publishing Co., Malabar, FL.

Thompson, J.A., and J.C. Bell. 1996. A color index for identifying hydric conditions for seasonally saturated conditions prairie soils in Minnesota. Soil Sci. Soc. Am. J. 60:1979–1988.

Thorp, J. 1948. How soils develop under grass, p. 55–66. *In* Yearbook of agriculture. U.S. Government Printing Office, Washington, DC.

Tiessen, H., J.W.B. Stewart, and J.R. Bettany. 1982. Cultivation effects on the amounts and concentration of carbon, nitrogen, and phosphorus in grassland soils. Agron. J. 74:831–835.

Troeh, F.R., and L.M. Thompson. 1993. Soils and soil fertility. Oxford University Press, New York.

Vepraskas, M.J., and S.W. Sprecher. 1997. Overview of aquic conditions and hydric soils, p. 1–22. *In* M.J. Vepraskas and S. Sprecher (eds.) Aquic conditions and hydric soils: The problem soils. SSSA Special Publication No. 50. SSSA, Madison, WI.

Walker, P.H. 1966. Postglacial environments in relation to landscape and soils on the Cary Drift, Iowa. IA Agric. Home Econ. Exp. Stn. Res. Bull. 549:838–875.

Warkentin, J. 1969. The western interior of Canada. McClelland and Stewart Ltd., Toronto, Canada.

Wright, H.E. 1970. Vegetational history of the central plains, p. 157–172. *In* W. Dort, Jr. and J.K. Jones, Jr. (eds.) Pleistocene and recent environments of the central Great Plains. University Press of Kansas, Lawrence, KS.

33.9 Spodosols

Randall Schaetzl

Willie Harris

33.9.1 Introduction

Spodosols are soils that typify the process of podzolization; they are roughly equivalent to Podzols in older U.S. classifications schemes (Baldwin et al., 1938; Muir, 1961; Petersen, 1976, 1984; McKeague et al., 1983). (The term Podzol is still widely used in many classification schemes outside of the United States.) In the process of podzolization, soluble compounds of Al and organic matter, often accompanied by Fe and Si, are translocated from an upper, eluvial zone to a lower, illuvial zone. As a result, a typical Spodosol profile has a bleached E horizon overlying a dark, reddish to brownish Bs, Bh, and/or Bhs horizon(s). The process is best expressed in humid climates, under vegetation assemblages that tend to produce acidic litter. Spodosols are most extensive in cool-cold climates, but also are common in subtropical and tropical regions, where eluvial and/or illuvial zones can be quite

thick and well expressed (Daniels et al., 1975; Thompson, 1992; Figure 33.58). These thick illuvial zones constitute a substantial pool of C (Holzhey et al., 1975; Stone et al., 1993), one that has not yet been definitively tallied in many estimates of global C stocks.

The original Podzol concept derived from the ashy, light colored E horizon, having been derived from the Russian words "pod" (under) and "zola" (ash) (McKeague et al., 1983; Sauer et al., 2007). This name has often been mistakenly applied to all soils with a light colored, high-chroma E horizon, regardless of the nature of the underlying horizons. Thus, soils with a well-developed E horizon and undergoing various degrees of podzolization needed to be separated from those that embodied the core concept of the true podzolization process. Therefore, in an attempt to restrict the order Spodosols to soils formed mainly via podzolization, the Soil Survey Staff (1975, 1999) placed the taxonomic focus on B horizon character, chemistry, and morphology, for example, Stanley and Ciolkosz (1981), Mokma (1993). Important aspects of the current Spodosol classification scheme for the United States (Soil Survey Staff, 1999) are discussed in the following.

33.9.2 Processes Involved in the Formation of Spodosols—the Podzolization Suite

In podzolization, a combination of organic matter and Al, often in association with Fe^{3+}, are translocated from the upper profile to a lower, illuvial horizon (Petersen, 1976; DeConinck, 1980; Buurman and van Reeuwijk, 1984; Courchesne and Hendershot, 1997, Lundström et al., 2000a). Some literature suggests that podzolization includes not only translocation of metals (sesquioxides) and organic matter, but also the associated processes of (1) chemical weathering by organic and carbonic acids and (2) clay destruction and/or translocation (Fridland, 1958; Schaetzl and Anderson, 2005). Regardless, in both definitions, podzolization processes operate under acidic conditions, where weathering is intense in the upper profile, and where the weathering byproducts are translocated to the B horizon, or removed from the profile. As a result, an acidic E horizon is formed—impoverished in
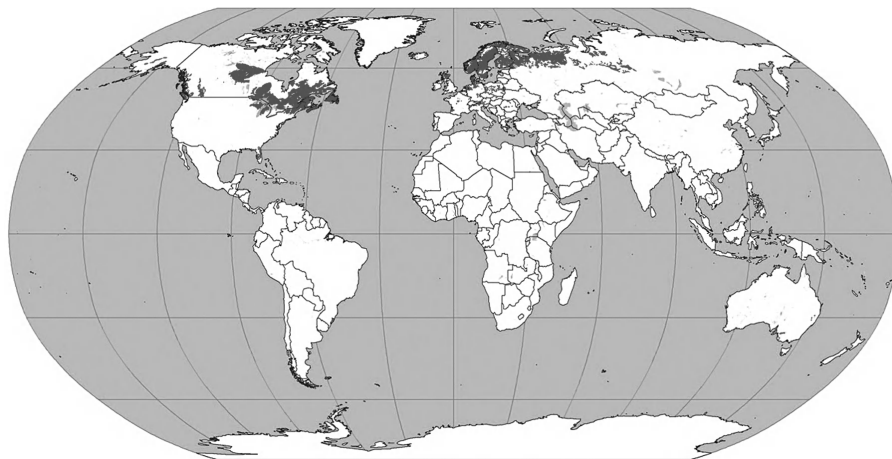
FIGURE 33.58 Global distribution of Spodosols. (Courtesy of USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC. 2010.)

FIGURE 33.59 Relict treethrow pits and mounds in a landscape dominated by Spodosols. This landscape was originally forested but is now in pasture. It is likely that the microtopography has been muted somewhat by the decades of grazing at this site. Location: Northeastern Wisconsin. (Photo by R. Schaetzl.)

Al, organic matter and usually Fe. In many soils, the result of podzolization is a profile that classifies as a Spodosol (or Podzol). It is important to note, however, that podzolization is also a subsidiary or background process in many soils. For example, it may occur after, or in association with, acidification, lessivage, brunification, and various intensities of redox processes (DeConinck and Herbillon, 1969; Guillet et al., 1975; Ugolini et al., 1977; Li et al., 1998; Harris and Hollien, 1999, 2000; Hseu et al., 2004).

33.9.3 Characteristics of a Spodosol Profile

Spodosols often have distinct horizons with relatively abrupt or clear boundaries, usually due to low rates of pedoturbation by bioturbators, for example, worms, mammals, insects, which can blur boundaries (Langmaid, 1964), or break and distort them, as occurs locally during tree uprooting (Figure 33.59; Schaetzl, 1986; Schaetzl et al., 1990). Coarse soil textures, acidic conditions, and cool soil temperatures act in unison to retard faunalturbation. It is common for E–Bh horizon boundaries to be wavy or irregular, sometimes conforming to the shapes of previous tree roots, which might have influenced the percolation pathways of water in these sandy soils (Figure 33.60).

Due to the low rates of decomposition more than to high rates of litter production, Spodosols typically have thin to thick (2–20 cm), usually continuous, O horizons. Particularly thick O horizons are found under coniferous forests and heath vegetation. Decomposition rates of the litter decrease with temperature, favoring greater litter accumulation for cool region (frigid and cryic) Spodosols. Slow litter decomposition is also a consequence of the acidic litter character and its high contents of slowly degradable compounds, for example, lignins and waxes (Vance et al., 1986). Spodosol O horizons vary in their degree of decomposition, although Oi and Oe horizons, also referred to as *mor* litter, are common. The slow decomposition in the O horizon is performed by fungi, usually resulting in mycelial mats.

The A horizons in Spodosols are typically thin, especially in coarser-textured soils, where macroorganism activities are low. Bioturbation of organics from the O horizon into the mineral

(a)   (b)   (c)
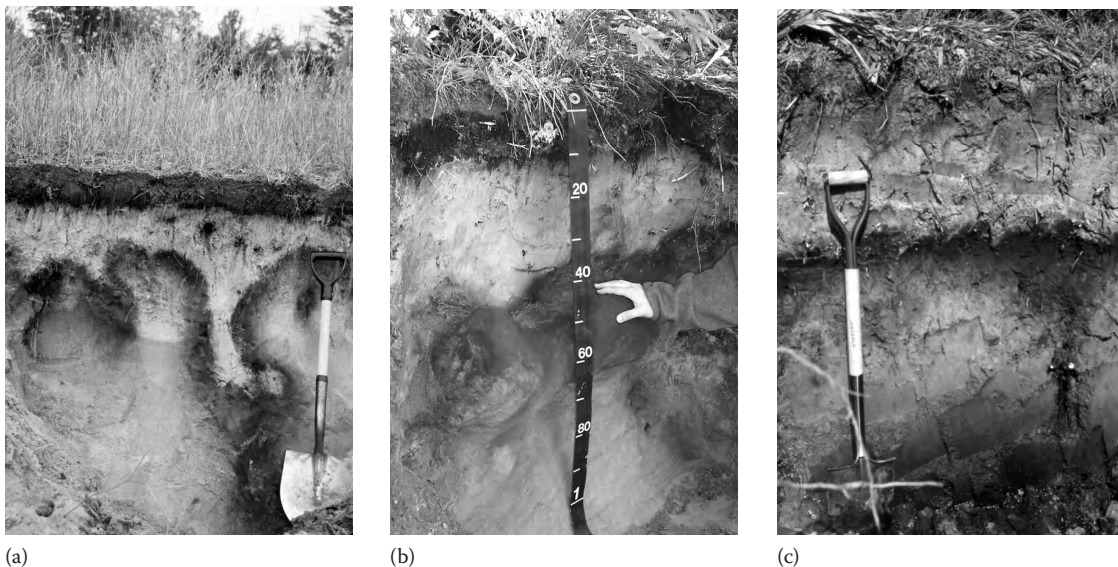
FIGURE 33.60 Morphology of the E and Bs horizons some representative, well-developed, Spodosols. (a) A sandy, mixed, frigid, Entic Haplorthod, showing pronounced tonguing of the E and Bs horizons. Note that this soil has been plowed in the past and exhibits an Ap horizon. Subtropical Alaquods commonly exhibit similar, irregular boundaries. Location: NE Lower Michigan. (Photo by R. Schaetzl.) (b) A sandy, mixed, frigid, shallow, ortstein, Typic Durorthod, showing a thick, cemented Bsm horizon (ortstein). Location: Northern Michigan. (Photo by R. Schaetzl.) (c) A sandy, siliceous, hyperthermic, Aeric Alaquod, showing an upper and a lower Bh horizon, with contrasting upper boundaries (abrupt vs. gradual, respectively). Note that the E horizon above the lower Bh has not been as thoroughly eluviated as has the upper E horizon. Location: Northern Florida. (Photo by L. Daniels.)

soil, typically important to A horizon formation, is minimal in Spodosols. Thus, the primary pathways of A horizon formation are via translocation of organic materials into the mineral soil by percolating water and by in situ root decay. Neither of these processes is strong in Spodosols. However, small organic molecules, with varying degrees of solubility, can move through the A horizon and participate in chemical reactions in the middle and lower profile. Many Spodosols lack an A horizon entirely; the O horizon is directly overlying an E horizon.

The E horizon is a morphologic hallmark of the Spodosol order and of podzolization. The E horizon is lighter colored (higher Munsell values) than other horizons in the profile because (1) coatings bound by oxides and organic matter have been stripped from, or inhibited from coating, the quartz grains that dominate the mineralogy and (2) most of the dark, ferromagnesian minerals in the parent material have been chemically weathered. Many Spodosol E horizons are bleached to the point of being almost pure white. The E horizon is commonly the most acidic horizon in the Spodosol profile, having been leached of most of its base cations (Lundström et al., 2000b). It is also coarser textured than underlying horizons, partially due to clay eluviation, as well as intense chemical weathering that destroys some of the clays. Well-developed E horizons qualify as albic horizons (Soil Survey Staff, 1999). Although most Spodosols do have a well-developed, high-chroma E horizons, *Soil Taxonomy* (Soil Survey Staff, 1999) has placed taxonomic emphasis on the illuvial (B) horizon, for two reasons: (1) the E horizon is often disturbed by plowing, bioturbation, or erosion, and (2) in soils formed in loamy parent materials, the E horizon is often not well expressed, although the strong B horizon is clearly indicative of ongoing podzolization (Figure 33.61).

The reddish, reddish brown, or nearly black, B horizon in Spodosols is usually a Bs, Bh, or Bhs, indicative of illuvial sesquioxides, humus, or both (Mokma and Evans, 2000). In Spodosols with a high water table (Aquods), the B horizon may be low in ferrous iron (Fe^{++}) compounds, having been lost to the groundwater. The zone of maximum illuvial humus accumulation in most Spodosol B horizons is near the top, variously decreasing with depth. Thus, Bhs–Bs1–Bs2 horizon sequences are common, as illuvial carbon decreases with depth in the horizon. The upper boundary of B horizons is typically abrupt, often with wavy boundaries or tonguing, contrasting to the typically diffuse lower horizon boundaries (Figure 33.60). The wavy upper boundary reflects the preferential flow of percolating water, as it moves in tongues along preferred zones in coarse-textured soils, or as it follows root pathways (Figure 33.60). Schaetzl (1986, 1990) determined that at least some of these tongues are due to preferential infiltration beneath microtopographic depressions and pits (present or former) in the forest floor. Well-developed B horizons qualify as spodic horizons if they meet certain chemical and morphological criteria (Soil Survey Staff, 1999). Spodic horizons are characterized by illuvial Al and organic matter; more details are provided in Section 33.9.7.

Strongly-developed spodic horizons, for example, Bhsm, Bsm, can be cemented into ortstein (Figure 33.60; Wang et al., 1978; Dubois et al., 1990; Mokma, 1997). Ortstein is commonly so well cemented that it restricts root penetration, but when sandy it still

FIGURE 33.61 Profile of a loamy, mixed, frigid, Typic Haplorthod from Maine. Note the weak E horizon in this soil, despite evidence from the B horizon of intense podzolization and illuviation of humus and Fe. This type of Spodosol morphology is common in parent materials where silt and clay contents are high; in such soils the high surface area limits the rate at which podzolization can "strip" all the particles in the E horizon of their coatings. (Photo by R. Schaetzl.)

maintains good permeability (Lambert and Hole, 1971). Ortstein is generally more planar in Aquods than in better drained soils and is usually associated with a water table (Mokma et al., 1994). The cementing agent in ortstein is generally assumed to be an amorphous, Al-dominated material (Lee et al., 1988), although Fe and Si may also play a role.

33.9.4 Spodosol Genesis: Details and Theories of the Podzolization Process

33.9.4.1 The General Suite of Processes

There are two primary theories of podzolization: (1) the protoimogolite theory and (2) the chelate-complex theory. Both endeavor to explain the mobilization, translocation and immobilization of the oxidized, metal ions (sesquioxides) and organic compounds in soils (Schaetzl and Anderson, 2005), as well as the variability in organic matter content in Spodosol B horizons. By association, podzolization theory must also explain the intense weathering typical of these soils. All podzolization theories assume that the soil parent material has been preconditioned or exists in a state conducive to the process: it has minimal clay and is acidic or has been acidified. It is also assumed that the vegetation produces litter rich in low-molecular-weight organic acids and fulvic acids, and that

the climate is humid enough to initiate deep percolation, at least at some times of the year.

Intense weathering of minerals by organic acids is a central part of the podzolization process (Van Hees et al., 2000). As minerals weather, various cations are released to the soil solution. Base cations must be depleted before Al and Fe can be rendered mobile. If the soil pH is not low enough, that is, if too many base cations exist in the soil, Al and Fe cations released by weathering will precipitate as solid, less mobile forms (oxides, oxyhydroxides, and hydroxides); they will not readily translocate. Defining the exact mechanism by which these metal cations become mobile and illuviate is where the two main theories of podzolization diverge.

The protoimogolite theory focuses on the inorganic, colloidal transport of Al, Si, and Fe. The theory was formulated by the observation that Al and Fe can exist in some humus-poor Spodosols as amorphous, inorganic compounds, for example, imogolite and allophane (Farmer et al., 1980; Anderson et al., 1982; Farmer, 1982; Childs et al., 1983; Gustafsson et al., 1995). Imogolite and allophone are Al–Si hydroxyl gels also known as imogolite-type materials, or ITM (Lundström et al., 2000a). In this theory, ITM form from weathering products in O and E horizons and move with percolating water until immobilized, usually in the B horizon where higher (>5) pH values exist. Al is transported as a positively charged hydroxy aluminum silicate complex (Anderson et al., 1982). The Al and Fe in ITM have presumably been dissolved/weathered by noncomplexing organic and inorganic acids, as well as by readily biodegradable, small, complexing organic acids (Lundström et al., 2000b). Next, negatively charged, colloidal organic matter migrates from the upper profile into the B horizon and precipitates onto the positively charged ITM that are already there. This process is supported by thin section data that show dark organs surrounding allans (allophane-rich cutans) (Freeland and Evans, 1993), or Al- and Fe-rich cutans overprinted onto Si-rich cutans in B horizons (Jakobsen, 1989). Dissolution/weathering processes continue to act on ITM in the B horizon. Because ITM are more easily weathered than crystalline Fe oxides, ITM will continue to eluviate, eventually leaving behind an Fe-rich B horizon, as the lower B horizon becomes enriched in Al by the continued dissolution and remobilization of ITM rich in Al. In a hybrid model of sorts, Ugolini and Dahlgren (1987) and Dahlgren and Ugolini (1991) proposed that ITM are formed in B horizons in situ, as organometallic complexes illuviate and interact with an Al-rich residue or protoimogolite formed thereby CO_2 weathering (Lundström et al., 2000b). Low soil organic matter contents tend to favor the formation of ITM, whereas conditions in soils with abundant organic matter appear to inhibit its formation (Wang et al., 1986).

In the more traditional, widely accepted, chelate-complex (or fulvate complex) theory, organic acids form chelate complexes with Fe and Al cations, rendering these normally insoluble cations (and Si) soluble for translocation (Wright and Schnitzer, 1963; DeConinck, 1980; Buurman and van Reeuwijk, 1984; Dahlgren and Ugolini, 1989; Sauer et al., 2007). Normally, Fe^{3+}

and Al^{3+} are not soluble in soils, but when chelated they become more soluble and are readily translocated in percolating water (Riise et al., 2000). In short, the process is driven by organic acids, as opposed to the dominantly inorganic pathway outlined by the protoimogolite theory.

Organic acids and phenolic compounds in soils generally can be categorized based on their molecular weights: low-molecular-weight (LMW) acids <1000 Da, fulvic acids (FAs) 1000–3000 Da, and humic acids >3000 Da (Krzyszowska et al., 1996; Lundström et al., 2000b). Many of these acids are readily produced during litter decay and carried in the soil solution. They also can be produced as root and fungal exudates. From experiments on aqueous extracts from plant litter, it has been long known that organic acids can dissolve ferric and aluminum oxides (Bloomfield, 1953; Schnitzer and Kodama, 1976; Kodama et al., 1983; Lundström et al., 1995). Many of these acids, however, also readily form metal chelates. Thus, central to the chelate-complex theory is the idea that organic acids produced in soils not only weather primary minerals but also chelate the cations released from them. As the chelate complexes move within the profile, they continue to chelate more metal cations, but remain soluble until a certain level of saturation (metal loading) is achieved (Petersen, 1976). At this point, the complex is rendered immobile and precipitates on ped faces, roots, and mineral surfaces. Immobilization (precipitation) of the chelate complexes occurs for any of a number of reasons: (1) increased pH values in the lower profile (Gustafsson et al., 1995), (2) microbial decomposition of the organometallic complex (Lundström et al., 1995), (3) stoppage of the percolating water at a water table, lithologic discontinuity or aquitard (DeConinck, 1980; Buurman and van Reeuwijk, 1984; Schaetzl, 1992; Schaetzl and Schwenner, 2006), (4) flocculation from continued metal loading and electrostatic charge reduction of complexes (DeConinck, 1980), and (5) sorption/precipitation of complexes within a matrix of finer illuvial materials also accumulated as part of the podzolization process (Holzhey et al., 1975; Dahlgren and Marrett, 1991; Harris et al., 1995; Li et al., 1998; Harris and Hollien, 1999, 2000). Once precipitated in the lower B horizon, aluminum can be released by additional microbial breakdown of the organometallic complexes, or it can combine with silica to form ITM. Iron so released will form less soluble ferric oxyhydroxides (Buurman and van Reeuwijk, 1984). With time, as more organometallic coatings accumulate in the B horizon, cutans on grain surfaces tend to thicken. Because they are not crystalline, the cutans tend to shrink and crack upon drying. Cracked grain coatings are diagnostic of illuvial organometallic compounds (DeConinck, 1980; Stanley and Ciolkosz, 1981).

In the initial stages of podzolization, Fe and Al cations are released from primary minerals (if present) so rapidly that any chelate complexes that form are quickly saturated and hence, immobilized. Eventually, however, a zone of depleted Fe and Al forms (the incipient E horizon). Newly formed organic chelates can then pass through this eluvial zone before becoming saturated. Therefore, the E horizon may be present in the early

stages of podzolization, but not visible until it deepens past the depth of organic matter accumulation, that is, the A horizon. In sandy, upland soils, mixing is so minimal that A horizons are thin, and distinct E horizons can exist immediately below, perhaps within a few millimeter of the surface. Over time, the E horizon thickens, as (1) metal cations are continually stripped from the top of the B, remobilized, and redeposited lower and (2) the base of the E grows downward, below the depth of melanization. The B horizon continues to accumulate organic matter, Fe, and Al as cutans and as microaggregates, but the majority of the illuvial compounds reside near the top of the B horizon.

33.9.4.2 Groundwater Podzolization: Alaquods

Most Spodosols formed by processes described above are found in cool (mesic to cryic), midlatitude sites, on freely draining, upland landscape positions. However, large areas of well-expressed Spodosols also occur in subtropical and tropical regions where podzolization is associated with water tables (Brasfield et al., 1973; Dubroeucq and Volkoff, 1998; Harris, 2000; Watts and Collins, 2008). These Spodosols commonly classify as Alaquods, reflecting their wetter conditions and and prevalence of Al over Fe in the spodic horizon (usually a Bh horizon). In the southeastern United States, these wet Spodosols typically occur on nearly level, Coastal Plain landforms called "flatwoods" (Brasfield et al., 1973; Brooks, 1981). Flatwoods are poorly drained uplands where periodic chemical reduction and leaching favor Fe depletion. Flatwoods vegetation includes pine species (e.g., *Pinus palustrus*; *Pinus taeda*) with an understory dominated by saw palmetto (*Serenoa repens*) and gallberry (*Ilex glabra*) (Abrahamson and Hartnett, 1990), a plant community consistent with podzolization. However, Alaquods are not restricted to this plant community or to flatwoods landscapes. They can be found along some lake margins, seepage zones, floodplains, etc. with the common denominator being a seasonal water table.

A clue to the water table linkage is the morphological transition between Alaquods and better drained soils upslope (Garman et al., 1981; Tan et al., 1999). The common trend is for the thickness and morphological prominence of E and Bh horizons to diminish with diminishing water table influence, for example, shorter periods of saturation, greater depths to saturation, etc. (Figure 33.62). The Bh lightens in color and its upper boundary becomes shallower as the mean water table depth increases. This divergence between Bh horizon and the water table at transitions confirms that Bh horizons in these Alaquods do not form as a simple consequence of C precipitation with groundwater-entrained Al in quartz sand. Instead, true podzolic E horizons form, similar to those in Spodosols of cooler regions. Similarity of cooler and warmer Spodosols is also evidenced by morphological properties such as common irregular upper boundaries, for example, the root disturbance legacy effect on preferential flow as discussed above (Figures 33.59 and 33.60). Furthermore, near-surface water saturation
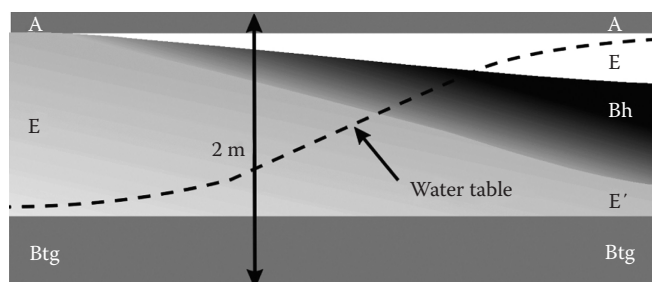
FIGURE 33.62 Schematic of a transition between Paleudult and Alaquod. Fading-upward of Bh and wedging out of E is typical of transitions between Aquods and better drained soils in northern Florida (Garman et al., 1981; Tan et al., 1999). The E of the Paleudult has sufficient coatings on sand grains to impart lower values and higher chromas than for the Alaquod E, in which grains have been essentially stripped of coatings.

is a dynamic seasonal condition rather than a static condition that would be required to explain abrupt white-to-black horizon boundaries.

The E–Bh boundary in Alaquods is also a boundary between zones of uncoated sand grains (A and E horizons) and coated grains (Bh and subjacent horizons; Figure 33.62). Sand-grain coatings are either stripped or prevented from forming in the upper profile under typical Alaquod hydrology, yet coatings occur in near-surface horizons of better drained sandy soils of the region (Harris et al., 1996). Hence, the origin, composition, and stability of grain coatings are pertinent to the mechanism of Alaquod formation. These coatings likely originate from mineral weathering in conjunction with the binding effects of secondary metal oxides. Most <50 μm-sized minerals in these soils are bound in coatings. The clay fraction of grain coatings in the Coastal Plain soils of the southeastern United States is commonly dominated by hydroxy-interlayered vermiculite, kaolinite, gibbsite, and quartz (Carlisle and Zelazny, 1974; Harris et al., 1987a, 1987b), with minor amounts of metal oxides that bind the coatings. These minerals, except for quartz, are nearly depleted in podzolic E horizons.

Experimental generation of Aquod-like E–Bh horizons has documented that colloidal transport, that is, eluviation–illuviation of sand-grain coatings, was a component of the process (Harris et al., 1995; Harris and Hollien, 2000). The coating destabilization occurred as a result of organo-complexation of the metal oxide-binding agents. Sand-grain stripping produced a light-gray podzolic E horizon and culminated in illuviation of the mobilized sand-grain coating materials to form a finer-textured zone subjacent to the E horizon. The illuvial zone (incipient Bh) provided conditions for sorption of organometal complexes and thus served as a template for Bh horizon development. The E–Bh horizon boundary contrast sharpened with time, from the darkening effect of continued C flux.

Colloidal transport in experimental formation of Aquod-like features does not prove that it occurs in the genesis of real Alaquods. However, it is consistent with the (1) crystalline clay mineralogy of Bh horizons, which typifies that of sand-grain

coatings in nonpodzolized areas (Harris and Carlisle, 1985; Harris et al., 1987a; Harris and Hollien, 2000), (2) upward-trending Bh horizons in areas transitional to better drainage, as described above (Figure 33.62), (3) elevated silt and clay contents in Bh horizons, and (4) abrupt, irregular E–Bh boundaries that can be more easily accounted for by the physical accretion of fine particulates than by chemical precipitation alone. Indeed, other studies have reported evidence of particle migration in Spodosols (Calhoun and Carlisle, 1973; Guillet et al., 1975; Ugolini et al., 1977; Li et al., 1998; Harris and Hollien, 1999).

The role of the water table in Alaquod formation may be to bring about conditions unfavorable for the stability or formation of sand-grain coatings. Artificial generation of E- and Bh-like horizons was achieved only (1) in predisposed coated sands collected on wetter landscape positions, that is, bordering Alaquods, and (2) under fluctuating water table conditions (Harris et al., 1995). Lower Fe oxide contents and lower Al crystallinity are factors that could explain the predisposition of wetter landscape materials to form Aquod-like features, compared with materials from drier summit positions. Lower Fe contents would be a predisposing factor for sand-grain coating destabilization because Fe oxides serve as binding agents. Aluminum oxides were not depleted in predisposed materials and still served to bind coatings. However, the lower crystallinity of Al oxides, possibly the result of higher organic acid activity in the wetter positions, would render them more vulnerable to organic acid complexation (Huang and Violante, 1986). Thus, wetter landscapes predispose sand-grain coatings to undergo destabilization and to participate in podzolization processes theorized above, should the groundwater influence intensify.

A fluctuating water table is still required to experimentally form artificial E–Bh sequences, even in predisposed material. Hence, the water table appears to have a role in not only predisposing, but also actually triggering, podzolization. The triggering mechanism, however, is still uncertain. It may relate to the effects of the frequency and duration of water saturation in the affected zone. For example, anaerobic conditions could inhibit microbial degradation of complexing organic anions, such that these anions reach sufficient activity to mobilize metals.

33.9.4.3 Groundwater Podzolization: Deep, Thick Bh Horizons

The previous discussion on Alaquod formation conforms closely to the chelate-complex theory of podzolization. However, the formation of deep, thick Bh horizons, found in some subtropical coastal regions (Daniels et al., 1975; Thompson, 1992), may involve elements of the protoimogolite theory. These horizons are morphologically distinct from shallower, thinner Bh horizons (Figure 33.60) and their upper boundaries may be located within, or below, 2 m. There are sometimes multiple Bh horizons with varying degrees of coalescence, across zones up to several meters thick. Deep Bh horizons of some regions (1) darken gradually from the upper boundary downward, rather than being

darkest at the top, (2) are not overlain by a well-expressed podzolic E horizons, (3) have consistently smooth upper boundaries, and (4) transgress soil drainage continua, extending below excessively drained sandy soils, as well as Alaquods. These distinctions suggest C immobilization by *in situ* inorganic Al (Farmer et al., 1983) at the approximate upper boundary of long-term saturation, adhering to the independent C-transport aspect of the latter theory. The absence of a well-expressed podzolic E horizon is evidence that immobilization of C by Al, as influenced by water table, could occur whether there has been C-induced transport of Al or not.

33.9.5 Factors That Influence the Formation of Spodosols

Podzolization occurs in humid climates, where precipitation exceeds evapotranspiration during some of the year, such that water frequently percolates through the profile, or stands in the profile at the water table (McKeague et al., 1983). Podzolization is best expressed under vegetation that produces acidic litter, such as coniferous forest and heather, and on coarse-textured parent materials that have minimal surface area. It is especially strong in the coniferous forests of the high midlatitudes. Cool temperatures in midlatitude and alpine areas keep evapotranspiration rates low and inhibit decomposition of the acidic litter. The discussion below focuses on how, as these factors change, the character of the profile also changes.

33.9.5.1 Vegetation

Podzolization is promoted under vegetation that produces acidic litter that decays slowly. The litter decay products, that is, organic acids of various kinds, not only help weather Fe- and Al-bearing primary minerals, but also can be efficient as metal chelates, because they release abundant amounts of water-soluble organic acids. Vegetation types that particularly promote podzolization include heath (heather), hemlock (*Tsuga canadensis*), kauri (*Agathis australis*), and pine (*Pinus* sp.), the litter from which releases acidic leachates (Bloomfield, 1953; Mackney, 1961; Nielsen et al., 1987a, 1987b; Madsen and Nørnberg, 1995; Mossin et al., 2001; Jongkind et al., 2007). Thick litter layers in various states of decay are typical in naturally vegetated Spodosol landscapes. Thick O horizons also help reduce variation in soil moisture contents and allow for more prolonged weathering and leaching of the mineral soil below. Barrett (1997) illustrated the importance of above-ground vegetation to podzolization by demonstrating that the process markedly slows after only a few decades of forest removal and litter disturbance (see also Hole, 1975).

Fire impacts Spodosol development by eliminating or reducing the O horizon volume, changing its chemistry and character, and altering the above-ground forest vegetation. Mokma and Vance (1989) observed that soils under vegetation that burns more frequently were much less developed than in areas of diminished fire frequency. Indeed, in parts of Michigan,

repeated (ca. 5–50 years) fires have so retarded spodic development that most of the soils are Entisols, even after 15,000 years (Schaetzl et al., 2006). Schaetzl (2002) also noted that the areas in Michigan having maximal fire frequencies and minimal soil development also receive much less snow than other areas with better soil development, suggesting that snowpack and snowmelt impacts on soils development are often as important as fire frequency. Indeed, the two are often mutually reinforcing. Subtropical Alaquods are an exception to the fire-suppressing effect on podzolization; they predominate in fire-maintained flatwoods ecosystems.

Because most Spodosols form under forest vegetation, mixing by tree uprooting is a common, albeit localized and often pedon-specific, form of disturbance (Dunn et al., 1983; Schaetzl et al., 1990; Jonsson and Dynesius, 1993). Pit and mound topography is the common surface expression of this process, remaining on the landscape for centuries after the treethrow event (Figure 33.59; Schaetzl and Follmer, 1990; Kabrick et al., 1997; Small, 1997). This microtopography dramatically impacts infiltration pathways, directly and indirectly (Schaetzl, 1990). Soils are often much better developed beneath pits, where infiltration is focused, and litter and snowpacks are thicker. Tree uprooting distorts, breaks, mixes, and even inverts soil horizons, leading to high levels of soil spatial variability (Veneman et al., 1984; Schaetzl, 1986; Meyers and McSweeney, 1995). The spatial variation in soil development continues to evolve and even strengthen as a consequence of the subsequent microtopography and its effects on the redistribution of water and litter.

33.9.5.2 Climate

Spodosols on upland sites are best developed in cool, humid, continental climates, especially those with a pronounced, snowy winter. Cool temperatures favor vegetation types, for example, conifers and heather, that produce acidic litter, which promotes podzolization. Low (but not freezing) mean annual temperatures, coupled with an extended period of subfreezing temperatures and snow in winter, tend to slow litter decay processes and maximize the potential for organic acid production and its slow, steady translocation into and through the mineral soil. Wet local conditions in the tropics similarly affect litter decay, promoting Spodosol development. In warmer and/or drier climates, much of the litter is quickly oxidized and mineralized, resulting in minimal amounts of acid leachate for podzolization. Globally, podzolization and spodic development (on upland sites) is maximal in the frigid soil temperature regime, eventually weakening at the extreme cold end of the climate spectrum, in the gelic and pergelic soil temperature regimes. Here, vegetation transitions to tundra, and soils are frozen for longer periods of time. Similarly, podzolization on upland sites is weak in the slightly warmer, mesic soil temperature regime, where litter decay is rapid and where forest vegetation tends to include more broadleaf trees and base cyclers (Cann and Whiteside, 1955; Schaetzl and Isard, 1991). For example, in midlatitude locations near the "warm" end of their range, Spodosols located on slightly cooler sites tend to be better developed, especially with regard

to B horizon character (Stanley and Ciolkosz, 1981). In the frigid soil temperature regime of northern Michigan, Hunckler and Schaetzl (1997) observed much better spodic development on cooler, northerly slope aspects. Thus, climate impacts podzolization mainly by affecting the vegetation that occupies the landscape and by impacting the nature of the pedohydrologic cycle.

In midlatitude areas, Schaetzl and Isard (1991) observed that podzolization is much more pronounced in areas that receive heavy, early snowfall, and accumulate thick snowpacks. This pattern suggests that snowpacks insulate the soils from freezing (Isard and Schaetzl, 1995), facilitating slow and steady infiltration during spring snowmelt. Snowmelt water is often rich in organic acids, having just percolated through fresh and slightly decomposed litter (Schaetzl and Isard, 1996). Additionally, areas with deep snowpacks have warmer soils, especially in winter, which foster weathering of primary minerals and release of metal cations to the soil solution (Isard et al., 2007). The melting snow also provides water for long, steady infiltration events, and thus, efficient translocation of metal chelates. In support of this assumption, Schaetzl (1990) observed that soil water during snowmelt contained more Fe and Al than water similarly sampled during warm season percolation events. Thicker snowpacks also reduce fire frequencies; fire has been shown to inhibit podzolization in some areas, as previously discussed.

33.9.5.3 Parent Material

Spodosols form best on acidic, coarse-textured parent materials, with low amounts of silicate clay and Fe-rich minerals. In these materials, base and metal cations are slowly released by weathering, and the metal cations are subject to chelation by organic acids in the profile and translocation to the B horizon. (The fate of base cations is varied; many are leached from the profile, some are biocycled.) In finer-textured, Fe-rich soils, cations (especially Fe) are released in abundance, although many are quickly adsorbed to clay minerals. As such, lessivage (clay translocation) and/or brunification (reddening), rather than podzolization, tends to dominate the pedogenic suite. Essentially, the great numbers of Fe and other cations released to the soil solution in these types of parent materials commonly overwhelm the ability of the soil to translocate them. As a result, distinct eluvial/illuvial zones do not readily form (Duchaufour and Souchier, 1978; Alexander et al., 1994; Figure 33.61). These soils commonly only exhibit a reddened, "color B" horizon with no, or only a thin, E horizon. Thus, as clay and Fe contents increase, translocation of metal cations and humus decreases and E horizons are less "bleached" (Cline, 1949; Gardner and Whiteside, 1952). Schaetzl and Mokma (1988) used this knowledge to develop a numerical, field- and morphology-based index of spodic development—the POD index.

Increased biological activity in finer-textured soils also makes it more difficult to retain a thick O horizon at the surface. Particularly, important is the increased amount of bioturbation in these finer-textured soils, which can blur or destroy incipient horizons. Basic (high pH) parent materials must generally be leached of free carbonates before podzolization can
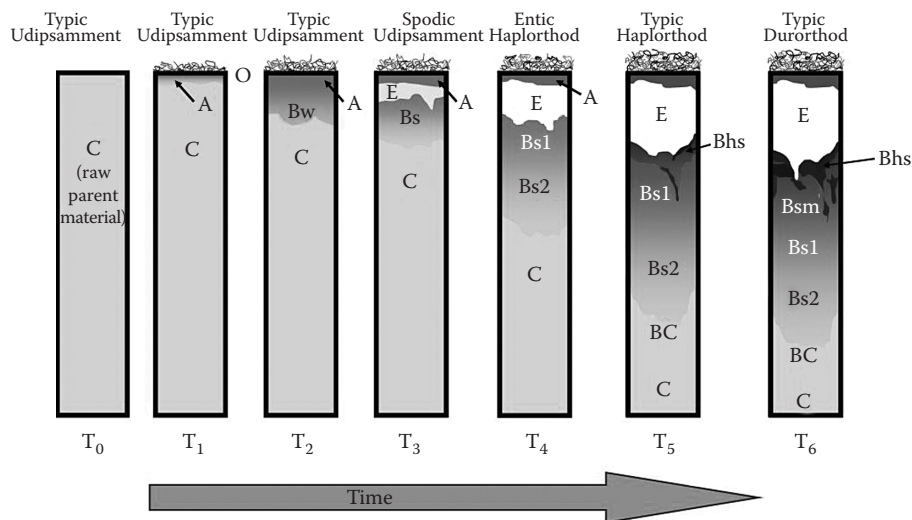
FIGURE 33.63 Typical horizonation/soil development sequence, from Entisols to well-developed Spodosols, on sandy, well-drained parent materials in the midlatitudes.

begin in earnest, as high pH values inhibit the formation of metal–chelate complexes and inhibit weathering and release of metal cations. Thus, podzolization will occur sooner in parent materials that are initially acidic. Coarse-textured parent materials also foster rapid infiltration and deep percolation, accelerating the development of an eluvial zone, and a deeper B horizon.

33.9.5.4 Time

Spodic morphologies develop along generally predictable pathways (Figure 33.63) and can form in as few as a few decades (Paton et al., 1976; Stützer, 1998), although it usually takes at least 3000–8000 years, under optimal conditions, for this development to reach the point where the soil classifies as a Spodosol (Franzmeier and Whiteside, 1963; Jauhiainen, 1973; Barrett and Schaetzl, 1992; Mokma et al., 2004; Sauer et al., 2008). E horizons typically form first, or are most apparent, often forming within a few centuries (VandenBygaart and Protz, 1995; Lichter, 1998; Barrett, 2001). Noticeable darkening and reddening of the B horizon take longer to occur, but is often visible within 1000–2000 years. After the Bs horizon has become well established, a Bhs horizon may form near the top of the illuvial zone. A chronosequence in beach deposits in Norway (Sauer et al., 2008) showed incipient podzolic features within 3800 years, but fully expressed spodoc horizons were observed only in the 6600-year-old soil. In some soils, cementation of parts of the B horizon (ortstein) may follow. Sauer et al. (2007) provide an excellent review of the rate of Spodosol formation.

33.9.5.5 Topography, Relief, and Drainage Class

In this chapter, we have differentiated between cool region Spodosols that commonly occur on freely drained upland landscape positions and subtropical Spodosols that are mainly restricted to wet upland landscapes. The commonality is "upland;" Spodosols, regardless of drainage, are located on

landscapes that experience leaching and throughflow of soil water. Flatwoods of the southeastern United States are an example of poorly drained, leached uplands. However, some poorly drained leaching soil environments are not uplands, for example, toeslopes of sandy hills, and lake margins. There is a strong hydrologic–topographic control of Spodosol occurrence in the southeastern United States. Spodosols dominate on landscapes with fluctuating water tables but are absent, or only of isolated occurrence, on other landscapes, for example, sandhills, karst terrain. Podzolization in all settings involves gravity-driven translocation, which can also be affected by preferential flow. Materials move from eluvial to illuvial horizons and, barring disturbances such as tree uprooting or bioturbation, and excluding cation biocycling, do not readily move back toward the surface, even in wetter Spodosols. Indeed, Schaetzl and Mokma (1988) found that Spodosols in somewhat poorly drained positions were generally the best developed Spodosols, among drainage classes. Further, Alaquods can have strongly expressed E horizons (Watts and Collins, 2008).

33.9.6 Distribution of Spodosols

When considered geographically, the state factors discussed previously outline the global distribution of Spodosols. Thus, Spodosols occur mainly in cool, humid climates, generally under forest vegetation and on freely draining parent materials (Figure 33.58). The most widespread and dense concentrations of Spodosols occur in Canada and the northern Great Lakes region, Russia, Scandinavia and northern Europe. At appropriate (high) elevations, Spodosols are found on just about all of Earth's major mountain ranges. Aquods are also common on wet landscapes of the humid and subhumid tropics and subtropics. In areas where the climate or vegetation is marginal for the formation of well-developed Spodosols, weakly developed Spodosols or soils with spodic-like profiles can still be observed,

provided the remaining state factors favor podzolization. For example, Spodosols exist in cold, polar deserts (Ugolini, 1986; Blume et al., 1997).

33.9.7 Characteristics and Classification of Spodosols

Most Spodosols meet the NRCS classification criteria by having a spodic horizon. The spodic horizon is characterized by illuvial accumulations of Al and organic matter, most of which is translocated from overlying eluvial horizons above, and accumulates in amorphous forms. Illuvial Fe is usually associated with the Al and organic matter, although Fe is not necessary to meet many of the various worldwide "spodic/podzol" classification criteria. These illuvial substances, rich in Al, Fe, and various forms of illuvial organic materials, are referred to as "spodic materials." By definition, spodic horizons must be composed of 85% or more spodic materials (Soil Survey Staff, 1999). The spodic material concept embodies the podzolization process by requiring the materials to have (1) relatively low pH values and high organic matter contents and (2) red colors (usually) if they underlie an albic (E) horizon. In soils that lack an albic horizon, the pH, organic matter content, and color criteria are still required, but the spodic materials must *also* exhibit one or more of the following properties: (1) cementation, (2) cracked coatings on sand grains, (3) significantly more Al and Fe than an overlying horizon, determined by ammonium oxalate extraction, or (4) high optical densities of the liquid ammonium oxalate extract (an indication of amorphous materials, especially organic matter; Stanley and Ciolkosz, 1981; Daly, 1982; Mokma, 1993).

Many soils with a spodic horizon also have andic properties and are placed in the Andisol order (Flach et al., 1980; Soil Survey Staff, 1999). Because Andisols and Spodosols are pedogenically similar, many Andic subgroups of Spodosols have been defined (Table 33.44), and vice versa. Provision is also made for soils in which the spodic horizon is shallow and thus has been plowed. If the Ap horizon of such soils contains 85% or more spodic materials, the soil may classify as a Spodosol instead of an Andisol.

Some Spodosols have only a placic horizon, or a placic and a spodic horizon. The placic horizon is a thin (1–25 cm thick), black to dark reddish, cemented pan with a sharp upper boundary but diffuse lower boundary; it is cemented by Fe (or Fe and Mn) and organic matter (Clayden et al., 1990; Hseu et al., 1999). The pan is often wavy and even may bifurcate. Placic horizons may form as reduced iron, mobilized in surface horizons, is translocated downward in the profile and

TABLE 33.44 Listing of All the Current (1999) Suborders, Great Groups, and Subgroups of Spodosols

Suborders	Great Groups	Subgroups
Aquods	Cryaquods	Aeric, Andic, Duric, Entic, Lithic, Pergelic, Placic, Sideric, Typic
	Alaquods	Aeric, Alfic, Alfic Arenic, Arenic, Arenic Ultic, Arenic Umbric, Duric, Grossarenic, Histic, Lithic, Typic, Ultic
	Fragiaquods	Argic, Histic, Plagganthreptic, Plaggeptic, Typic
	Placaquods	Andic, Typic
	Duraquods	Andic, Histic, Typic
	Epiaquods	Alfic, Andic, Histic, Lithic, Typic, Ultic, Umbric
	Endoaquods	Andic, Argic, Histic, Lithic, Typic, Umbric
Cryods	Placocryods	Andic, Humic, Typic
	Duricryods	Andic, Aquandic, Aquic, Humic, Oxyaquic, Typic
	Humicryods	Andic, Aquandic, Aquic, Lithic, Oxyaquic, Pergelic, Typic
	Haplocryods	Andic, Aquandic, Aquic, Entic, Lithic, Oxyaquic, Pergelic, Typic
Humods	Placohumods	Andic, Cryic, Typic
	Durihumods	Andic, Typic
	Fragihumods	Typic
	Haplohumods	Andic, Arenic, Arenic Ultic, Entic, Grossarenic, Grossarenic Entic, Lithic, Plagganthreptic, Plaggeptic, Typic, Ultic
Orthods	Placorthods	Typic
	Durorthods	Andic, Typic
	Fragiorthods	Alfic, Alfic Oxyaquic, Aquentic, Aquic, Cryic, Entic, Humic, Oxyaquic, Plagganthreptic, Plaggeptic, Typic, Ultic
	Alorthods	Alfic, Arenic, Arenic Ultic, Entic, Entic Grossarenic, Grossarenic, Grossarenic Entic, Oxyaquic, Plagganthreptic, Plaggeptic, Typic, Ultic
	Haplorthods	Alfic, Alfic Oxyaquic, Andic, Aqualfic, Aquentic, Aquic, Duric, Entic, Entic Lithic, Fragiaquic, Fragic, Humic, Lamellic, Lamellic Oxyaquic, Lithic, Oxyaquic, Oxyaquic Ultic, Typic, Ultic

Source: Soil Survey Staff. 1999. Soil taxonomy. USDA-NRCS agriculture handbook no. 436. U.S. Government Printing Office, Washington, DC.

oxidized and precipitated in the B horizon. This precipitated Fe can adsorb soluble organic matter, but not necessarily form organometallic complexes (Hseu et al., 1999, Soil Survey Staff, 1999). The iron is usually present as ferrihydrite and poorly crystalline goethite. Placic horizons typically form in soils at landscape positions where conditions change from reducing to oxidizing. The reducing conditions, necessary for placic horizon genesis, are associated with (1) high rainfall and cool temperatures, typical of perhumid climates (and often also associated with Histosols) and (2) a slowly permeable subsurface layer (Lapen and Wang, 1999).

33.9.7.1 Taxonomic Groupings of Spodosols

In a review of Spodosols, Rourke et al. (1988) commonly referred to a "warm" (thermic, hyperthermic, or isohyperthermic temperature regimes) vs. "cold" (frigid, mesic, or isomesic temperature regimes) Spodosol dichotomy. Although this level of distinction is not taxonomically observed at the suborder

or great group level, we agree it is a useful way of considering Spodosols because, globally, these soils are mainly found in two distinctly different climatic settings: subtropical/tropical and cool-cold midlatitude locations. In subtropical and tropical locations, they tend to be located on wet sites (Goldin and Edmonds, 1996), whereas in cooler locations Spodosols occur in all landscape positions. Most "warm" Spodosols are Aquods, Orthods, and Humods; many of these soils lack a Bs horizon and have Bh horizons instead (Rourke et al., 1988). "Cool" Spodosols tend to have thinner profiles, and Bs, Bh, and/or Bhs horizons. Table 33.44 provides a complete listing of the suborders, great groups, and subgroups of Spodosols. Figure 33.64 shows the typical profile morphologies for the most common Spodosol subgroups (Soil Survey Staff, 1999).

33.9.7.1.1 Orthods

Orthods are the most common Spodosol suborder (Rourke et al., 1988). They typically have A–E–Bhs–Bs–BC–C horizon
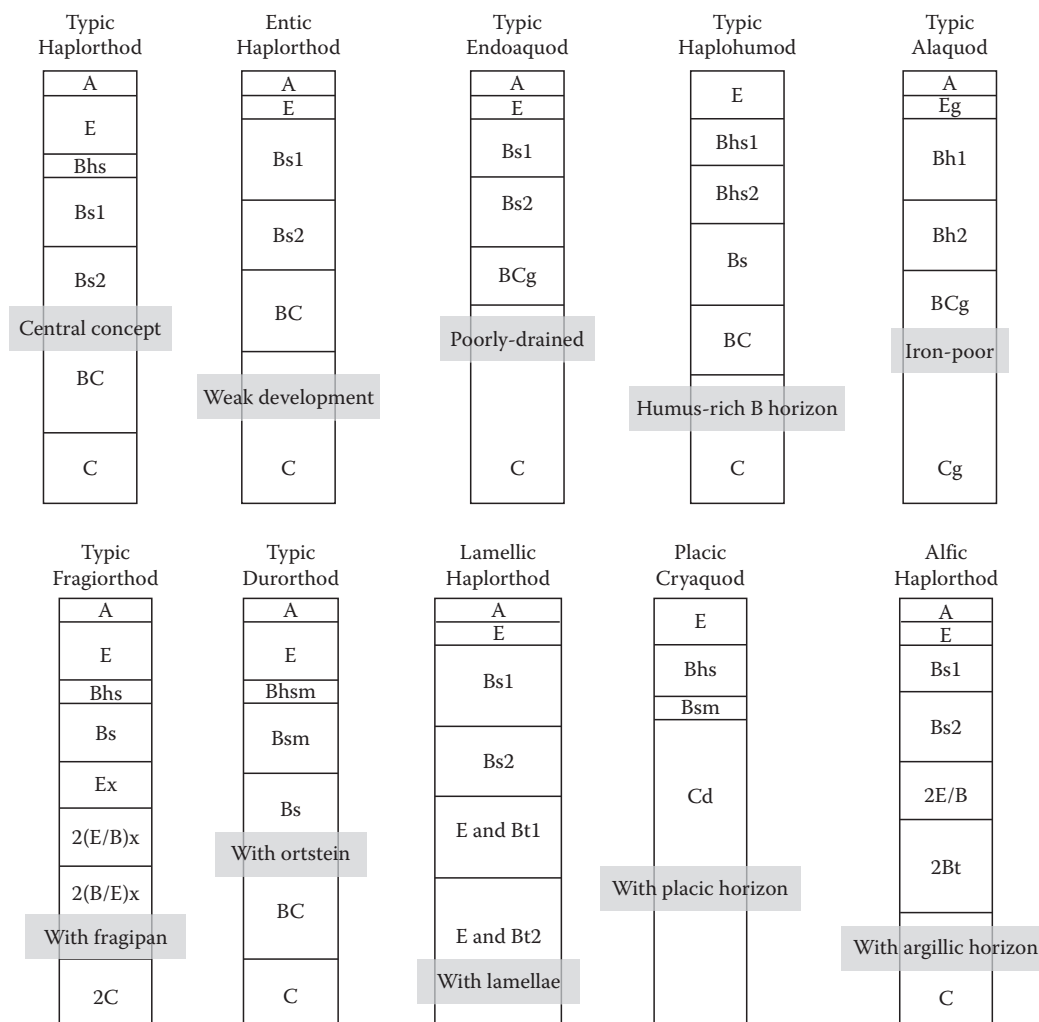
FIGURE 33.64 Typical horizonation patterns in common Spodosol profiles, illustrating the variations found naturally in extragrade and intergrade subgroups. Natural undulations in horizon boundaries, as well as O horizons, have been omitted to keep the diagrams simple and straightforward.

sequences, although the Bhs horizon is usually lacking in weakly developed pedons (Base and Brasher, 1990; Schaetzl, 2002).

33.9.7.1.2 Aquods

Aquods have a shallow, often fluctuating, water table. They are the most prevalent Spodosol suborder in subtropical and tropical regions and are a common admixture in other Spodosol landscapes in low-lying landscape positions.

33.9.7.1.3 Humods

Humods have high organic matter contents in the B horizon, usually concentrated in the upper B. In mesic and colder climates, Humods are most common in areas with very humid climates and lush vegetative growth.

33.9.7.1.4 Cryods

Cryods are Spodosols with a cryic soil temperature regime and include soils that have a mean annual temperature <8°C, but lack permafrost.

33.9.8 Management of Spodosol Landscapes

Many Spodosol landscapes are in natural or managed forest systems, where harvest of forest products, recreation, and tourism are the major economic activities. The cool climate of these landscapes, coupled with low soil fertility and sandy textures, tends to limit agricultural options on Spodosols. Particularly, restrictive are the low pH values of these soils, as well as low CEC and available P levels. Many Spodosols adsorb P and K, thereby restricting the ability of plants to obtain these nutrients (Bartlett and McIntosh, 1969; Bartlett, 1972; Villapando and Graetz, 2001). Some types of agriculture have been successful on Spodosols, particularly potatoes, small grains such as rye and barley, some hay crops, and irrigated berries, in some areas, especially where the soils are sandy loam or finer in texture. Blueberries are particularly well suited to cool, acid, sandy Aquods. A long growing season and climatic suitability for high value crops, for example, citrus, warm season vegetables, etc., makes it economically feasible to manage subtropical Spodosols (mainly Aquods) for intensive agriculture or horticulture (Harris et al., 2010). In these agricultural systems, management for some crops requires artificial drainage. Managing Aquods to minimize environment impacts of agricultural P is a challenge due to the minimal P retention capacity of their A and E horizons (Harris et al., 1996) that can result in P transport to water bodies (Allen, 1987; Mansell et al., 1991).

References

Abrahamson, W.G., and D.C. Hartnett. 1990. Pine flatwoods and dry prairies, p. 103–149. *In* R.L. Myers and J.J. Ewel (eds.) Ecosystems of Florida. University of Central Florida Press, Orlando, FL.

Alexander, E.B., C.L. Ping, and P. Krosse. 1994. Podzolisation in ultramafic materials in southeast Alaska. Soil Sci. 157:46–52.

Allen, L.H., Jr. 1987. Dairy-siting criteria and other options for wastewater management on high water-table soils. Soil Crop Sci. Soc. Fla. Proc. 47:108–127.

Anderson, H.A., M.L. Berrow, V.C. Farmer, A. Hepburn, J.D. Russell, and A.D. Walker. 1982. A reassessment of podzol formation processes. J. Soil Sci. 33:125–136.

Baldwin, M., C.E. Kellogg, and J. Thorp. 1938. Soil classification, p. 979–1001. *In* Soils and men. Yearbook of agriculture, USDA. U.S. Government Printing Office, Washington, DC.

Barrett, L.R. 1997. Podzolization under forest and stump prairie vegetation in northern Michigan. Geoderma 78:37–58.

Barrett, L.R. 2001. A strand plain soil development sequence in Northern Michigan, USA. Catena 44:163–186.

Barrett, L.R., and R.J. Schaetzl. 1992. An examination of podzolization near Lake Michigan using chronofunctions. Can. J. Soil Sci. 72:527–541.

Bartlett, R.J. 1972. Field test for spodic character based on pH-dependent phosphorous adsorption. Soil Sci. Soc. Am. Proc. 36:642–644.

Bartlett, R.J., and J.L. McIntosh. 1969. pH-dependent bonding of potassium by a spodosol. Soil Sci. Soc. Am. Proc. 33:535–539.

Base, S.R., and B.R. Brasher. 1990. Properties of United States Spodosols, p. 19–28. *In* J.M. Kimble and R.D. Yeck (eds.) Proc. Fifth Int. Soil Correlation Meet. (ISCOM) Characterization, classification, and utilization of Spodosols, Maine, Massachusetts, New Hampshire, New York, Vermont, and New Brunswick. October 1–14, 1987. Part A: Papers. USDA-SCS, Lincoln, NE.

Bloomfield, C. 1953. A study of podzolization. Part II. The mobilization of iron and aluminum by the leaves and bark of *Agathis australis* (Kauri). J. Soil Sci. 4:17–23.

Blume, H.-P., L. Beyer, M. Bölter, H. Erlenkeuser, E. Kalk, S. Kneesch, U. Pfisterer, and D. Schneider. 1997. Pedogenic zonation in soils of the southern circum-polar region. Adv. GeoEcol. 30:69–90.

Brasfield, J.F., V.W. Carlisle, and R.W. Johnson. 1973. Spodosols—Soils with a spodic horizon, p. 57–60. *In* S.W. Buol (ed.) Soils of the Southern States and Puerto Rico. Southern Cooperative Series Bulletin. Vol. 174. USDA.

Brooks, H.K. 1981. Guide to the physiographic regions of Florida. Cooperative Extension Service, University of Florida, Gainesville, FL.

Buurman, P., and L.P. van Reeuwijk. 1984. Proto-imogolite and the process of podzol formation: A critical note. J. Soil Sci. 35:447–452.

Calhoun, F.G., and V.W. Carlisle. 1973. Statistical analysis of Spodosol parameters. Soil Crop Sci. Soc. Fla. Proc. 33:139–143.

Cann, D.B., and E.P. Whiteside. 1955. A study of the genesis of a podzol-gray-brown podzolic intergrade soil profile in Michigan. Soil Sci. Soc. Am. Proc. 19:497–501.

Carlisle, V.W., and L.W. Zelazny. 1974. Pedon mineralogy of representative Florida Typic Quartzipsamments. Soil Crop Sci. Soc. Fla. Proc. 34:43–47.

Childs, C.W., R.L. Parfitt, and R. Lee. 1983. Movement of aluminum as an inorganic complex in some podzolised soils, New England. Geoderma 29:139–155.

Clayden, B., B.K. Daly, R. Lee, and G. Mew. 1990. The nature, occurrence and genesis of placic horizons, p. 88–104. *In* J.M. Kimble and R.D. Yeck (eds.) Proc. Fifth Int. Soil Correlation Meeting (ISCOM) Characterization, Classification, and Utilization of Spodosols, Maine, Massachusetts, New Hampshire, New York, Vermont, and New Brunswick. October 1–14, 1987. Part A: Papers. USDA-SCS, Lincoln, NE.

Cline, M.G. 1949. Profile studies of normal soils of New York. I. Soil profile sequences involving brown forest, gray–brown podzolic, and Brown podzolic soils. Soil Sci. 68:259–272.

Courchesne, F., and W.H. Hendershot. 1997. La Genese des podzols. Geogr. Phys. Quat. 51:235–250.

Dahlgren, R.A., and D.J. Marrett. 1991. Organic carbon sorption in arctic and subalpine Spodosol B horizons. Soil Sci. Soc. Am. J. 55:1382–1390.

Dahlgren, R.A., and F.C. Ugolini. 1989. Aluminum fractionation of soil solutions from unperturbed and tephra-treated Spodosols, Cascade Range, Washington, USA. Soil Sci. Soc. Am. J. 53:559–566.

Dahlgren, R.A., and F.C. Ugolini. 1991. Distribution and characterization of short-range-order minerals in Spodosols from the Washington Cascades. Geoderma 48:391–413.

Daly, B.K. 1982. Identification of podzol and podzolized soils in New Zealand by relative absorbance of oxalate extracts of A and B horizons. Geoderma 28:29–38.

Daniels, R.B., E.E. Gamble, and C.S. Holzhey. 1975. Thick Bh horizons in the North Carolina coastal plain: I. Morphology and relation to texture and soil ground water. Soil Sci. Soc. Am. Proc. 39:1177–1181.

DeConinck, F. 1980. Major mechanisms in formation of spodic horizons. Geoderma 24:101–128.

DeConinck, F., and A. Herbillon. 1969. Evolution mineralogique et chemique des fractions argileuses dans des Alfisols et dos Spodosols de la Campine (Belgiuque). Pedology 19:159–272.

Dubois, J.-M., Y.A. Martel, D. Côté, and L. Nadeau. 1990. Les ortsteins du Québec: Répartition Géographique, relations géomorphologiques et essai de datation. Can. Geog. 34:303–317.

Dubroeucq, D., and B. Volkoff. 1998. From Oxisols to Spodosols and Histosols: Evolution of the soil mantles in the Rio Negro basin (Amazonia). Catena 32:245–280.

Duchaufour, P.H., and B. Souchier. 1978. Roles of iron and clay in genesis of acid soils under a humid, temperate climate. Geoderma 20:15–26.

Dunn, C.P., G.R. Guntenspergen, and J.R. Dorney. 1983. Catastrophic wind disturbance in an old-growth hemlock-hardwood forest, Wisconsin. Can. J. Bot. 61:211–217.

Farmer, V.C. 1982. Significance of the presence of allophane and imogolite in podzol Bs horizons for podzolization mechanisms: A review. Soil Sci. Plant Nutr. 28:571–578.

Farmer, V.C., J.D. Russell, and M.L. Berrow. 1980. Imogolite and proto-imogolite allophane in spodic horizons: Evidence for a mobile aluminum silicate complex in podzol formation. J. Soil Sci. 31:673–684.

Farmer, V.C., J.O. Skjemstad, and C.H. Thompson. 1983. Genesis of humus B horizons in hydromorphic humus podzols. Nature 304:342–344.

Flach, K.W., C.S. Holzhey, F. DeConinck, and R.J. Bartlett. 1980. Genesis and classification of Andepts and Spodosols, p. 411–426. *In* B.K.G. Theng (ed.) Soils with variable charge. New Zealan Society of Soil Science, Palmerston North, NZ.

Franzmeier, D.P., and E.P. Whiteside. 1963. A chronosequence of podzols in northern Michigan. I. Ecology and description of pedons. Mich. State Univ. Agric. Exp. Stn. Quart. Bull. 46:2–20.

Freeland, J.A., and C.V. Evans. 1993. Genesis and profile development of success soils, northern New Hampshire. Soil Sci. Soc. Am. J. 57:183–191.

Fridland, V.M. 1958. Podzolization and illimerization (clay migration). Sov. Soil Sci. 1:24–32.

Gardner, D.R., and E.P. Whiteside. 1952. Zonal soils in the transition region between the podzol and gray–brown podzolic regions in Michigan. Soil Sci. Soc. Am. Proc. 16:137–141.

Garman, C.R., V.W. Carlisle, L.W. Zelazny, and B.C. Beville. 1981. Aquiclude-related spodic horizon development. Soil Crop Sci. Soc. Fla. Proc. 40:105–110.

Goldin, A., and J. Edmonds. 1996. A numerical evaluation of some Florida Spodosols. Phys. Geogr. 17:242–252.

Guillet, B., J. Rouiller, and B. Souchier. 1975. Podzolization and clay migration in Spodosols of eastern France. Geoderma 14:223–245.

Gustafsson, J.P., P. Bhattacharya, D.C. Bain, A.R. Fraser, and W.J. McHardy. 1995. Podzolisation mechanisms and the synthesis of imogolite in northern Scandinavia. Geoderma 66:167–184.

Harris, W.G. 2000. Hydrologically-linked Spodosol formation in the Southeastern United States, p. 331–342. *In* J.L. Richardson and M.J. Vepraskas (eds.) Wetland soils: Their genesis, hydrology, landscape, and separation into hydric and nonhydric soils. CRC Press, Boca Raton, FL.

Harris, W.G., and V.W. Carlisle. 1985. Clay mineralogical relationships in Florida haplaquods. Soil Sci. Soc. Am. J. 51:481–484.

Harris, W.G., V.W. Carlisle, and S.L. Chesser. 1987a. Clay mineralogy as related to morphology of Florida soils with sandy epipedons. Soil Sci. Soc. Am. J. 51:1673–1677.

Harris, W.G., V.W. Carlisle, and K.C.J. Van Rees. 1987b. Pedon zonation of hydroxy-interlayered minerals in ultic haplaquods. Soil Sci. Soc. Am. J. 51:1367–1372.

Harris, W.G., M. Chrysostome, T.A. Obreza, and V.D. Nair. 2010. Soil properties pertinent to horticulture in Florida. HortTechnology 20:10–19.

Harris, W.G., S.H. Crownover, and N.B. Comerford. 1995. Experimental formation of aquod-like features in sandy coastal plain soils. Soil Sci. Soc. Am. J. 59:877–886.

Harris, W.G., and K.A. Hollien. 1999. Changes in quantity and composition of crystalline clay across E–Bh boundaries of alaquods. Soil Sci. 164:602–608.

Harris, W.G., and K.A. Hollien. 2000. Changes across artificial E–Bh boundaries generated under simulated fluctuating water tables. Soil Sci. Soc. Am. J. 64:967–973.

Harris, W.G., R.D. Rhue, G. Kidder, R.C. Littell, and R.B. Brown. 1996. P retention as related to morphology of sandy coastal plain soils. Soil Sci. Soc. Am J. 60:1513–1521.

Hole, F.D. 1975. Some relationships between forest vegetation and podzol B horizons in soils of Menominee tribal lands, Wisconsin, U.S.A. Sov. Soil Sci. 7:714–723.

Holzhey, C.S., R.B. Daniels, and E.E. Gamble. 1975. Thick Bh horizons in the North Carolina coastal plain: II. Physical and chemical properties and rates of organic additions from surface sources. Soil Sci. Soc. Am. Proc. 39:1182–1187.

Hseu, Z.-Y., Z.-S. Chen, and Z.-D. Wu. 1999. Characterization of placic horizons in two subalpine forest inceptisols. Soil Sci. Soc. Am. J. 63:941–947.

Hseu, Z.-Y., C.G. Tsai, C.W. Lin, and Z.-S. Chen. 2004. Transitional soil characteristics of Ultisols and Spodosols in the subalpine forest of Taiwan. Soil Sci. 169:457–467.

Huang, P.M., and A. Violante. 1986. Influence of organic acids on crystallization and surface properties of precipitation products of Al, p. 159–221. *In* P.M. Huang and M. Schnitzer (eds.) Interactions of soil minerals with natural organics and soil microbes. Special Publication No. 17. SSSA, Madison, WI.

Hunckler, R.V., and R.J. Schaetzl. 1997. Spodosol development as affected by geomorphic aspect, Baraga County, Michigan. Soil Sci. Soc. Am. J. 61:1105–1115.

Isard, S.A., and R.J. Schaetzl. 1995. Estimating soil temperatures and frost in the lake effect snowbelt region, Michigan, USA. Cold Reg. Sci. Technol. 23:317–332.

Isard, S.A., R.J. Schaetzl, and J.A. Andresen. 2007. Soils cool as climate warms in Great Lakes region, USA. Ann. Assoc. Am. Geogr. 97:467–476.

Jakobsen, B.H. 1989. Evidence for translocations into the B horizon of a subarctic podzol in Greenland. Geoderma 45:3–17.

Jauhiainen, E. 1973. Age and degree of podzolization of sand soils on the coastal plain of northwest Finland. Comment. Biol. 68:5–32.

Jongkind, A.G., E. Velthorst, and P. Buurman. 2007. Soil chemical properties under kauri (*Agathis australis*) in the Waitakere Ranges, New Zealand. Geoderma 141:320–331.

Jonsson, B.G., and M. Dynesius. 1993. Uprooting in boreal spruce forests: Long-term variation in disturbance rate. Can. J. For. Res. 23:2383–2388.

Kabrick, J.M., M.K. Clayton, A.B. McBratney, and K. McSweeney. 1997. Cradle-knoll patterns and characteristics on drumlins in northeastern Wisconsin. Soil Sci. Soc. Am. J. 61:595–603.

Kodama, H., M. Schnitzer, and M. Jaakkimainen. 1983. Chlorite and biotite weathering by fulvic acid solutions in closed and open systems. Can. J. Soil Sci. 63:619–629.

Krzyszowska, A.J., M.J. Blaylock, G.F. Vance, and M.B. David. 1996. Ion chromatographic analysis of low molecular weight organic acids in Spodosol forest floor solutions. Soil Sci. Soc. Am. J. 60:1565–1571.

Lambert, J.L., and F.D. Hole. 1971. Hydraulic properties of an ortstein horizon. Soil Sci. Soc. Am. Proc. 35:785–787.

Langmaid, K.K. 1964. Some effects of earthworm invasion in virgin podzols. Can. J. Soil Sci. 44:34–37.

Lapen, D.R., and C. Wang. 1999. Placic and ortstein horizon genesis and peatland development, southeastern Newfoundland. Soil Sci. Soc. Am. J. 63:1472–1482.

Lee, F.Y., T.L. Yuan, and V.W. Carlisle. 1988. Nature of cementing materials in ortstein horizons of selected Florida Spodosols: I. Constituents of cementing materials. Soil Sci. Soc. Am. J. 52:1411–1418.

Li, S.Y., Z.-S. Cheng, and J.C. Liu. 1998. Subalpine loamy spodosols in Taiwan: Characteristics, micromorphology, and genesis. Soil Sci. Soc. Am. J. 62:710–716.

Lichter, J. 1998. Rates of weathering and chemical depletion in soils across a chronosequence of Lake Michigan sand dunes. Geoderma 85:255–282.

Lundström, U.S., N. van Breemen, and N. Bain. 2000a. The podzolization process: A review. Geoderma 94:91–107.

Lundström, U.S., N. van Breemen, D.C. Bain, P.A.W. van Hees, R. Giesler, J.P. Gustafsson, H. Ilvesniemi et al. 2000b. Advances in understanding the podzolization process resulting from a multidisciplinary study of three coniferous forest soils in the Nordic countries. Geoderma 94:335–353.

Lundström, U.S., N. van Breemen, and A.G. Jongmans. 1995. Evidence for microbial decomposition of organic acids during podzolization. Eur. J. Soil Sci. 46:489–496.

Mackney, D. 1961. A podzol development sequence in oakwoods and heath in central England. J. Soil Sci. 12:23–40.

Madsen, H.B., and P. Nørnberg. 1995. Mineralogy of four sandy soils developed under heather, oak, spruce and grass in the same fluvioglacial deposit in Denmark. Geoderma 64:233–256.

Mansell, R.S., S.A. Bloom, and B. Burgoa. 1991. Phosphorus transport with water flow in acid, sandy soils. *In* B. Jacob and M.Y. Corapcioglu (eds.) Transport processes in porous media. Kluwer Academic Publishers, Dordecht, the Netherlands.

McKeague, J.A., F. DeConinck, and D.P. Franzmeier. 1983. Spodosols, p. 217–252. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy. Elsevier, New York.

Meyers, N.L., and K. McSweeney. 1995. Influence of treethrow on soil properties in Northern Wisconsin. Soil Sci. Soc. Am. J. 59:871–876.

Mokma, D.L. 1993. Color and amorphous materials in Spodosols from Michigan. Soil Sci. Soc. Am. J. 57:125–128.

Mokma, D.L. 1997. Ortstein in selected soils from Washington and Michigan. Soil Survey Horiz. 38:71–75.

Mokma, D.L., J.A. Doolittle, and L.A. Tornes. 1994. Continuity of ortstein in sandy Spodosols, Michigan. Soil Surv. Horiz. 35:6–10.

Mokma, D.L., and C.V. Evans. 2000. Spodosols, p. E-307–E-321. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Mokma, D.L., and G.F. Vance. 1989. Forest vegetation and origin of some spodic horizons, Michigan. Geoderma 43:311–324.

Mokma, D.L., M. Yli-Halla, and K. Lindkvist. 2004. Podzol formation in sandy soils of Finland. Geoderma 120:259–272.

Mossin, L., B.T. Jensen, and P. Nørnberg. 2001. Altered podzolization resulting from replacing heather with Sitka Spruce. Soil Sci. Soc. Am. J. 65:1455–1462.

Muir, A. 1961. The podzol and podzolic soils. Adv. Agron. 13:1–57.

Nielsen, K.E., K. Dalsgaard, and P. Nornberg. 1987a. Effects on soils of an oak invasion of a *Calluna* heath, Denmark, I. Morphology and chemistry. Geoderma 41:79–95.

Nielsen, K.E., K. Dalsgaard, and P. Nornberg. 1987b. Effects on soils of an oak invasion of a *Calluna* heath, Denmark, II. Changes in organic matter and cellulose composition. Geoderma 41:97–106.

Paton, T.R., P.B. Mitchell, D. Adamson, R.A. Buchanon, M.D. Fox, and G. Bowman. 1976. Speed of podzolisation. Nature 260:601–602.

Petersen, L. 1976. Podzols and podzolization. Royal Veterinary and Agricultural University, Copenhagen, Denmark.

Petersen, L. 1984. The podzol concept, p. 12–20. *In* P. Buurman (ed.) Podzols. Van Nostrand Reinhold Co., New York.

Riise, G., P. Van Hees, U. Lundstrom, and L.T. Strand. 2000. Mobility of different size fractions of organic carbon, Al, Fe, Mn and Si in Podzols. Geoderma 94:237–247.

Rourke, R.V., B.R. Brasher, R.D. Yeck, and F.T. Miller. 1988. Characteristic morphology of U.S. Spodosols. Soil Sci. Soc. Am. J. 52:445–449.

Sauer, D., H. Sponagel, M. Sommer, L. Giani, R. Jahn, and K. Stahr. 2007. Podzol: Soil of the year 2007. A review on its genesis, occurrence, and functions. J. Soil Sci. Plant Nutr. 170:581–597.

Sauer, D., I. Sshulli-Maurer, R. Sperstad, R. Sorenson, and K. Stahr. 2008. Podzol development with time in sandy beach deposits. J. Soil Sci. Plant Nutr. 171:483–497.

Schaetzl, R.J. 1986. Complete soil profile inversion by tree uprooting. Phys. Geogr. 7:181–189.

Schaetzl, R.J. 1990. Effects of treethrow microtopography on the characteristics and genesis of Spodosols, Michigan, USA. Catena 17:111–126.

Schaetzl, R.J. 1992. Beta spodic horizons in podzolic soils (lithic haplorthods and haplohumods). Pedology 42:271–287.

Schaetzl, R.J. 2002. A Spodosol–Entisol transition in northern Michigan: Climate or vegetation? Soil Sci. Soc. Am. J. 66:1272–1284.

Schaetzl, R.J., and S.N. Anderson. 2005. Soils: Genesis and geomorphology. Cambridge University Press, New York.

Schaetzl, R.J., S.F. Burns, T.W. Small, and D.L. Johnson. 1990. Tree uprooting: Review of types and patterns of soil disturbance. Phys. Geogr. 11:277–291.

Schaetzl, R.J., and L.R. Follmer. 1990. Longevity of treethrow microtopography: Implications for mass wasting. Geomorphology 3:113–123.

Schaetzl, R.J., and S.A. Isard. 1991. The distribution of Spodosol soils in southern Michigan: A climatic interpretation. Ann. Assoc. Am. Geogr. 81:425–442.

Schaetzl, R.J., and S.A. Isard. 1996. Regional-scale relationships between climate and strength of podzolization in the Great Lakes region, North America. Catena 28:47–69.

Schaetzl, R.J., L.R. Mikesell, and M.A. Velbel. 2006. Soil characteristics related to weathering and pedogenesis across a geomorphic surface of uniform age in Michigan. Phys. Geogr. 27:170–188.

Schaetzl, R.J., and D.L. Mokma. 1988. A numerical index of podzol and podzolic soil development. Phys. Geogr. 9:232–246.

Schaetzl, R.J., and C. Schwenner. 2006. An application of the Runge "Energy Model" of soil development in Michigan's upper peninsula. Soil Sci. 171:152–166.

Schnitzer, M., and H. Kodama. 1976. The dissolution of micas by fulvic acid. Geoderma 15:381–391.

Small, T.W. 1997. The Goodlett–Denny mound: A glimpse at 45 years of Pennsylvania treethrow mound evolution with implications for mass wasting. Geomorphology 18:305–313.

Soil Survey Staff. 1975. Soil taxonomy. USDA agriculture handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1999. Soil taxonomy. USDA-NRCS agriculture handbook no. 436. U.S. Government Printing Office, Washington, DC.

Stanley, S.R., and E.J. Ciolkosz. 1981. Classification and genesis of Spodosols in the Central Appalachians. Soil Sci. Soc. Am. J. 45:912–917.

Stützer, A. 1998. Early stages of podzolisation in young aeolian sediments, western Jutland. Catena 32:115–129.

Stone, E.L., W.G. Harris, R.B. Brown, and R.J. Kuehl. 1993. Carbon storage in Florida Spodosols. Soil Sci. Soc. Am. J. 57:179–182.

Tan, Z., W.G. Harris, and R.S. Mansell. 1999. Water table dynamics across an aquod–udult transition in Florida flatwoods. Soil Sci. 164:10–17.

Thompson, C.H. 1992. Genesis of podzols on coastal dunes in southern Queensland: I. Field relationships and profile morphology. Aust. J. Soil Res. 30:593–613.

Turcotte, D.E., and T.H. Butler, Jr. 2006. Folists and humods off the coast of eastern Maine. Soil Surv. Horiz. 47:5–9.

Ugolini, F.C. 1986. Pedogenic zonation in the well-drained soils of the arctic regions. Quat. Res. 26:100–120.

Ugolini, F.C., and R. Dahlgren. 1987. The mechanism of podzolization as revealed by soil solution studies, p. 195–203. *In* D. Righi and A. Chauvel (eds.) Podzols et Podzolisation. Association Francaise pour l'Etude du Sol, and Institut National de la Recherche Agronomique, Paris, France.

Ugolini, F.C., H. Dawson, and J. Zachara. 1977. Direct evidence of particle migration in the soil solution of a podzol. Science 198:603–605.

Van Hees, P.A.W., U.S. Lundström, and R. Giesler. 2000. Low molecular weight organic acids and their Al-complexes in soil solution—Composition, distribution and seasonal variation in three podzolized soils. Geoderma 94:173–200.

Vance, G.F., D.L. Mokma, and S.A. Boyd. 1986. Phenolic compounds in soils of hydrosequences and developmental sequences of Spodosols. Soil Sci. Soc. Am. J. 50:992–996.

VandenBygaart, A.J., and R. Protz. 1995. Soil genesis on a chronosequence, Pinery Provincial Park, Ontario. Can. J. Soil Sci. 75:63–72.

Veneman, P.L.M., P.V. Jacke, and S.M. Bodine. 1984. Soil formation as affected by pit and mound microrelief in Massachusetts, USA. Geoderma 33:89–99.

Villapando, R.R., and D.A. Graetz. 2001. Phosphorus sorption and desorption properties of the spodic horizon from selected Florida Spodosols. Soil Sci. Soc. Am. J. 65:331–339.

Wang, C., G.J. Beke, and J.A. McKeague. 1978. Site characteristics, morphology and physical properties of selected ortstein soils from the Maritime Provinces. Can. J. Soil Sci. 58:405–420.

Wang, C., J.A. McKeague, and H. Kodama. 1986. Pedogenic imogolite and soil environments: Case study of Spodosols in Quebec, Canada. Soil Sci. Soc. Am. J. 50:711–718.

Watts, F.C., and M.E. Collins. 2008. Soils of Florida. SSSA, Book and Multimedia Publishing Committee, Madison, WI.

Wright, J.R., and M. Schnitzer. 1963. Metallo-organic interactions associated with podzolization. Soil Sci. Soc. Am. Proc. 27:171–176.

33.10 Aridisols

H. Curtis Monger

Randal J. Southard

Janis L. Boettinger

33.10.1 Concept of Aridisols

Vast areas of the Earth, roughly one-third of the total land surface, are in arid regions. Such aridity results from their position relative to mountain ranges that scavenge moisture, their latitude around 30° where convective Hadley cells create enduring high pressures, or their great distances from large bodies of water (Ahrens, 1991; Figure 33.65). Rainfall in these dry climates is insufficient to maintain perennial streams, and the soils typically have little organic matter, contain pedogenic carbonate, and because their soil profiles are of the "nonflushing type"

(Rode, 1962), contain soluble minerals otherwise leached from humid soils (Southard, 2000). The dominant soil order in arid regions is the Aridisol order, which, with the exception of the Entisol order, covers more of the Earth's surface than any soil order in the *Soil Taxonomy* system (Wilding, 2000).

The central idea of Aridisols is that they do not have water available to mesophytic plants for long periods. This lack of soil moisture may be exacerbated by low water-holding capacity due to a shallow soil depth or restricted infiltration or by low osmotic potential due to salinity. The lack of water also inhibits soil formation.

33.10.1.1 Aridisols vs. Desert Soils

As the order name implies, Aridisols are associated with arid climates, but not all soils in arid regions are Aridisols. In fact, Aridisols occupy less than half of the arid land area (Figure 33.65). Other soil orders in arid climates are the Entisols, Vertisols, Oxisols, and Andisols (Table 33.45). The rationale for having Entisols in arid climates was to preserve the *azonal* concept, which identified soils lacking genetic horizons regardless of the bioclimatic zone in which they occur (Smith, 1986). The rational for having Oxisols in arid climates was that oxic properties, even if irrigated, would remain limiting and the management difficulties would be similar to other Oxisols; thus, they should grouped together with other Oxisols despite their occurrence in dry climates (Smith, 1986). Similar reasons were presented for the shrink–swell properties of Vertisols and the andic properties of Andisols. In some minor areas within the aridic moisture regime, there are soils with a mollic epipedon. These soils are grouped with the Mollisols rather than Aridisols (Smith, 1986; Ahrens and Eswaran, 2000).

33.10.1.2 History of the Aridisol Order

The rationale for the Aridisol order of *Soil Taxonomy* was to separate on a soil map the "sown from the unsown"; that is, the land that can be cultivated from the land that can only be grazed (Smith, 1986). Using climate as the basis for a soil order is somewhat unique to the *Soil Taxonomy* system. Other classification
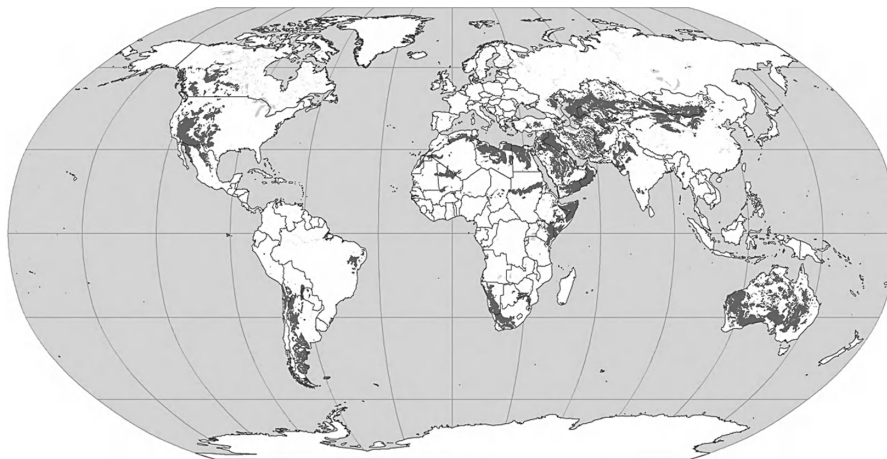
FIGURE 33.65 Global distribution of Aridisols. (Courtesy of USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC, 2010.)

TABLE 33.45 Global Extent of Aridisols (km²) and Vertisols, Oxisols, Andisols, and Entisols Occurring within the Aridic Soil Moisture Regime

Soil Order	Suborder	Africa	Asia	Australia/ Oceania	Europe	South America	Central America	North America	Caribbean	Global
Aridisols										14,264,316
	Cryids	549	398,041	—	100	142,088	—	516,759	—	1,057,537
	Salids	138,542	954,801	69,044	7,993	28,304	—	9,988	31	1,208,703
	Durids	—	—	—	—	—	—	—	—	—
	Gypsids	308,871	298,414	—	1,806	—	—	—	—	609,091
	Argids	292,435	1,107,480	1,395,126	66,339	471,738	2,039	1,012,842	1,446	4,349,445
	Calcids	1,598,740	1,988,228	538,948	1,053	96,171	57	178,254	52	4,401,503
	Cambids	715,612	1,049,670	294,538	29,594	388,308	486	158,233	1,596	2,638,037
Vertisols	Torrerts	161,019	54,772	508,848	—	5,204	133	35,705	1,337	767,018
Oxisols	Torrox	7,481	—	3,527	—	13,461	—	—	1,316	25,785
Andisols	Torrands	942	134	—	—	64	—	103	—	1,243
Entisols	In Aridic moisture regimes									11,266,409
								Total arid soils		26,324,771
								Total ice-free land		130,268,185

Sources: Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS. U.S. Government Printing Office, Washington, DC. Data courtesy of USDA-Natural Resources Conservation Service, Soil Survey Division, World Soil Resources, 2009.

systems, such as the World Reference Base, do not have an equivalent to the Aridisol order (IUSS Working Group WRB, 2006). Yet the delineation based on climate was inherited from the concept of *zonal* soils, the idea that certain kinds of soils are associated with certain kinds of bioclimatic zones—a concept developed by N.M. Sibirtsev in the late 1800s (Sibirtsev, 1898; Buol et al., 1980).

By the mid-1930s, soils in the arid zone of the United States were mapped as "northern gray desert soils" and "southern gray desert soils" (Marbut, 1935). With the publication of a revised classification system in 1938, the northern gray desert soils were changed to "sierozems" and the southern desert soils, which were associated with creosotebush, were changed to "red desert soils" (Baldwin et al., 1938). With the publication of *Soil Taxonomy* (1975), these soils became Aridisols with two suborders: Argids, if an argillic or natric horizon was present, or Orthids if not (Witty, 1990). In the 1980s and 1990s, the International Committee on Aridisols (ICOMID) developed seven suborders based on important diagnostic horizons and temperature: Cryids, Salids, Durids, Gypsids, Argids, Calcids, and Cambids.

33.10.1.3 Concept of the Aridic Moisture Regime

The concept of the aridic moisture regime is not totally synonymous with the definition of *arid*, which has several definitions. *Arid* in some definitions is based on precipitation alone, as in the arid-semiarid boundary at 250 mm of annual rainfall (Bull, 1991). In other definitions, the arid boundary is based on a combination of precipitation and temperature, as in the aridity index of 10 (Schmidt, 1979). Still, other definitions, such as the Köppen system, combine precipitation, temperature, and seasonality of precipitation (Strahler and Strahler, 1987). In contrast, the soil moisture regime is based on the number of days a soil is dry (i.e., soil moisture held at tensions less than −1500 kPa). Soil moisture is used because the availability of soil

moisture to plants is affected by runoff–runon, infiltration, and available water-holding capacity rather than annual precipitation per se. Generally speaking, however, most soils that have an aridic moisture regime do occur in areas of arid climates (Soil Survey Staff, 1999). Exceptions include semiarid soils with physical properties that keep them dry, such as physical crusts that prevent infiltration, steep slopes where runoff is high, low water-holding capacity in shallow soils, or low osmotic potential in strongly saline soils.

The aridic moisture regime is based on the duration of dryness in the *soil moisture control section* during the period the soil is warm enough for plant growth. The soil moisture control section will vary among soils because it is based on how deep water infiltrates. The top boundary of the soil moisture control section is the depth reached by the wetting front when 2.5 cm (1 in.) of water is applied for 24 h. The bottom boundary is the depth reached by the wetting front when 7.5 cm (3 in.) of water is applied for 48 h. Because these depths are largely a function of soil texture, finer-textured soils, such as those with fine-loamy, coarse-silty, fine-silty, or clayey textural classes, have moisture control sections that lie between approximately 10 and 30 cm (Herbel and Gile, 1973). In contrast, coarse-loamy soils have moisture control sections roughly between 20 and 60 cm, and sandy textural classes have boundaries between approximately 30 and 90 cm depending on coarse fragments, which deepen these limits for all textures.

By definition, a soil has an aridic moisture regime if that soil's control section is dry more than half the year (in normal years) when the temperature at 50 cm is above 5°C. The criterion of "above 5°C" was added because cold temperatures, rather than dryness, can be the limiting factor to plant growth (Smith, 1986). Additionally, the soil moisture control section is "Moist in some or all parts for less than 90 consecutive days when the
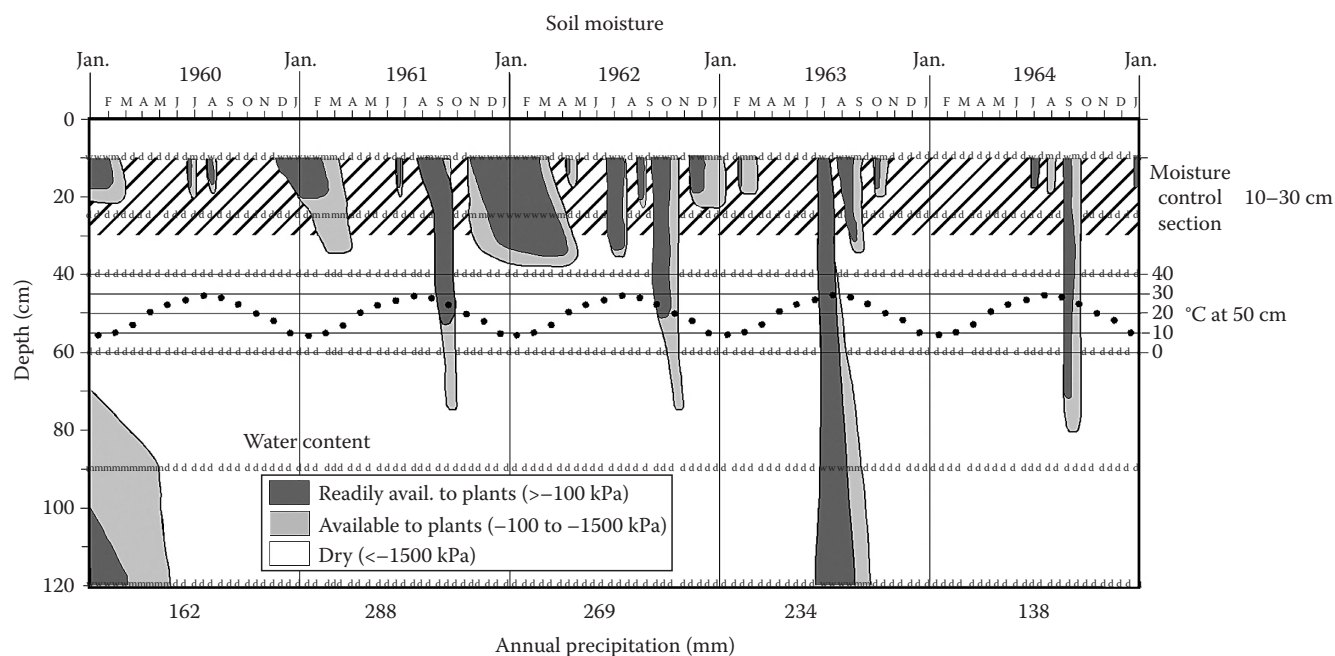
FIGURE 33.66 Soil with an aridic soil moisture regime. The moisture control section (10–30 cm) is dry for over 50% of the time in "normal years" (Soil Survey Staff, 1999). The year 1962, however, was not a normal year and does not meet 50%-dry criterion. Neither did it meet the criterion that the soil cannot be moist for more than 3 months (90 consecutive days). The temperature at 50 cm is above 5°C and 8°C (Soil Survey Staff, 1999). Moisture data from Herbel et al. (1994). Temperature data from Gile and Grossman (1979). Soil is a fine, mixed, superactive, thermic Ustic Calciargid. (From Soil Survey Staff. 1999. Soil taxonomy—A basic system of soil classification for making and interpreting soil surveys. USDA agriculture handbook no. 436. USDA-SCS. U.S. Government Printing Office, Washington, DC.)

soil temperature at 50 cm is above 8°C" (Soil Survey Staff, 1999). In other words, if a soil is moist, "not dry" (i.e., water held at tensions greater than −1500 kPa), for more than 3 months (90 consecutive days), the soil is too moist to have an aridic moisture regime. A graphical illustration of an aridic soil moisture regime is shown in Figure 33.66.

33.10.2 Geographical Distribution

The total ice-free land area on Earth is 130,268,185 km². Of that area, 20% or 26,324,771 km² is occupied by soils that occur within the aridic moisture regimes. Aridisols occupy about half (54%) of this area (Table 33.45; Figure 33.65). The rest is occupied by Entisols (43%), Vertisols (3%), Oxisols (0.1%), and minute amounts of Andisols (0.005%) with even smaller areas of Gelisols and Mollisols not shown in Table 33.45. Of the total extent of Aridisols, most (41%) occur in Asia, stretching eastward from the Middle East to northern China (Figure 33.65). Africa contains the next largest extent of Aridisols (21%), followed by Australia (16%), North America (13%), South America (8%), and 1% or less in Europe and Central America. In Asia, most suborders are Calcids, Argids, and Cambids that occur in northern China, Mongolia, Kazakhstan, Iran, Afghanistan, Pakistan, Iraq, and Saudi Arabia, with smaller amounts of Salids in Uzbekistan, Turkestan, and Iran, and Gypsids in localized areas throughout Asia, especially Syria and Turkey. Extensive areas of Cryids exist in the higher elevations of China and Mongolia. Substantial areas of Cryids also exist in North and South

America where they occur at both high elevations and high latitudes. Africa, Australia, North and South America all have expansive areas of Calcids, Argids, and Cambids with smaller areas of Salids and Gypsids, which also occur in a few areas of southern Europe (Eswaran and Zi-Tong, 1991).

Aridisols in mountainous regions form complex patterns with semiarid soils that have ustic and xeric moisture regimes at higher elevations and in topographic low-lying areas that receive run-on water. These topographically controlled soil patterns are also influenced by seasonality of precipitation. For example, in North America, the most southerly Aridisols occur in the subtropical high pressure belt that receives most of its precipitation in the summer from subtropical convective thunderstorms and border the ustic soil moisture regime. Farther north, Aridisols of the intermountain west occur in more continental climates, but precipitation is still dominated by summer thunderstorms driven by orographic lifting in the Rocky Mountains. To the west, Aridisols of the Great Basin region receive most of their precipitation in the winter from Pacific frontal storms and have Aridisols that border the xeric moisture regime.

33.10.3 Diagnostic Horizons and Classification

33.10.3.1 Epipedons

Virtually, all Aridisols have the ochric epipedon, which by definition is too thin, too dry, too light in color, contains too little organic carbon, has too high an *n* value or melanic index, or is too hard to qualify as any of the other epipedons (Soil

TABLE 33.46 Abbreviated Definitions of Diagnostic Subsurface Horizons in Aridisols

Horizon	Properties
Argillic	Clay accumulation relative to overlying horizon, evidence of clay translocation, minimum thickness of 7.5 cm if loamy or clayey, 15 cm if sandy (Bt)
Calcic	$CaCO_3$ accumulation, generally ≥15 cm thick, ≥15% $CaCO_3$ equivalent, ≥5% more $CaCO_3$ than underlying horizon. In coarse-textured soil > 5% $CaCO_3$ equivalent (Bk or Bkk)
Cambic	Evidence of alteration, reddening, development of soil structure, loss of carbonates (Bw, B, or Bq)
Duripan	Silica-cemented, does not slake in water or acid, evidence of opal accumulation (Bqm)
Gypsic	Gypsum accumulation, ≥15 cm thick, ≥5% gypsum, ≥1% visible secondary gypsum, thickness in centimeter multiplied by % gypsum has product ≥ 150 (By or Byy)
Petrocalcic	$CaCO_3$-cemented, ≥10 cm thick, or ≥1 cm thick if a laminar cap is underlain directly by bedrock, does not slake in water but dissolves in acid, very hard or harder when dry (Bkm or Bkkm)
Petrogypsic	Gypsum-cemented, ≥10 cm thick, ≥5% gypsum and thickness in cm multiplied by % gypsum has product ≥50, roots cannot enter except in vertical fractures that have a horizontal spacing of ≥10 cm (Bym or Byym)
Natric	Properties of argillic horizon, plus columnar or prismatic structure, or degraded blocky structure, sodium adsorption ratio (SAR) ≥ 13 or exchangeable sodium percentage (ESP) ≥ 15 (Btn)
Salic	Accumulation of salts more soluble than gypsum, ≥15 cm thick, electrical conductivity (EC) ≥ 30 dS/m, thickness in centimeter multiplied by EC has product ≥ 900 (Az, Bz, or Cz)

Sources: Soil Survey Staff. 1999. Soil taxonomy—A basic system of soil classification for making and interpreting soil surveys. USDA agriculture handbook no. 436. USDA-SCS. U.S. Government Printing Office, Washington, DC; Soil Survey Staff. 1975. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. Agriculture handbook no. 436. USDA-SCS. U.S. Government Printing Office, Washington, DC, 1975.

Horizon designations typically associated with diagnostic horizons are given in parentheses.

Survey Staff, 1999). Dark anthropic epipedons can also occur in Aridisols, but are very rare and localized. If irrigated for long time periods, a mollic-like epipedon can form, but is grouped with the anthropic epipedons rather than the mollic for taxonomic purposes (Soil Survey Staff, 1999). Surface horizons of Aridisols are uniquely different from surface horizons of more humid soils because of their desert pavement, desert varnish, vesicular A horizon, or biological soil crusts.

33.10.3.2 Subsurface Horizons

Subsurface horizons include the cambic, argillic, natric, calcic, petrocalcic, gypsic, petrogypsic and salic horizons, and the duripan. None of these horizons is unique to Aridisols, although the occurrence of salic, gypsic, and petrogypsic horizons in other soil orders is rare. A summary of the properties of these horizons is given in Table 33.46. Many combinations of these horizons are possible in Aridisols depending on the interaction of the five soil-forming factors. As an example, Figure 33.67 shows a petrocalcic, calcic, and argillic horizons, as well as the ochric epipedon in two Aridisols in the southwestern United States. Despite being in arid climates, Aridisols can have prominent pedogenic horizons that display advanced stages of soil development (cf., Huenneke and Noble, 1996).

33.10.3.3 Suborders, Great Groups, and Subgroups

The nine subsurface horizons (Table 33.46) and the cryic soil temperature regime are used to differentiate seven suborders, presented below in the sequence they are keyed out in *Soil Taxonomy*: (1) Cryids (Aridisols in cold areas), (2) Salids (accumulations of salts more soluble than gypsum), (3) Durids (accumulations of and cementation by secondary silica), (4) Gypsids (accumulations of gypsum), (5) Argids (accumulations of illuvial clay), (6) Calcids

(accumulations of pedogenic carbonates), and (7) Cambids (weak translocation or transformation of subsoil material). The Cryids occur at high latitudes and high elevations and are common in the intermountain western region of the United States (Hipple et al., 1990). The Argids, Durids, and Petrocalcids (petrocalcic horizon) occur on the oldest geomorphic surfaces, usually late Pleistocene and older, as on the least eroded parts of dissected alluvial fans. Cambids and other Calcids are found on geologically younger side slopes and surfaces of intermediate age, usually latest Pleistocene or younger. Gypsids and Salids often occur near playa margins, where salts are concentrated at or near the soil surface by upward flux from a water table. Less commonly, they occur in association with Cambids and Calcids where parent materials are saline or where gypsum is added with eolian material (Reheis, 1987; Eswaran and Gong, 1991).

The keying sequence follows, for the most part, properties that have the greatest constraints to use and management (Ahrens and Eswaran, 2000). The cold climate of Cryids, for example, presents a limitation that cannot be overcome by reclamation efforts. Similarly, the Salids present special problems because drainage and high-quality water are required before productive agriculture is possible. These 7 suborders are subsequently subdivided into 28 great groups, which in turn are subdivided into more than 269 subgroups (Table 33.47). The great groups are differentiated on the basis of a number of other properties, including aquic conditions, cementation of horizons by gypsum or carbonates, and presence of diagnostic subsurface horizons not identified at the suborder level, generally at depths between 100 and 150 cm. Subgroups are differentiated by a large number of properties. Some common criteria are shallow lithic contacts, presence of nodules and concretions, soil properties that grade toward Andisols or Vertisols, and soil moisture regimes that border on ustic and xeric.
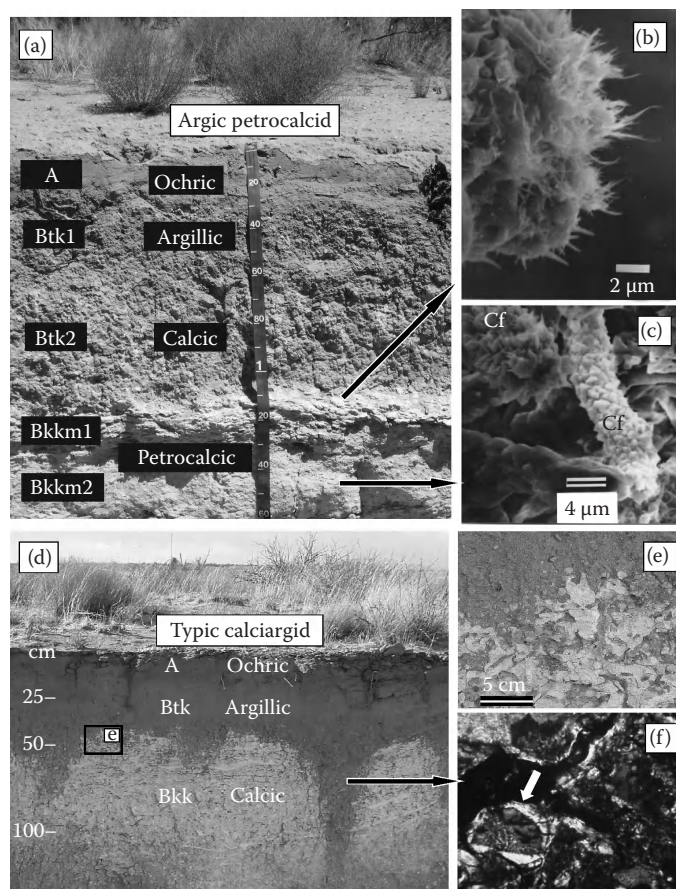
FIGURE 33.67 Morphological and micromorphological features of two Aridisols in southern New Mexico. (a) Illustration of a petrocalcic horizon containing (b) authigenic palygorskite fibers radiating into pore space and (c) calcified fungal hyphae filaments (cf.) suggesting biomineralization of calcite by soil microbes. (d) Illustration of the degradation of a calcic horizon indicated by K-fabric crosscut by noncalcareous material (e) and a vertical pipe containing grain argillans suggesting clay illuviation (f).

33.10.4 Properties

33.10.4.1 Physical Properties

Physical properties of Aridisols can vary dramatically within a local landscape as a consequence of differences in the five soil-forming factors. Particle size distribution, for example, in an area $32 \times 32\,km^2$ ($400\,mi^2$) area in southern New Mexico contains soils with nine particle size classes: sandy-skeletal, loamy-skeletal, clayey-skeletal, sandy, coarse-loamy, fine-loamy, fine-silty, fine, and very fine (Gile et al., 2003). In the same area, bulk density values ranges from a high of $2.19\,g\,cm^{-3}$ in the laminar zone of a petrocalcic horizon to a low of $0.99\,g\,cm^{-3}$ in a gravelly loam B horizon. Infiltration is highly variable as well. When a silty soil surface becomes barren upon the loss of desert grasses and becomes sealed via a physical crust, the amount of water entering the soil is significantly reduced compared with neighboring soils still coved with grasses (Herbel and Gibbens, 1989; Bhark and Small, 2003). Consequently, positive feedback loops develop (Ludwig et al., 2005; Monger and Bestelmeyer, 2006) that exacerbate desertification and create heterogeneous vegetative patterns, such as banded vegetation and localized "islands of fertility" (Schlesinger et al., 1990; Tongway et al., 2001; Rango et al., 2006).

33.10.4.2 Chemical Properties

Organic C and nitrogen contents in Aridisols are among the lowest for all soils. Maximum concentration of organic C in ochric epipedons is often <1% on a mass basis due to low biomass production and rapid decomposition. In Aridisols, as in other soil orders, organic C is inversely related to mean annual soil temperature. Thus, in cool and cold arid regions (frigid and cryic soil temperature regimes), organic C contents rival those of warmer semiarid to subhumid climates (Southard, 2000). Nitrogen occurs mostly in organic matter and is quickly nitrified after the organic matter is decomposed. The NO_3^- produced is susceptible to some leaching, but more importantly, to runoff during overland flow and accumulation in playa environments where denitrification can occur (Schlesinger et al., 1990, 2006).

Soluble salts have accumulated in some, but not all, Aridisols as the result of limited leaching and high evaporation near the soil surface. The presence of salts in Aridisol profiles reflects their high solubility and indicates their relative susceptibility

TABLE 33.47 Suborders, Great Groups, and Subgroups of Aridisols

Suborders	Great Groups	Subgroups
Cryids	Salicryids	Aquic, Typic
	Petrocryids	Xereptic, Duric Xeric, Duric, Petrogypsic, Xeric, Ustic, Typic
	Gypsicryids	Calcic, Vitrixerandic, Vitrandic, Typic
	Argicryids	Lithic, Vertic, Natric, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
	Calcicryids	Lithic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
	Haplocryids	Lithic, Vertic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
Salids	Aquisalids	Gypsic, Calcic, Typic
	Haplosalids	Duric, Petrogypsic, Gypsic, Calcic, Typic
Durids	Natridurids	Vertic, Aquic Natrargidic, Aquic, Natrixeralfic, Natrargidic, Vitrixerandic, Vitrandic, Xeric, Typic
	Argidurids	Vertic, Aquic, Abruptic Xeric, Abruptic, Haploxeralfic, Argidic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
	Haplodurids	Aquicambidic, Aquic, Xereptic, Cambidic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
Gypsids	Petrogypsids	Petrocalcic, Calcic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
	Natrigypsids	Lithic, Vertic, Petronodic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
	Argigypsids	Lithic, Vertic, Calcic, Petronodic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
	Calcigypsids	Lithic, Petronodic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
	Haplogypsids	Lithic, Leptic, Sodic, Petronodic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
Argids	Petroargids	Petrogypsic Ustic, Petrogypsic, Duric Xeric, Duric, Natric, Xeric, Ustic, Typic
	Natrargids	Lithic Xeric, Lithic Ustic, Lithic, Xeretic, Ustertic, Vertic, Aquic, Durinodic Xeric, Durinodic, Petronodic, Glossic Ustic, Haplic Ustic, Haploxeralfic, Haplic, Vitrixerandic, Vitrandic, Xeric, Ustic, Glossic, Typic
	Paleargids	Vertic, Aquic, Arenic Ustic, Arenic, Calcic, Durinodic Xeric, Durinodic, Petronodic, Ustic, Petronodic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
	Gypsiargids	Aquic, Durinodic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
	Calciargids	Lithic, Xerertic, Ustertic, Vertic, Aquic, Arenic Ustic, Arenic, Durinodic Xeric, Durinodic, Petronodic Xeric, Petronodic Ustic, Petronodic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
	Haplargids	Lithic Ruptic-Entic, Lithic Xeric, Lithic Ustic, Lithic, Xerertic, Ustertic, Vertic, Aquic, Arenic Ustic, Arenic, Durinodic Xeric, Durinodic, Petronodic Ustic, Petronodic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
Calcids	Petrocalcids	Aquic, Natric, Xeralfic, Ustalfic, Argic, Calcic Lithic, Calcic, Xeric, Ustic, Typic
	Haplocalcids	Lithic Xeric, Lithic Ustic, Lithic, Vertic, Aquic Durinodic, Aquic, Duric Xeric, Duric, Durinodic Xeric, Durinodic, Petronodic Xeric, Petronodic Ustic, Petronodic, Sodic Xeric, Sodic Ustic, Sodic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
Cambids	Aquicambids	Sodic, Durinodic Xeric, Durinodic, Petronodic, Vitrixerandic, Vitrandic, Fluventic, Xeric, Ustic, Typic
	Petrocambids	Sodic, Vitrixerandic, Vitrandic, Xeric, Ustic, Typic
	Anthracambids	Typic
	Haplocambids	Lithic Xeric, Lithic Ustic, Lithic, Xerertic, Ustertic, Vertic, Durinodic Xeric, Durinodic, Petronodic Xeric, Petronodic Ustic, Petronodic, Sodic Xeric, Sodic Ustic, Sodic, Vitrixerandic, Vitrandic, Xerofluventic, Ustifluventic, Fluventic, Xeric, Ustic, Typic

Source: Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS. U.S. Government Printing Office, Washington, DC.

to downward leaching vs. upward movement of groundwater due to evaporation. Typically, well-drained Aridisols have chromatographic-type accumulations of the most soluble salts (chlorides and sulfates) at greater depths than those that are less soluble (gypsum and calcium carbonate), whereas more poorly drained Aridisol have soluble salts closest to the surface. The distribution of these salts within the soil profile is reflected in the electrical conductivity (EC) (Table 33.48; Figure 33.68).

Exchangeable sodium and pH can be uniquely higher in Aridisols than in other soils. The high accumulation of Na on the cation exchange complex is a common phenomenon in Aridisols (exchangeable Na is related to Na in solution and is reflected in the Na adsorption ratio [SAR]) (Table 33.48). Hydrolysis of exchangeable Na, particularly when electrolyte concentrations

are low (i.e., little or no soluble salts), can produce alkaline soil reactions near pH 10. Swelling and dispersion of soil clays can cause plugging of soil pores and reducing permeability (Levy, 2000). Clay dispersion also enhances the mobility of clay particles in the soil solution and increases the accumulation of clay in subsoil horizons, particularly for natric horizons.

33.10.4.3 Mineralogical Properties

Mineral diversity is high in Aridisols compared with other soil orders for several reasons. First, Aridisols are widely distributed throughout a wide range of parent materials over vast areas of Earth's surface (Figure 33.65; Nettleton and Peterson, 1983; Allen, 1990). Second, in young Aridisols, there has been relatively little weathering of inherited minerals since the arid

TABLE 33.48   Selected Properties[a] of Representative Aridisols

Depth (cm)	Horizon	Sand (%)	Silt (%)	Clay (%)	OC (%)	N (%)	CCE (%)	BD (g cm⁻³)	pH	EC (dS m⁻¹)	SAR
Xeric Calcicryid											
0–10	A	32	56	12	2.98	0.275	15	0.95	7.7	1.7	tr
10–18	Bk1	30	55	15	1.83	0.189	80		8.2	1.0	1
18–33	Bk2	34	44	22	0.78	0.078	41	1.17	7.8	12.9	18
33–58	Bk3	35	41	24	0.37	0.037	73	1.27	8.0	19.0	21
58–84	Bkq1	31	34	35	0.38		54	1.11	7.9	19.6	19
84–107	Bkq2	31	30	39	0.46		88		7.7	30.2	19
107–152	Bkq3	38	32	30	0.23		62		7.8	24.3	18
Typic Haplosalid											
0–11	A	1	50	49	0.69	0.064	4		8.4[b]	22.5	67
11–15	Cyz	1	44	55	0.38	0.039	2		8.7	116.7	281
15–20	Czy	2	55	43	0.60	0.063	3		8.8	122.5	553
20–36	Cz	10	53	37	0.44	0.046	4		8.4	83.8	201
36–56	C1	17	54	29	0.27	0.030	6		8.4	66.7	110
56–76	C2	2	58	40	0.26	0.033	13		8.6	63.0	110
76–105	C3	4	78	18	0.18	0.019	11		8.6	37.7	74
105–150	C4	4	53	43	0.21	0.028	13		8.6	38.3	34
Typic Haplodurid											
0–3	A1	86	12	2	0.11	0.013	3		8.2[b]	0.4	1
3–10	A2	81	14	5	0.08	0.015	5		8.3	0.4	1
10–41	A3	75	14	11	0.15	0.021	7		8.3	0.3	1
41–58	Bqkm1	87	7	6	0.20	0.019	14		8.1	0.4	4
58–76	Bqkm2	81	13	6	0.03		9		8.4	0.4	10
76–94	C	87	9	4	0.01		3		8.5	0.5	15
94–140	Cqkm	84	12	4	0.01		14		8.3	1.0	7
140–175	C′	90	8	2			4		8.4	0.9	6
Typic Calcigypsid											
0–0.5	A1	31	43	26	1.27	0.112	12		7.9	2.5	
0.5–5	A2	28	42	30	0.96	0.094	12		7.9	3.8	3
5–13	A3	29	42	29	0.93	0.088	12	1.37	7.9	4.0	3
13–33	Bk1	25	40	35	0.63	0.066	19		7.8	9.2	9
33–56	Bk2	22	40	38	0.63	0.068	26	1.22	7.4	29.1	8
56–84	Cy1	20	30	50	0.26	0.025	10	1.30	7.6	21.3	
84–115	Cy2	10	50	39	0.14		8		7.7	17.4	2
115–140	Cy3	15	41	44	0.16		14		7.8	16.6	6
140–180	Cy4	14	43	43	0.18		11		7.9	21.3	7
Typic Haplargid											
0–3	A1	58	28	14	0.98	0.095	—	1.49	6.9[b]	0.5	
3–12	A2	60	23	17	0.26	0.038	—	1.50	6.6	0.5	
12–23	BAt	57	22	21			—	1.63	6.9	0.3	
23–46	Bt	46	24	30			—	1.47	7.5	0.2	
46–66	Btk1	40	31	29			3	1.39	7.9	0.2	
66–84	Btk2	33	31	36			8	1.41	8.0	0.2	
84–110	Btk3	31	39	29			15	1.34	8.1	0.3	
110–130	BCtk1	34	40	26			16	1.45	8.1	0.2	
130–160	BCtk2	33	44	23			18	1.60	8.3	1.0	7
160–180	BCtk3	27	49	24			12	1.42	8.1	2.7	11
180–230	C	32	49	19			9	1.37	7.9	6.8	10

(*continued*)

TABLE 33.48 (continued)   Selected Properties[a] of Representative Aridisols

Depth (cm)	Horizon	Sand (%)	Silt (%)	Clay (%)	OC (%)	N (%)	CCE (%)	BD (g cm^{-3})	pH	EC (dS m^{-1})	SAR
Argic Petrocalcid											
0–5	A	85	5	10	0.25		—				
5–18	BAt1	87	4	9	0.13		—	1.86^c			
18–25	BAt2	82	5	13	0.17		—	1.78			
25–36	Bt	80	5	15	0.14		1	1.70			
36–48	Btk	77	6	17	0.24		16	1.53			
48–74	Bkm1	68	9	23	0.15		75				
74–100	Bkm2	71	11	18	0.05		65	1.68			
100–150	Bkm3	75	6	19	0.05		51	1.66			
Typic Haplocambid											
0–6	A	47	43	10	0.52		tr	1.41	8.0		
6–15	Bw1	44	45	12	0.31		tr	1.26	8.3		
15–33	Bw2	46	43	11	0.27		tr	1.22	8.5		
33–48	Bk1	44	44	12	0.35		4	1.22	8.7	0.7	5
48–60	Bk2	65	29	6	0.28		tr	1.24	8.7	1.7	10
60–87	Bk3	49	44	7	0.22		tr	1.17	8.2	7.8	15
87–135	C1	39	52	9	0.22		tr	1.24	7.9	13.3	14
135–150	C2	66	25	9	0.12		tr	1.30	7.8	11.8	12
150–175	C3	62	29	9	0.12		tr	1.30	7.7	11.1	11
175–205	C4	64	22	14	0.10		tr	1.30	7.9	10.6	9

Source: Data from Soil Survey Staff. 1975. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. Agriculture handbook no. 436. USDA-SCS. U.S. Government Printing Office, Washington, DC. USDA-NRCS, 1997c.

[a] OC, organic carbon; N, nitrogen; CCE, calcium carbonate equivalent; BD, bulk density at 33 kPa water potential; EC, electrical conductivity; and SAR, sodium adsorption ratio.

[b] pH of 1:1 soil:water mixture, unless noted by "b," then pH of saturated paste.

[c] Bulk density at 33 kPa water potential, unless noted by "c," then air-dry.

climate provides little surplus water for mineral hydrolysis or for leaching of weathering products. Third, the dry climate enables soluble minerals to accumulate that would otherwise be flushed from soils of more humid climates. Fourth, older Aridisols span multiple glacial–interglacial cycles of the Pleistocene; thus, the mineralogy of wetter paleoclimates is superimposed on the mineralogy of arid recent climates, such as the kaolinitic soils in regions of West Africa and Australia.

Consequently, family mineralogic classes for Aridisols in *Soil Taxonomy* include the following categories: Mixed, Smectitic, Carbonatic, Gypsic, Siliceous, Glassy, Illitic, and Kaolinitic (Southard, 2000). Mineralogy of the sand and silt fractions generally reflects relatively little weathering with a silicate mineral suite that is dominated by quartz, feldspar, amphiboles, and micas, with a few Aridisols containing abundant inherited zeolites (Allen, 1990; Ming and Boettinger, 2001). Mineralogy of the clay fraction is usually dominated by smectites (usually montmorillonite), as reflected by the family mineralogical classes of *Soil Taxonomy* (Southard, 2000). However, even within a small region, the suite of clay minerals can be diverse. Analysis of 154 horizons from 72 pedons revealed the presence of illite, kaolinite, smectite, palygorskite, sepiolite, chlorite, interstratified mica-montmorillonite, and vermiculite (Monger and Lynn, 1996). In these Aridisols, illite and kaolinite were found to be ubiquitous and their concentrations changed little with depth or age, unlike smectite whose concentration formed a bulge in

B horizons and became more abundant with age. Palygorskite and sepiolite occurred in well-developed petrocalcic horizons, whereas chlorite, interstratified mica-montmorillonite, and vermiculite were more common in soils formed in igneous parent material than in soils derived from limestone.

Mineralogy of Aridisols may have significant contents of relatively soluble, nonsilicate minerals, including calcite ($CaCO_3$), gypsum ($CaSO_4 \cdot 2H_2O$), halite ($NaCl$), thenardite (Na_2SO_4), mirabilite ($Na_2SO_4 \cdot 10H_2O$), epsomite ($MgSO_4 \cdot 7H_2O$), nahcolite ($NaHCO_3$), and trona ($Na_2CO_3 \cdot 10H_2O$) (Doner and Lynn, 1989; Buck et al., 2006).

33.10.5 Pedogenesis of Aridisols

The most distinguishing characteristic about soil formation in Aridisols is the limited water available for plant growth, mineral weathering, and leaching of soluble weathering products. A typical water budget is dominated by a long period of deficit during the summer when potential evapotranspiration greatly exceeds precipitation followed by partial recharge of water-holding capacity, then rapid utilization of that stored water during the brief growing season. The lack of a surplus of precipitation over evapotranspiration and water-holding capacity prevent wholesale leaching of weathering products from Aridisols. Nonetheless, many Aridisols have well-developed soil profiles due to (1) the concentration of the limited
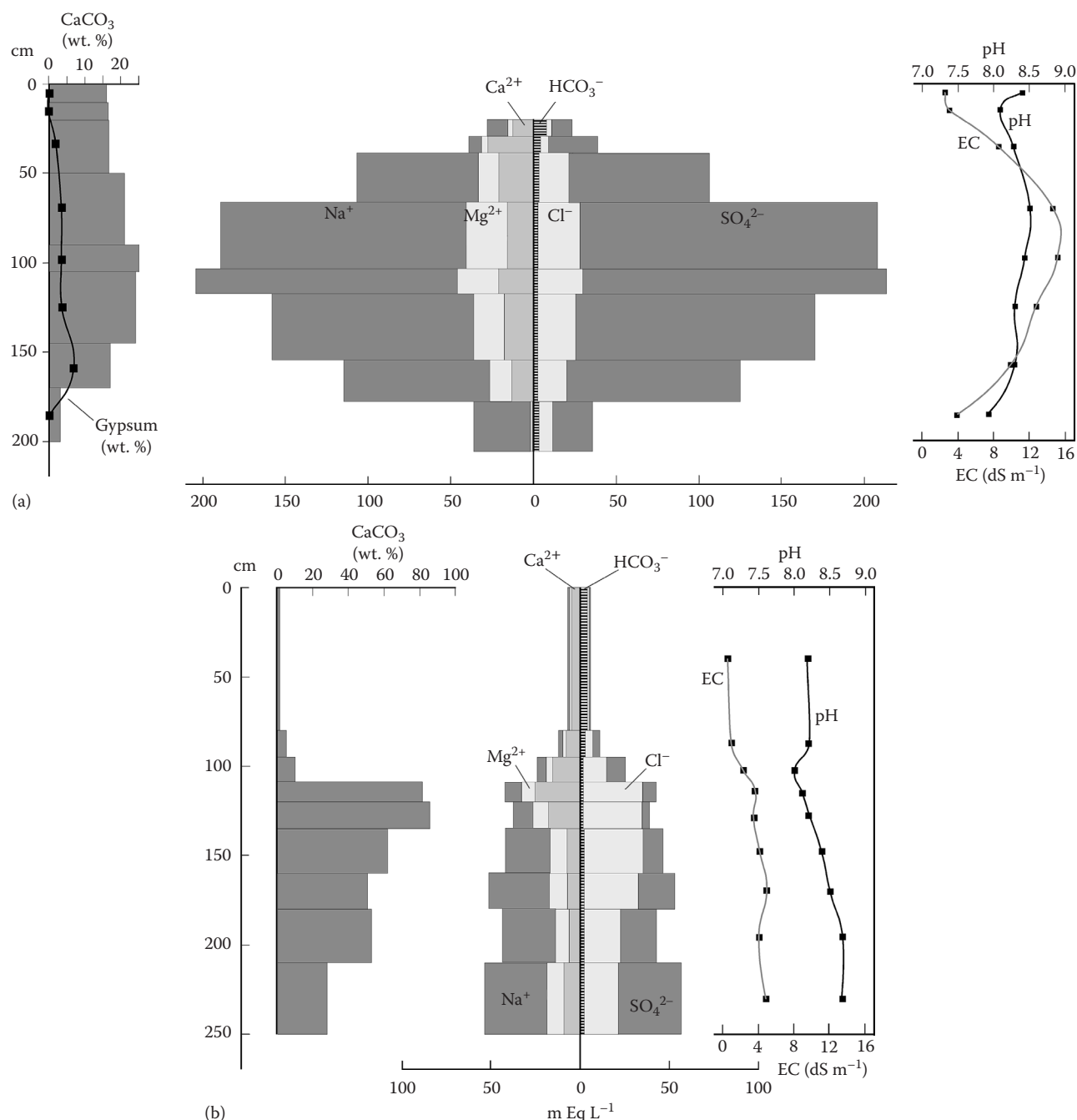
FIGURE 33.68 Salinity profiles, pH, EC, $CaCO_3$, and gypsum content illustrating chemical properties of two Aridisols in southern New Mexico. (a) Vertic Haplocambid formed in the clayey sediments of a playa with high soluble salt concentrations and (b) a neighboring Argic Petrocalcid formed sandy alluvium on a basin floor.

water in a relatively small volume of the upper soil profile, (2) occasional periods of relatively high precipitation in the winter months when deeper leaching could occur, and (3) wetter paleoclimates. The depth of leaching and horizon development may reflect the rainfall of the extreme years, rather than that of average years (Archibold, 1995).

Working in concert with the effects of soil moisture is the effects of soil temperature. Soil temperature regimes of Aridisols range from cryic to isohyperthermic (Soil Survey Staff, 1999).

Although the rate of chemical weathering increases as temperature increases (Buol et al., 2002), Aridisols in cooler temperatures retain water longer and dissolve more CO_2 (Breecker et al., 2009). Most Aridisols develop in sediments transported by wind or water on relatively low gradient landscapes. Lithology of the sediments varies widely. Truncation of steep landforms and slow weathering rates retard the formation of subsurface soil horizons and result in Entisols on most erosional landforms (Nordt et al., Section 33.4). Some Aridisols have formed in residual rock

material on erosional landforms, for example, on old and highly weathered granitic pediments in the Mojave Desert (Boettinger and Southard, 1991, 1995) and on stable, steep hills where episodic erosion has been minimal (Nettleton and Peterson, 1983).

The four generalized soil-forming processes of additions, transformations, translocations, and losses (Simonson, 1959) are operating in Aridisols, as in more humid soils, but at different rates. Of particular importance, as discussed in the sections below, are eolian accretions and rain chemistry (additions), neoformation of silicate clay, silica, carbonate, and gypsum (transformation), and illuviation of silicate clay (translocations).

33.10.5.1 Atmospheric Additions

Because Aridisols that are of the "nonflushing type" (Rode, 1962), their profiles are cumulative repositories of airborne particles and rain-borne cations and anions. These atmospheric additions have major impacts on the physical and chemical properties and provide a historical record of paleoenvironmental changes (e.g., Chadwick and Davis, 1990; Reynolds et al., 2001; Graham et al., 2008).

33.10.5.1.1 Eolian Accretions

Eolian accretions in many Aridisols are well documented (Yaalon and Ganor, 1973; Gile and Grossman, 1979; Reheis, 1987; Reheis et al., 1995; Valentine and Harrington, 2006). The role of eolian dust is particularly important in many Holocene Aridisols, where much of the carbonate, gypsum, and opaline silica are derived from an external source (Harden et al., 1991). In some cases, eolian material constitutes nearly the entire fine-earth fraction and accounts for the nearly stone-free vesicular horizons below many desert pavements (McFadden et al., 1987). Eolian accretion in Aridisols is enhanced if they have a surface structure that can trap dust. For example, landscapes derived from rock types that tend to weather to soils with few surface clasts are less likely to have significant eolian accretion because an effective dust trap is absent (Boettinger and Southard, 1991).

Eolian accretion of carbonate dust is often required for the development of thick petrocalcic horizons in soils formed in noncalcareous parent materials given the limited amount of water necessary for silicate mineral weathering (Gile et al., 1966; Birkeland, 1984). Ruhe (1967), for example, calculated that there was insufficient calcium in rhyolite alluvium to account for the large amount of calcium contained in the calcic and petrocalcic horizons formed in rhyolitic alluvium. Moreover, even if the rhyolitic parent material had contained enough calcium, the rocks were only slightly weathered and therefore seemed probable that the source of calcium was atmospheric, especially since depth to groundwater was over 100 m. Eolian accretion of carbonate dust, however, is not prerequisite to the formation of all petrocalcic and calcic horizons. Aridisols formed on limestone bedrock or in limestone alluvium may have some accumulation of calcareous eolian dust, but its impact in terms of carbonate source may be of lesser importance in these soils (Rabenhorst and Wilding, 1986).

33.10.5.1.2 Rainwater Chemistry

Rainwater typically contains various amounts of Ca^{2+}, Mg^{2+}, K^+, Na^+, SO_4^{2-}, and NO_3^-. An important consequence of this impure rain is that it can be an important source of nutrients for arid ecosystems (Schlesinger, 1997). Another important consequence is that calcium in rain, like calcium in dust, is a potential source of calcium for the formation of petrocalcic horizons. For example, a 10-year analysis of dust trap data revealed that the calcareous dust, combined with water-soluble calcium extracted from the dust, could generate 0.4 g of soil $CaCO_3$ per meter squared per year (Gile and Grossman, 1979). Rain, however, could generate 1.5 g m^{-2} $year^{-1}$, assuming 200 mm of annual rainfall and a Ca^{2+} concentration of 3 mg L^{-1} (Gile et al., 1981).

33.10.5.2 Surface Features

Pedogenic processes in Aridisols produce unique surface features not found in soils of more humid regions (Cooke et al., 1993). Of special interest are (1) desert pavement, (2) desert varnish, (3) vesicular horizons, and (4) biological soil crusts. In each case, the open surface exposure to sun, wind, rain, temperature fluctuations, and lack of vascular plant roots are involved with their formation.

33.10.5.2.1 Desert Pavement and Varnish

Desert pavement, a surficial layer of rock fragments, usually one or two fragments thick, occurs on many Aridisols (Figure 33.69a and b). The origin of desert pavement has been attributed to loss of fine particles by wind deflation, water erosion, splitting of stones, vertical sorting of coarse fragments by repeated wetting and drying, and by rafting of surface fragments by infiltrating eolian material (Springer, 1958; Cooke, 1970; McFadden et al., 1987). In the case of rafting of surface fragments (Figure 33.69b), the pavement itself serves as a dust trap (Yaalon and Ganor, 1973; Gile, 1975a; Peterson, 1977). The clasts shown in Figure 33.69a, for example, act as a dust trap to catch eolian materials derived from nearby playa and alluvial fans. The effectiveness of desert pavement decreases as surface fragments become more interlocked. When the land surface is nearly completely covered by pavement, the surface neither retains much new eolian material nor loses much material to erosion, unless disturbed.

Desert varnish is a yellowish- or reddish-brown to nearly black coating on desert pavement rock fragments. The varnish is commonly black on pebble surfaces aboveground, but reddish-brown, red, or yellowish red belowground in soils with argillic horizons (Gile and Grossman, 1979). Its composition includes 2:1 silicate clay minerals and oxyhydroxides of Fe and Mn. These constituents are probably of eolian origin and appear to be microbially precipitated (Dorn and Oberlander, 1981; Broecker and Liu, 2001). The varnish occurs most frequently on stable, mafic and various other rock fragments and is generally thickest (as much as 100 μm thick) and darkest on Pleistocene pavements (Nettleton and Peterson, 1983; Garvie et al., 2008). Desert varnish can form in little more than a decade and may serve as an armor that protects microbial weathering fronts from harsh
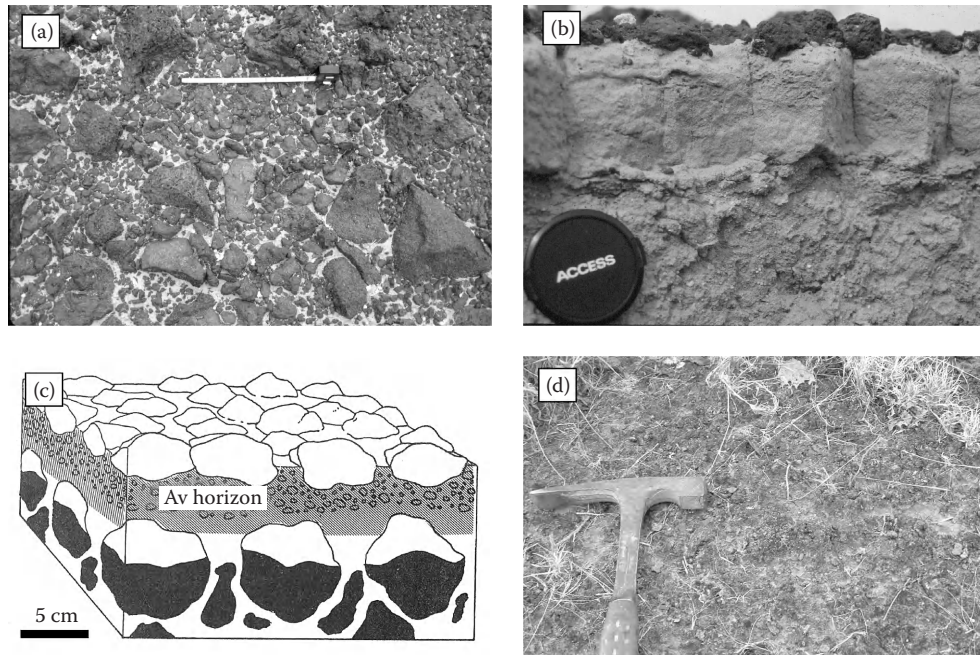
FIGURE 33.69 (a) Desert varnished on clasts interlocking to form a nearly continuous desert pavement on a Pleistocene basalt flows in the Mojave Desert, California. Tape length is about 30 cm. (b) Vesicular surface horizon (an Av horizon) rafting atop desert-varnished basalt pebbles of a desert pavement in the Cima volcanic field in the Mojave Desert. Lens cap is about 60 mm. (c) Diagram of desert pavement underlain by a thin, vesicular A horizon. (From Gile, L.H., and R.B. Grossman. 1979. The desert project soil monograph. USDA-SCS, Washington, DC.) (d) Biological soil crust covering an otherwise bare inter-vegetative surface of a silty soil in the Chihuahuan Desert, New Mexico.

desert environments of irradiation, extreme temperature fluctuations, and low moisture conditions (Krumbein and Jens, 1981). Comparison of desert varnish on exposed stones of adjacent geomorphic surfaces, the same age shows desert varnish to be absent on limestone (Gile and Grossman, 1979), unless the limestone contains exposed deposits of chert or silicified veins.

33.10.5.2.2 *Vesicular A Horizons*

Beneath the pavement, especially in silty eolian material, a horizon may form with many fine, discontinuous vesicular pores. These horizons are widespread feature of Aridisols and are generally referred to as an Av horizon ("v" in this context is not to be confused with "v" used to describe horizons with plinthite) (Figure 33.69c). The vesicles result from repeated saturation and desiccation and progressive enlargement of air bubbles trapped beneath the thin, saturated surface layer (Springer, 1958; Miller, 1971). Av horizons coevolve with the formation of desert pavement (Anderson et al., 2002). The rate at which vesicular A horizons form can be quite rapid based on the observation that well-developed vesicular horizon formed in tire tracks known to be less than 2 years old (Gile and Grossman, 1979). The discontinuous porosity of Av horizons, in addition to their platy structure, leads to low infiltration rates and shallow leaching that affect ecologic function (Turk and Graham, 2009).

33.10.5.2.3 *Biological Soil Crusts*

A continuum of crust types occupies surfaces of many Aridisols depending on climate, soil properties, and disturbance history

(Johnston, 1997; Belnap and Lange, 2001). On one end is the abiotic physical crust; on the other end are mosses. Soil biological crusts (a.k.a., microbiotic or cryptogamic crusts) include the cyanobacterial crusts, green algae crusts, lichen crusts, and mosses. Abiotic physical crusts are widespread, but not present everywhere, and appear as a surficial layer of uncemented, coherent fine earth that can be broken free from underlying soil. The crust is generally less than 10–20 mm thick and is often platy in the upper part and massive below. Water infiltration into physical crust is slow, in contrast to the rapid infiltration into uncrusted coppice sand dunes. Physical crusts form by raindrop impacts, localized sheet erosion, and by repeated wetting and drying of loamy soil material (Nettleton and Peterson, 1983).

Biological soil crusts occupy many habitats. They occur on top of the soil surface, within soil, beneath translucent stones, within rock fissures and cracks, and within the airspaces of porous rocks. Functions of biological soil crusts include stabilizing soil surfaces (Belnap and Gillette, 1998), increasing mineral weathering, fixing nitrogen (Malam Issa et al., 2001; Belnap, 2002), modifying evaporation (Verecchia et al., 1995), decreasing seed germination until periodically disturbed (Prasse and Bornkamm, 2000), influencing populations of other soil microorganisms, such as nematodes (Zhi et al., 2009), and affecting surface hydrology (Belnap, 2006). Biological soil crusts can decrease as well as increase runoff depending on the crust type and soil setting. In some sandy soils, cyanobacterial crusts increases runoff as the result of hydrophobic properties and pore plugging, in contrast to silty soils where the roughness

of lichen crusts slows overland flow and increases infiltration (Kidron et al., 2003; Kidron, 2007).

33.10.5.3 Weathering and Silicate Clay Formation

Physical weathering in Aridisols in the form of disintegration of rock results from many processes, most prominent of which are thermal expansion and contraction, ice crystal growth, root growth in cracks, and salt crystal growth (Birkeland, 1999). Chemical weathering, despite being hampered by long periods of dryness, can be substantial depending on soil age. In Holocene Aridisols of North America chemical weathering is minimal due to the aridity of the Holocene climate (Antevs, 1955). In other regions, such as South Africa, mid-Holocene climates during the Altithermal were wetter and warmer (Linacre and Geerts, 1997). But even in North America, ferrugination (reddening) in arid Holocene soils has been sufficient to create cambic horizons, thus converting Torriorthents into Haplocambids (Gile et al., 1995). Aridisols on Pleistocene landscapes in North America show clear signs of weathering, such as the transformation of primary silicate minerals to Ca- and Mg-rich clays (Harden et al., 1991), interstratified mica-vermiculite (Eghbal and Southard, 1993a), vermiculite and Al-rich smectite (Boettinger and Southard, 1995), and Ca and silica in calcareous duripans (Boettinger and Southard, 1991; Eghbal and Southard, 1993c). In addition, the deeply weathered granitic pediments of the Mojave Desert, alteration of biotite to vermiculite are extensive. These alteration and neosynthesis of Al-rich smectites from feldspar weathering are the main sources of phyllosilicates in the clay fraction (Boettinger and Southard, 1995). Similar evidence of silicate clay neoformation during more humid conditions of the Pleistocene was found in the granitic soils in the Chihuahuan Desert (Monger and Lynn, 1996), as discussed in Section 33.10.4.3.

The pedogenic formation of the fibrous clays, palygorskite [ideal structure: $Si_8Mg_5O_{20} \cdot (OH)_2(OH_2^-)_4 \cdot 4H_2O$], and sepiolite [ideal structure: $Si_{12}Mg_8O_{30}(OH)_4(OH_2)_4 \cdot 8H_2O$] (Singer, 1989) has been documented in the alkaline, Si-preserving environment of petrocalcic horizons. Scanning electron microscopy of palygorskite fibers radiating into pore spaces has been used as evidence of in situ formation, rather than illuviation of palygorskite (Figure 33.67b; Monger and Daugherty, 1991a).

33.10.5.4 Pedogenic Silica

Pedogenic silica in Aridisols accumulates as opal ($SiO_2 \cdot nH_2O$) (Flach et al., 1969; Chadwick et al., 1987a), opal-CT (Monger and Kelly (2002)), and possibly as chalcedony and microcrystalline quartz (Flach et al., 1973; Smale, 1973; Milnes et al., 1991). Incipient cementation produces horizons that are very hard when dry and brittle when moist. More complete cementation of the soil by opal produces a duripan. Pedogenic silica, as with pedogenic carbonate, has stages whereby progressively more complex silica accumulations occur on progressively older geomorphic surfaces (Taylor, 1986). But unlike pedogenic carbonate, which cements by engulfing soil particles like mortar between bricks, pedogenic silica cements by bonding soil particles at contacts like glue (Chadwick and Nettleton, 1990). Relatively small amounts of silica (about 10% by weight) cause significant cementation of clay particles (Nettleton and Peterson, 1983), which can significantly change water retention characteristics and CEC. Cementation by silica tends to be less disruptive of soil fabric than cementation by carbonate.

Climate is an important factor for the accumulation of pedogenic silica. It must provide enough water for weathering, but not enough for intensive leaching (Flach et al., 1973). This condition is met in many arid and semiarid climates. The second condition is a readily available source of $H_4SiO_4°$. This condition is met in many areas where easily weatherable volcanic ash and ultramafic materials rapidly release $H_4SiO_4°$ (Alexander et al., 1994; Alexander, 1995).

Secondary silica in the form of opaline veins and stringers is a frequent feature in many highly indurated petrocalcic horizons (Franks and Swineford, 1959). The source of this silica may be derived by weathering of siliceous rocks (Boettinger and Southard, 1991), volcanic ash (Chadwick et al., 1987b), or eolian dust (Harden et al., 1991). It may also originate from the replacement of quartz and feldspar grains by pedogenic calcite. As revealed by thin section petrography, the original shapes of some silicate grains are gone, occupied instead by calcite—hence, the term "replaced" (Swineford et al., 1958; Reheis, 1988; Monger and Daugherty, 1991b). Replacement of silicate grains has been attributed to *force-of-crystallization* (Maliva and Siever, 1988a, 1988b). This process, like pressure solution in metamorphic rocks, is based on the principle that minerals are more soluble at contact points when under stress. The force-of-crystallization hypothesis in petrocalcic horizons is postulated to occur as calcite progressively crystallizes and engulfs detrital grains. It accounts for why some surface pits in silicate grains match the shapes of calcite crystals impacted against them, and it eliminates the need to have unverifiable high pH to account for silica dissolution (Maliva and Siever, 1988b). Silica released by force-of-crystallization would not only precipitate as opaline phases, but would also be a source of Si for authigenic palygorskite and sepiolite, which are common clay minerals in petrocalcic horizons (Vanden Heuvel, 1966; Monger and Daugherty, 1991a; Reheis et al., 1992).

33.10.5.4.1 Duripans and Silcretes

Most duripans in Aridisols have a considerable content of $CaCO_3$, as much as 70% by weight (Southard et al., 1990), but at least half the volume of a duripan does not slake in acid, as petrocalcic horizons do. Many of these duripans have indurated laminar layers as much as 2 cm thick that overlie massive and less strongly cemented material that is highly calcareous. These duripans can be distinguished from petrocalcic horizons only by acid treatment. In many cases, only the upper laminar layer survives the acid treatment intact.

Duripans in Aridisols commonly occur beneath or within an argillic horizon (Flach et al., 1973). In soils with both an argillic horizon and underlying calcic horizon, duripans often straddle the two horizons, superimposed on the lower portion of the argillic and the upper portion of the calcic, thus, forming

a chromatographic column of illuviated constituents (Chadwick et al., 1987a). Silica in duripans is typically some form of opal or chalcedony that is finely distributed throughout the matrix, or as coatings, diffuse isotropic flocs, nodules, glaebules, and stringers (Flach et al., 1973; Boettinger and Southard, 1990; Monger and Adams, 1996). High pH promotes duripan formation in granitic terrain and, in some cases, precludes the requirement for volcanic ash as a source of silica (Boettinger and Southard, 1990).

Silcretes or duricrust in Aridisols is more indurated and generally older than duripans (Jackson, 1957). Their occurrence in the stratigraphic record may stretch back to the Precambrian (Summerfield, 1983). Silcretes develop when sediments become silicified in areas (1) where drainage of silica-rich groundwater is poor and has a fluctuating water table, (2) where silica-rich solutions evaporate, and (3) where upward moving silica-rich solutions meet downward percolating less alkaline or more salty waters (Smale, 1973; Summerfield, 1983; Milnes et al., 1991). Mineralogically, silcretes are a combination of opal-A, opal-CT, and microcrystalline quartz (microquartz) depending on the concentration of $H_4SiO_4°$ (Milnes et al., 1991). Like duripans, high $H_4SiO_4°$ concentrations ($>60\,mg\,L^{-1}$) favor opal-A; intermediate concentrations ($20–60\,mg\,L^{-1}$) favor opal-CT; and low $H_4SiO_4°$ concentrations ($<20\,mg\,L^{-1}$) favor microquartz (Chadwick et al., 1987b).

33.10.5.4.2 Biogenic Silica

Silica cycling by plants, especially grasses, is an important component of Aridisols. Desert plants, especially grasses, use silica for structural stability, mineral nutrition, and antiherbivore defense mechanisms (Iler, 1979; Massey et al., 2006). When silica-containing plants die and decompose, they release silica into the soil as phytoliths. Phytoliths are ubiquitous constituent of not only Aridisols, but most soils (Wilding et al., 1977; Meunier et al., 1999), ranging from 8 to $300\,kg\,ha^{-1}$ depending on biomass production and mean annual precipitation (Pease, 1967; Kelly, 1989). Except for certain silica-rich geological environments, such as hot springs, most soil opal has a biogenic origin that exerts a major control on silica in the soil solution (Alexandre et al., 1997). The marine record indicates that biogenic silica underwent a spectacular expansion during the middle Cenozoic when grasses proliferated across terrestrial environments (Lowenstam and Weiner, 1989). Vertically in soils, phytoliths are most abundant in surface horizons and decreases with depth, unless buried paleosols are encountered. In paleosols, phytolith morphology and isotopic composition can be a useful method for providing information about ecological and archaeological changes (Wilding and Drees, 1974; Kelly et al., 1991; Rosen and Weiner, 1994; Schaetzl and Anderson, 2005).

33.10.5.5 Pedogenic Carbonate

Several names have been given to pedogenic carbonate: caliche, calcrete, croute calcaire, tosca, caprock, crust, calcic horizons, and petrocalcic horizons (Gile, 1961; Goudie, 1973; Dregne, 1976; Soil Survey Staff, 1999). Although some pedgenic carbonate has been reported to be dolomite (e.g., Capo et al., 2000), the vast majority is calcite (e.g., Doner and Lynn, 1989; Kraimer et al.,

2005). Pedogenic carbonate is a subset of *soil carbonate*, which is the total amount of carbonate measured in the soil, typically by acid dissolution methods, and reported as calcium carbonate equivalent (Soil Survey Staff, 1996). Soil carbonate is a broad category containing *primary carbonate* (if present) and *pedogenic carbonate* (if present). Primary carbonate (also termed geogenic, lithogenic, inherited, or nonpedogenic) is $CaCO_3$ originally precipitated in aquatic environments as limestone. Pedogenic carbonate (also termed secondary or authigenic carbonate) is $CaCO_3$ developed in soil and can occur as either authigenic (in situ) $CaCO_3$ or allogenic (ex situ) $CaCO_3$, where allogenic $CaCO_3$ is derived from preexisting soils upslope or downwind. The age of pedogenic carbonate spans vast time periods, ranging from carbonates in Paleozoic paleosols (Mora et al., 1996) to carbonate formed on artifacts (Pustovoytov, 2002) and living plant roots (Monger et al., 2009).

Of the global amount of soil carbonate, Aridisols contain the largest proportion (48%), followed by Entisols (28%), and Mollisols (12%) (Eswaran et al., 2000). At the suborder level, however, Orthents contain the most (21.4%) because many Orthents are formed by the truncation of Aridisols (see Nordt et al., Section 33.4). Other suborders containing notable amounts of soil carbonates are Calcids (17.4%), Argids (14.2%), Cambids (10.0%), Ustolls (9.9%), Psamments (3.5%), Ustalfs (3.1%), Fluvents (2.8%), Gypsids (2.4%), and Salids (2.2%) (Eswaran et al., 2000).

Calcium carbonate is present in almost all Aridisols, except for some Aridisols on stable platforms in West Africa and Australia (Southard, 2000). Inputs of Ca from mineral weathering or from the atmosphere, either dissolved in rainwater or as particles of Ca-bearing minerals, coupled with relatively low solubility, cause $CaCO_3$ to accumulate. Once in soil solution, Ca can react with HCO_3^- from rainwater or produced via soil CO_2 (from respiration) with water. Loss of water by evaporation, or reduction of the partial pressure of CO_2 (P_{CO_2}), drives the precipitation of $CaCO_3$ from solution.

Aridisols on Holocene or latest Pleistocene landscapes in the southwestern United States, for example, are often calcareous in all horizons because the relatively dry conditions of the Holocene prevented leaching of carbonate from surface horizons (e.g., Typic Haplodurid and Typic Calcigypsid in Table 33.48). In contrast, Aridisols on older Pleistocene landscapes are often carbonate free in the upper horizons, presumably due to more effective leaching during the cooler, moister Pleistocene (e.g., Typic Haplargid and Argic Petrocalcid in Table 33.48). Aridisols of West Africa are acidic and lack carbonates because they form on old preweathered platform surfaces or from more recent polycycled sediments derived therefrom.

Soil taxonomy systems have used pedogenic carbonate as an important feature to classify soils (e.g., Pedocal vs. Pedalfer, Marbut, 1928). Pedogenic carbonate has also provided important clues about paleoclimate, paleoecology, paleoatmospheric composition, relative ages of geomorphic surfaces, landscape evolution, carbon sequestration, and vegetation patterns in arid landscapes (e.g., Mora et al., 1996; Ekart et al., 1999; Monger and Martinez-Rios, 2001; Duniway et al., 2007).

33.10.5.5.1 *Pedogenic Carbonate Forms*

Forms of pedogenic carbonate accumulation have been described as filamentary, concretionary, cylindroidal, nodular, platy, blocky, bedded, massive, veined, and flaky, as well as root casts, bands, joint fillings, coatings, pendants, beds, plugged horizons, laminar horizons, laminae, pisoliths, and ooliths (Gile, 1961; Harden et al., 1991; Eghbal and Southard, 1993b; Soil Survey Division Staff, 1993). These carbonate forms range from nonindurated, which slake when placed in water, to very strongly indurated, which do not slake in water and cannot be scored with a knife. On the low end of carbonate accumulation, horizons with carbonate filaments can contain as little as 1% $CaCO_3$ and, depending on texture, have typical bulk densities (e.g., 1.68 g cm^{-3}) and high infiltration rates (e.g., 12.4 cm h^{-1}, Gile, 1961). On the high end of carbonate accumulation, laminar horizons can contain over 90% $CaCO_3$ with high bulk densities (e.g., 2.22 g cm^{-3}) and low infiltration rates (e.g., 0.1 cm h^{-1}).

Pedogenic carbonate is an important indicator of soil age because progressively older geomorphic surfaces contain progressively greater amounts of carbonate. It was observed that with time, four morphogenetic stages of carbonate develop, progressing from having no visible carbonate to having stage I filaments, then stage II nodules, a stage III plugged horizon, and eventually a stage IV laminar horizon developed atop the plugged horizon (Gile et al., 1966; Figure 33.70). More recently, two additional stages of carbonate accumulation have been added: stages V and VI (Machette, 1985). Stage V contains thick laminae (>1 cm) and pisolites, and stage VI contains multiple generations of laminae, breccia, and pisolites. The time required to reach a certain morphogenetic stage depends on soil texture. Gravelly soils pass through the stages more quickly than fine-textured soils because gravelly soils have lower surface area and less pore space. Associated with an increase in pedogenic carbonate is a large expansion of the volume of the calcic and petrocalcic horizons as carbonate precipitates and forces matrix grains apart. Cementation of horizons occurs when carbonate content reaches 25%–60% (Machette, 1985).

Until about 1965, horizons with carbonate accumulation were designated as C horizons with a "ca" suffix (or "cam" suffix if indurated) following the procedure of the *Soil Survey Manual* (Soil Survey Division Staff, 1951). Subsequent work showed that carbonate accumulation was a pedogenic process (Gile et al., 1965). Therefore, horizons with significant amounts of pedogenic carbonate were afterward designated B horizons,

Schematic diagram of diagnostic carbonate morphology for the stages of carbonate accumulation in the two morphogenetic sequences (left below). Stages are described (right below), including stages V and VI not shown in diagram

Carbonate morphogenetic stages   Definitions of carbonate morphogenetic stages

	Diagnostic carbonate morphology	
Stage	Gravelly sequence[a]	Nongravelly sequence
I	Thin pebble coatings, continuous in last part of stage.	Few[b] filaments, masses, or faint coatings
II	Continuous pebble coatings, some interpebble masses.	Few to common nodules or masses.
III	Many interpebble masses; plugged horizon[c] develops in last part of stage.	Many nodules and internodular masses; plugged horizon develops in last part of stage.
IV	Laminar horizon overlying plugged horizon.	Laminar horizon overlying plugged horizon.
V	Very strongly cemented to indurated with >1 cm thick laminae; may contain pisoliths and have vertical faces and fractures coated with laminated carbonate.	
VI	Indurated with multiple generations of recemented laminae, breccia, and pisoliths.	

[a] Morphologies are best expressed where "nongravelly" materials contain less than about 20% by volume of rock fragments (fragments 2 mm or larger in diameter), and "gravelly" materials contain more than about 60% by volume of rock fragments. Materials that have between 20% and 60% by volume of rock fragments have intermediate morphologies.

[b] Few <2, common 2 to <20, many = 20 or greater percent of area covered.

[c] Plugged horizon contains 50% or more pedogenic carbonate (by vol).

FIGURE 33.70 Illustration and definitions of the morphogenetic stages of pedogenic carbonate accumulation in gravelly soils and soils without gravels. Stages V and VI are not shown in diagram. (From Gile, L.H., F.F. Peterson, and R.B. Grossman. 1966. Morphological and genetic sequences of carbonate accumulation in desert soils. Soil Sci. 101:347–360; Machette, M.N. 1985. Calcic soils of the southwestern United States, p. 1–21. *In* D.L. Weide (ed.) Soils and Quaternary geology of the southwestern United States. Geological Society of America. Special Paper 203.)

not C horizons. The prominence of carbonate horizons in arid and semiarid regions led to the proposed K horizon as a master horizon, such as the O, A, B, C, and R horizons at that time. The K horizon was based on the presence of K-fabric, which is defined as "fine-grained authigenic carbonate that coats or engulfs skeletal pebbles, sand, and silt grains as an essentially continuous medium" (Gile et al., 1965). The K horizon, by definition, contained 90% or more K-fabric. The K horizon was not adopted by the National Cooperative Soil Survey, except that the suffix "k" eventually replaced the "ca" suffix. In other earth science disciplines, however, the K horizon has been used (e.g., Machette, 1985; Birkeland, 1999). Recently, the "kk" suffix, which corresponds to the stage III plugged horizon or higher of the carbonate morphogenetic stages, has been formally adopted by the National Cooperative Soil Survey (Soil Survey Staff, 2010).

33.10.5.5.2 Origin of Pedogenic Carbonate

Models of pedogenic carbonate formation can broadly be classified into four groups (Monger and Wilding, 2002): (1) the *per ascensum*-groundwater model in which carbonate develops from upward rising groundwater via capillarity, (2) the *per descensum*-leaching model in which pedogenic carbonate develops from downward redistribution of calcareous parent material, (3) the *per descensum*-dust model in which pedogenic carbonate develops from atmospheric additions of dust and calcium in rainwater, and (4) the in situ-biogeochemical model in which carbonates develops from biological weathering, Ca upward translocation by hyphae and roots, and precipitation of carbonates within the soil profile via both inorganic and biogenic precipitation. The formation of pedogenic carbonate in any particular soil may involve several of these models working together at different magnitudes.

The *per ascensum*-groundwater model accounts for pedogenic carbonate formation as the result of precipitation of carbonate minerals from groundwater containing high concentrations of Ca^{2+} and HCO_3^- drawn up from shallow water tables by capillarity (Sobecki and Wilding, 1983). This model includes groundwater calcretes. These differ from pedogenic calcretes, that is petrocalcic horizons, in two ways. The groundwater calcretes have coarse (often poikilotopic) interlocking carbonate crystals that engulf detrital grains in contrast to pedogenic calcretes that have fine silt-size carbonate crystals. Second, groundwater calcretes impregnate sediments in a manner that preserves the original sedimentary structure in contrast to pedogenic calcretes that obliterates sedimentary structure.

The *per descensum*-leaching model accounts for pedogenic carbonate formation as the result of dissolution of indigenous carbonates from upper horizons, downward transport of ions in solution, and precipitation of secondary carbonates from soil solution in lower horizons. Biogenic CO_2 raises P_{CO_2} to levels as high as 1.2%, and the relative rates of CO_2 production and loss to the atmosphere by diffusion generally cause P_{CO_2} to increase with soil depth (McFadden et al., 1991). Plant roots take up water preferentially over dissolved HCO_3^-, causing $CaCO_3$ to precipitate due to desiccation. This process occurs in the upper profile

of many calcareous loess and alluvial deposits and was invoked to explain the correlation of progressively shallower carbonate horizons in progressively drier climates (Jenny and Leonard, 1934). Later, this model was used as the basis for calculating the number of wetting fronts required to leach carbonates to a particular depth (Arkley, 1963). In both cases, it was assumed that carbonate was uniformly distributed in parent material at the beginning of pedogenesis.

The *per descensum*-dust model was developed in recognition that pedogenic carbonate also occurs in soils that do not have indigenous calcareous parent materials. In southern New Mexico, for example, prominent calcic and petrocalcic horizons occur in soils with siliceous and only-slightly weathered parent materials. In this case, atmospheric additions were judged to be the source of pedogenic carbonates (Gile et al., 1966). Initially, calcareous dust was considered to be the source. Later, it was recognized that Ca^{2+} in rain was an additional, and even larger, source of Ca^{2+} for reaction with soil HCO_3^- to form pedogenic carbonates (Section 33.10.5.1.2). The depth and rate of $CaCO_3$ accumulation in many Aridisols illustrate the importance of varying eolian input and P_{CO_2} (McFadden et al., 1991). Building on these *per descensum* concepts, compartmental models have been constructed that compute the depth, amount, and distribution of pedogenic carbonate as a function of climate and time (Arkley, 1963; McFadden and Tinsley, 1985; Hirmas et al., 2009; Section 33.10.6).

The in situ-biogeochemical model focuses on the biogenic activity of plants and microorganisms in harvesting Ca from parent material and transporting it to the land surface. This process occurs in all soils, but its effect is more apparent in arid and semiarid soils because of their nonflushing nature and nonacidic pH. The in situ-biogeochemical model operates in both noncalcareous and calcareous parent materials. In noncalcareous parent materials, evidence that various plants play a role in carbonate formation comes from chemical analysis of soil profiles showing that plants transfer Ca^{2+} to the land surface from subsoil, rock, or groundwater (Goudie, 1973). Depending on the chemical environment, a portion of this Ca^{2+} is available for precipitation with HCO_3^- derived from CO_2 generated by root and microbial respiration. Moreover, references in the Russian literature note carbonate formation in plant tissue followed by its downward translocation with wetting fronts (Labova, 1967). Evidence that microorganisms precipitate carbonates is based on microscopy of calcified fungal hyphae (West et al., 1988) and in vitro laboratory experiments (e.g., Philips et al., 1987; Monger et al., 1991). Termites may also precipitate carbonate in large termite mounds of Africa and southeast Asia (Thorp, 1949), but not in small sheaths common in North American deserts (Liu et al., 2007). In formed on limestone bedrock, pedogenic carbonate forms via the progressive transformation of limestone into micritic pedogenic carbonate as a result of short-range carbonate dissolution and reprecipitation processes (Rabenhorst and Wilding, 1986), which is a unique process of pedogenic carbonate accumulation because the content of pedogenic carbonate is less than the carbonate content of the original limestone.

33.10.5.6 Pedogenic Gypsum and Soluble Salts

Soils containing gypsum ($CaSO_4 \cdot 2H_2O$) in the form of gypsic or petrogypsic horizons are estimated to cover some 207 million ha around the world (Eswaran and Zi-Tong, 1991). These soils can be grouped into two categories: *gypsiferous*(in which the suffix-ferous means "yielding" or "containing") or *gypseous* (in which the suffix-ous means "abounding in" or "consisting of"; Herrero and Porta, 2000). Unlike pedogenic carbonates that can form in any soil given a source of free calcium, bicarbonate, and nonacidic pH, gypsum formation is restricted to a parent material source of gypsum.

Soils with gypsum have unique properties because of gypsum's softness (2 on the Mohs scale), its solubility (2.6 g L^{-1} at 25°C), and its ability to corrode iron and disintegrate concrete. The disintegration of concrete occurs when sulfate reacts with Na to form mirabilite or thenardite or with Ca and Al to form ettringite $[Ca_6Al_2(SO_4)_3(OH)_{12} \cdot 26H_2O)]$ and thaumasite $[Ca_3Si(CO_3)(SO_4)(OH)_6 \cdot 12(H_2O)]$ (Herrero et al., 2009). Crystallization of these sulfate minerals increases the volume of solids, thus leading to disintegration.

Gypseous and gypsiferous soils are not necessarily salt-affected soils because gypsum does not significantly increase osmotic potential or ionic toxicity. In fact, gypsum has long been used as a source of calcium and sulfur nutrients, and a material used to reclaim sodic soils (Shainberg et al., 1989). Native vegetation on gypseous soils is distinguished from vegetation of saline soils by having more abundant biomass and greater number species. However, gypseous and gypsiferous can also be saline (Herrero et al., 2009). Salts more soluble than gypsum (soluble salts) most commonly accumulate in soils where there is a source of easily dissolvable minerals or ions in parent material (Gumuzzio et al., 1982), a source of saline water, and a climate with relatively high evapotranspiration rates. Pedogenic accumulation of soluble salts is often associated with gypsum and/or carbonates, depending on the initial ionic composition of the evaporating solution (Boettinger and Richardson, 2001).

Gypsic and salic horizons are typical of basins with playas (Driessen and Schoorl, 1973) where leaching is limited, water tables are high, and where eolian salts are recycled between playas and surrounding terrain. These salts in soil profiles are often seasonally dynamic due to their high solubility and the ease with which they are moved downward by leaching during the wet season and upward by evapotranspiration during the dry season (Eghbal et al., 1989). Aridisols with salic and gypsic horizons may also occur where there is throughflow and seeps (e.g., Buck et al., 2006) and artesian springs in arid to semiarid climates (Boettinger and Richardson, 2001).

33.10.5.7 Illuvial Clay and the Argillic Horizon in Aridisols

The processes by which clay accumulates in Aridisols are similar to those in other soil orders and include disaggregation, dispersion, and illuviation of clays in eolian dust, in situ weathering of primary minerals, illuviation of dispersed clay from surface horizons, and neosynthesis of clay minerals from soil solution (Nikiforoff, 1937; Brown and Drosdoff, 1940; Agriculture Experiment Station-Soil Conservation Service, 1964). Argillic and natric horizons in Aridisols can range in texture from sand to clay with thickness from 7.5 to 75 cm. Argillic and natric horizons generally begin at shallow depths (4–25 cm below the soil surface), which are shallower than argillic and natric horizons in soils of more humid regions. Argillic horizons typically are just above or extend into calcic horizons.

Mineral weathering that produced clays susceptible to dispersion and translocation is primarily a relict of wetter Pleistocene climatic conditions (Nettleton and Peterson, 1983). Wetter conditions of the Pleistocene pluvials were probably also required to dissolve and leach carbonates from surface horizons, the removal of Ca being a prerequisite for dispersion and translocation of the clays (Goss et al., 1973; Southard, 2000). This was shown by the absence of argillic horizons in Aridisols with abundant fragments of limestone in contrast to neighboring Aridisols of the same age formed in igneous alluvium (Gile and Hawley, 1972). The explanation apparently lies in the flocculating effect that carbonates have on clay movement, although clay illuviation has been reported in some calcareous soils (Goss et al., 1973; Pal et al., 2003). The absence of argillic horizons in many Aridisols can be the result of obliteration by landscape dissection (Nordt et al., Section 33.4), engulfment by carbonate (Gile et al., 1969), or faunal mixing (Gile, 1975a,b).

Many argillic horizons in Aridisols appear to be illuvial, even though they lack clay skins (argillans) on peds. This conclusion has been based on several lines of evidence: (1) Soils have thin, grayish E-like horizon with less clay than the underlying reddish-brown or red B horizon; (2) the argillic horizons when traced laterally have prominent clay skins in pipes that penetrate calcic horizon (Figure 33.67f); (3) reddish coatings of silicate clay occur on and in cracks in the tops of petrocalcic horizons and bedrock that underlie argillic horizons without clay skins at shallow depths; and (4) distinct linear bodies of oriented clay occur within peds, interpreted as former clay skins (Buol and Yesilsoy, 1964; Gile and Grossman, 1968; Smith and Buol, 1968; Nettleton et al., 1969, 1975; Gile et al., 1981).

33.10.6 Links to Past Climates

Aridisols are some of the oldest, if not the oldest soils in the world. They range from Cretaceous age or older in Australia (Pillans, 2005), through the Miocene (Hawley, 2005), the Pliocene (e.g., Gardner 1972; Mack et al., 1996), and early Pleistocene in many regions of the world (e.g., Nettleton and Peterson, 1983; Machette, 1985; Alsharhan et al., 1998; Kapur et al., 2000; Matmon et al., 2009). Such ancient soils bear the marks of many paleoclimatic changes at both the landscape scale and profile scale.

At the landscape scale, periods of landscape stability (soil formation) alternate with periods of landscape instability (i.e., erosion and sedimentation), as described by Ruhe (1962):

> By analogy with present condition … and on a theoretical basis … Past pluvial environments in present arid regions, correlative of glacial episodes elsewhere, should have resulted in greater vegetative cover on relatively stable landscapes … and soil formation. Interpluvial environments, as at present, should have resulted in increased aridity, lesser vegetative cover, unstable landscapes subject to severe erosion and to sediment transport.

Because shrublands have much bare ground, erosion increases. In contrast, during wetter climates, grasslands and woodlands (which have little bare ground) increase causing erosion to decrease (Langbein and Schumm, 1958). Such cycles commonly produce stacked sequences of buried paleosols in intermontane basins. Along floodplains, glacial–interglacial climate cycles commonly produce stepped sequences of geomorphic surfaces (Figure 33.71a). In the American Southwest, the model is a follows: (1) Rivers downcut during glacial periods when greater precipitation gives rise to more water in river channels and denser vegetative cover across landscapes, (2) during waning glacial and early interglacial times, river valleys backfill (partially) as a result of less water in the river channels and greater amounts of sediment supplied by erosion from sparsely vegetated landscapes, and (3) aggradation ceases and base levels stabilize during interglacial times, until the cycle began again when downcutting is renewed by the waxing phase of the next glacial cycle (Hawley et al., 1976).

At the pedon scale, mineralogy is a function of climate. Highly weathered minerals can be relics from a wetter climate
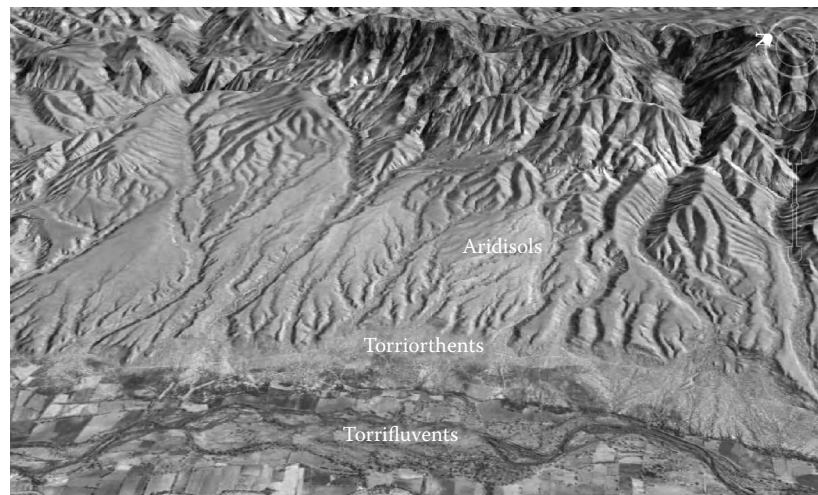
(a)
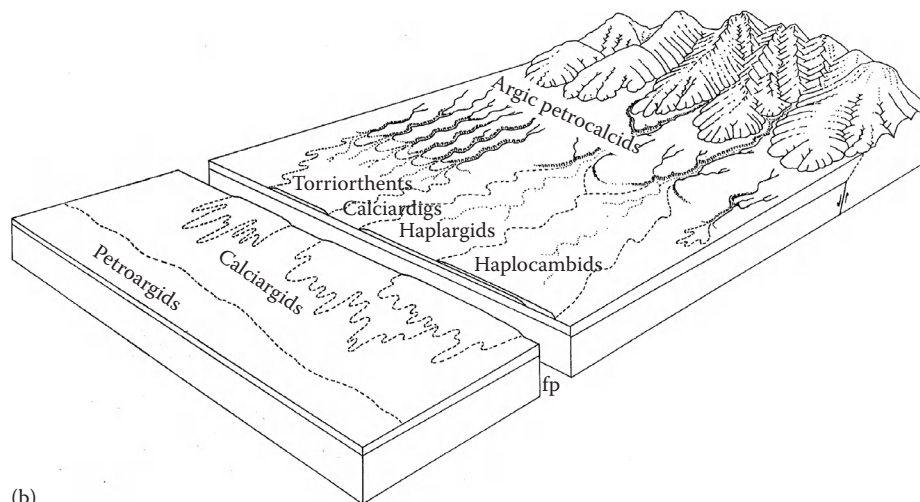
(b)

FIGURE 33.71 (a) Aridisols and Entisols (Torriorthents) on stepped geomorphic surfaces bordering a river flood plain (Torrifluvents) in Sonora, Mexico. The stepped surfaces record climate-controlled entrenchment and partial backfilling. Older land surfaces are topographically higher and can persist for hundreds of thousands of years, especially if armored by petrocalcic horizons. (b) Landscape relationships of different types of Aridisols in the Basin and Range Province, western United States. (Modified from Peterson, F.F. 1980. Holocene desert soil formation under sodium salt influence in a playa-margin environment. Quat. Res. 13:172–186.)

(Section 33.10.4.3). In contrast, soluble minerals reflect dry climates. Pedogenic carbonate is particularly valuable for providing evidence of climate change in at least three forms: its presence or absence, its depth, and its carbon and oxygen isotopes signatures.

Pedogenic carbonate is a feature of aridity (Mack, 1992; Guo et al., 2006). Although it can exist in humid climates under certain circumstances of high water tables and calcareous parent materials (Sobecki and Wilding, 1983), an analysis of 1168 soil profiles with carbonate showed that 95% existed in climates where the annual precipitation is less than 760 mm (Royer, 1999). A typical value of 500 mm (20 in.) has been used as the general boundary between soils with carbonate and soils without carbonate (Birkeland, 1999). The inference, therefore, can be made that paleosols containing pedogenic carbonate formed in dry climates.

Depth of pedogenic carbonate is commonly, but not always, proportional to annual rainfall. In general, greater precipitation corresponds to greater depths carbonate horizons (Jenny and Leonard, 1934; Arkley, 1963; Gile, 1975c). However, erosion, runoff, and run-on can confound this general relationship, as does the seasonal distribution of precipitation. Arid climates with winter rainfall often produce deeper leaching of carbonates than those with a similar amount of summer rainfall (Yaalon, 1983). It has been shown using the USDA-NRCS database that no statistically significant correlation exists between carbonate depth and rainfall, especially for shallow carbonate in arid and semiarid climates (Royer, 1999). Still, within a local region where erosion, runoff, and run-on can be held constant, carbonate depth is likely to be deeper in wetter environments than in drier environments.

Deeper wetting fronts in the Pleistocene are probably responsible for vertical, karst-like pipes that cross-cut petrocalcic and calcic horizons (Figure 33.67d). Similarly, carbonate filaments in B horizons overlying petrocalcic and calcic horizons in soils of Pleistocene age are probably the result of an upward shift in the depth of wetting during subsequent drier climates based on depths of carbonates in soils of Holocene age and radiocarbon dates of the carbonate crystals themselves (Gile et al., 1981; Monger, 2003). Other evidence for climatically driven shifts in carbonate depth includes engulfment of argillic horizons by calcic horizons as the depth of wetting shifts upward with increasing aridity (Nettleton and Peterson, 1983), and micromorphologic evidence of episodic deposition of carbonates, opal, and clay in argillic horizons and duripans (Eghbal and Southard, 1993c).

Simulation models are also useful for understanding links between carbonate depth and climate. Such models have a vertical sequence of compartments, each with a specified texture, bulk density, water-holding capacity, mineralogy, CO_2 content, ionic strength, and temperature (McFadden and Tinsley, 1985). Similarly, models by Marion et al. (1985) and later models by Marion and Schlesinger (1994) and Hirmas et al. (2009) have used stochastic precipitation, evapotranspiration, chemical thermodynamics, soil parameterization, and soil water movement to simulate carbonate distribution. The depth of carbonate for a soil in New Mexico, for example, was modeled as being about 15 cm deeper during wetter Pleistocene climates than its present arid conditions (Marion et al., 1985).

Carbon ($^{13}C/^{12}C$) and oxygen ($^{18}O/^{16}O$) isotopes in pedogenic carbonate contain information about paleoclimate in two ways: paleoecological and paleotemperature. First, $^{13}C/^{12}C$ ratios are linked to the relative abundance of C_4 to C_3 plants (i.e., plants that use either the C_4 or C_3 photosynthetic pathways; Cerling, 1984; Amundson et al., 1988; Ehleringer, 1988; Quade et al., 1989; Monger et al., 2009). Aridisols are associated with both groups: C_4 plants are warm-season desert grasses (although one shrub type is included, *Atriplex* spp.) and C_3 plants are desert shrub species (Sage et al., 1999). Pedogenic carbonate formed in a pure C_3 ecosystem has $\delta^{13}C$ values of around −12‰ to −9‰, in contrast to pedogenic carbonate formed under pure C_4 vegetation which has $\delta^{13}C$ values of around 1‰–3‰ (Cerling, 1999). Aridisols are also associated with a third group—the CAM plants (crassulacean acid metabolism)—which include the cacti and have isotopic values intermediate between C_4 and C_3 plants.

Second, oxygen isotopes have been used to make inferences about paleotemperatures and rainwater sources (Amundson et al., 1996; Liu et al., 1996). With higher mean annual temperatures, rainwater becomes preferentially enriched in ^{18}O, which is preserved in pedogenic $CaC^{18}O_3$ (e.g., Cerling and Quade, 1993). The isotopic signature of rain, however, is complicated because it is influenced by several factors: (1) latitudinal location (isotopically lighter rain occurs poleward); (2) seasonality (winter rain is isotopically lighter and percolates deep in soil); (3) mountain ranges and distance from the coast (isotopically lighter rain at higher altitudes and greater distance from coast); (4) the amount of rain (rain becomes lighter as the amount of rain increases, this includes both monthly rainfall and individual rainfall events); and (5) storm trajectories (source area, including vapor from land, yields different $\delta^{18}O$ values; Grootes, 1993).

33.10.7 Agriculture

Agricultural use of Aridisols is limited chiefly by the lack of irrigation water. Center pivot irrigation, however, has brought vast areas of Aridisols on upland landforms into agricultural production (e.g., Navajo Indian and Saudi Arabia irrigation projects). Where irrigation systems have been developed, most Aridisols present some problems and generally are less well suited for irrigation than are Entisols, such as Torriorthents and Torrifluvents, which are commonly associated with Aridisols (Figure 33.71a). Land leveling to allow flood, furrow, or sprinkler irrigation often exposes duripans, calcic, petrocalcic, natric, or argillic horizons.

Under irrigation, control of salinity in the rooting zone is critical to the long-term success of any crop management system. Flushing of salts by an amount of water in excess of the crop evapotranspiration demand (a leaching fraction) helps reduce salinity in the root zone. The excess water and dissolved salts must be removed by subsurface drainage. The presence of shallow subsurface horizons with slow permeability (duripans, natric or petrocalcic horizons) compounds the irrigation and drainage management problems for crop production, as well as for septic tank leach fields and

irrigation of lawns in urbanized areas. Soil subsidence, due to solution of gypsum, and corrosion of concrete are common problems where soils containing gypsum are irrigated (Nettleton et al., 1982; Section 33.10.5.6). Other than N, most major plant nutrients are abundant in Aridisols, although micronutrient availability is often limited by low solubility at alkaline pH. Careful management of P is also often needed due to sorption of P by $CaCO_3$.

33.10.8 Ecosystem Services

The term *ecosystem services* refer to ways humans benefit economically from processes and resources that are supplied by natural ecosystems for free, such as clean drinking water and pollination of crops (Withgott and Brennan, 2009). Aridisols provide many ecosystem services. Three prominent categories include *provisioning*, *regulating*, and *cultural* ecosystem services.

33.10.8.1 Provisioning Ecosystem Services of Aridisols

Grazing by livestock (mainly cattle, sheep, and goats) and wildlife is the common use of Aridisols where water is not available for irrigation. More than half of the Earth's land surface is grazed (Follett et al., 2001); most of which occurs in the dry regions (e.g., Sobecki et al., 2001). In the United States, for example, there are approximately 308 million ha of rangeland, about 31% of the total area (Havstad et al., 2007). Most vegetation on Aridisols, however, is dominated by drought-escaping annual grasses, forbs, and drought-enduring evergreen shrubs (Archibold, 1995). These have some grazing value, but less than the grazing value of Aridisols transitional to xeric or ustic soil moisture regimes, or frigid or cryic soil temperature regimes, that often have a significant component of perennial grasses. Forage on Aridisols is highly variable (Eckert et al., 1978; Schlesinger and Jones, 1984), ranging up to 10-fold differences in forage production (Ludwig, 1987). Still, huge amounts of beef have historically been generated on indigenous vegetation of Aridisols (Fredrickson et al., 1998). Overgrazing on some Aridisols has been accompanied by a drastic change in the structure of vegetation giving rise to geomorphic feedbacks, such as the formation of coppice dunes (Gile, 1966; Otterman and Gornitz, 1983; Schlesinger et al., 1990; Reid et al., 1993). In the past few decades in the United States, food production from the livestock industry using Aridisols has declined because of poor economic returns, increased regulation, an aging rural population, and increasing use of drylands for diverse purposes, such as exurban development (Havstad et al., 2007).

Medicinal plants are another provisioning service associated with Aridisols. Many desert plants, especially evergreen perennials, have evolved chemical defenses that may be prime sources of anticancer compounds (Donaldson and Cates, 2004). For example, extracts from a screen of 63 Sonoran desert plants revealed that 42% of evergreen species, 37% of woody perennials, 36% herbaceous perennials, and 23% of annuals showed greater than 50% inhibition against HeLa cells. Other potential sources of anticancer compounds are found in the rhizosphere of plant roots growing in Aridisols (He et al., 2004) and in plants with genetic variability among the populations unique to Aridisol habitat, such as *Anemopsis californica* (Medina-Holguín et al., 2007, 2008).

33.10.8.2 Regulating Ecosystem Services of Aridisols

Aridisols are important in regulating air quality at both the local and global scales. At the local scale, dust storms can reduce visibility on highways and aggravate respiratory problems in areas where Aridisols have been disturbed by construction or agriculture (e.g., Bar-Ziv and Goldberg, 1974). Globally, Aridisols are important sources of atmospheric dust that not only affect air quality, but also influence rain chemistry, ocean fertilization, and albedo (Schlesinger, 1997; Kaufman et al., 2005). The key variable linking Aridisols and air quality are ground cover in the form of desert pavement, biological soil crusts, or vascular plants. When ground cover is removed or disturbed, soil particles, especially in the very fine sand and silt fractions, are vulnerable to entrainment and movement by wind (e.g., Belnap and Gillette, 1998).

Aridisols play a similarly important role in regulating water quality. While forested mountains are the source of most water in regions with Aridisols, a significant proportion is also generated by lower elevation rangelands during infrequent and often torrential rainfall events (Havstad et al., 2007). Aridisols at lower elevations that are occupied by shrublands generally produce more runoff than those occupied by grasslands (Schlesinger et al., 2000; Wilcox and Thurow, 2006). As overland flow moves across shrublands, it carries more sediment than in grasslands. Moreover, soil moisture is more heterogeneous in shrublands than in grasslands because soil moisture is deeper beneath rills and gullies (Dick-Peddie, 1993). As with air quality, a highly disturbed landscape of Aridisols will give rise to lower quality water than neighboring nondisturbed Aridisols.

Carbon sequestration by soil is an important ecosystem service of global significance. Aridisols, however, have low contents of organic carbon, only 4% of the total, but contain the greatest amount of soil inorganic carbon (SIC, calcium carbonate), 48% of the total (Eswaran et al., 2000). Soil inorganic carbon is greatest in Aridisols because, as discussed earlier (Section 33.10.5.5), its occurrence is mainly limited to dry climates. Of the global 946 Pg of SIC, arid regions contain 77.8%, followed by semiarid regions (14.2%), Mediterranean climates (5.4%), permafrost (1.9%), humid (0.5%), and perhumid areas (0.2%) (Eswaran et al., 2000). Although the global amount of SIC is large, containing more carbon than the amount in terrestrial vegetation (560 Pg) or the atmosphere (820 Pg), the rate at which SIC forms is relatively low: 0.3–15 g $CaCO_3$ m^{-2} year^{-1} (Machette, 1985; Eghbal and Southard, 1993b; Reheis et al., 1995; Monger and Gallegos, 2000). The loss of CO_2 from SIC also appears to be low even when petrocalcic horizons are exposed at the land surface by erosion (Serna-Perez et al., 2006). However, because some SIC is linked to microbial and root precipitation, labile pools of SIC may exist, but presently remain unmeasured.

33.10.8.3 Cultural Ecosystem Services of Aridisols

Early civilizations arose on Torrifluvents and neighboring Aridisols, such as those in Sumeria in the fourth millennium

B.C. when irrigated agriculture in that arid climate encouraged stable settlements, led to surplus food, freed people to pursue specialized trades and develop social order and cultural creativity (Durant, 1935). Similar cultural developments arose in arid and semiarid climates along the Indus River of ancient India, Hoang-Ho (Yellow River) of ancient China, and in dryland climates of the western Hemisphere, such as those occupied by the Inca, Anasazi, and Hohokan cultures. Today, arid climates are well known for their archaeological preservation (e.g., Egypt) as well as other cultural ecosystem services including wildlands, solitude, recreation, esthetic experiences, and places for scientific discovery.

33.10.9 Desertification

Desertification refers to the broad set of environmental degradation processes that results in a persistent decrease in the productivity of drylands—a condition that impoverishes an estimated 1 billion people (Verstraete et al., 2009). Such degradation typically involves the loss of productive grasslands and their replacement by shrublands accompanied by erosion. In addition to erosional degradation, desertification of Aridisols also includes chemodesertification in which improper irrigation causes salt accumulation (Dregne, 1983). Desertification of grazingland Aridisols can result in the spatial distribution of plants becoming very heterogeneous (Eckert et al., 1978; Schlesinger and Jones, 1984), commonly giving rise to islands of fertility (Schlesinger et al., 1990). These islands tend to be relatively stable, resistant to erosion (Otterman and Gornitz, 1983) and often trap eolian very fine and fine sand to form coppice dunes (Gile, 1966; Reid et al., 1993). These coppice dunes can range in height from a few centimeters around the base of shrubs to 1 m or more, in which case they dominate the landscape microtopography.

Desertification involves ecogeomorphic thresholds. When these thresholds are exceeded, the system is slow to recover in large part because the soil resource has changed (Bestelmeyer et al., 2009). Forces that drive desertification are coupled environmental and human systems, such as short-term climate variability and exploitation of scarce resources in areas remote from centers of political power (Reynolds et al., 2007). Recent proposals to combat desertification involve the effort to view desertification as (1) a provider of ecosystem services, (2) a coupled human ecological system, and (3) a system that must be viewed at multiple scales in order to understand its causes and consequences (Verstraete et al., 2009).

References

Agricultural Experiment Stations-Soil Conservation Service. 1964. Soils of the western United States. Washington State University, Pullman, WA.

Ahrens, C.D. 1991. Meteorology today: An introduction to weather, climate, and the environment. 4th edn. West Publishing Co., St. Paul, MN.

Ahrens, R.J., and H. Eswaran. 2000. The international committee on aridisols: Deliberations and rationale. Soil Surv. Horiz. 41:110–117.

Alexander, E.B. 1995. Silica cementation in serpentine soils in the humid Klamath Mountains, California. Soil Surv. Horiz. 36:154–159.

Alexander, E.B., R.C. Graham, and C.L. Ping. 1994. Cemented ultramafic till beneath a podzol in southeast Alaska. Soil Sci. 157:53–58.

Alexandre, A., J.D. Meunier, F. Colin, and J.M. Koud. 1997. Plant impact on the biogeochemical cycle of silicon and related weathering processes. Geochim. Cosmochim. Acta 61:677–682.

Allen, B.L. 1990. Mineralogy of aridisols, p. 191–196. *In* J.M. Kimble and W.D. Nettleton (eds.) Proc. 4th Int. Soil Correl. Meet. (ISCOM) charact. classif. util. aridisols. Part A: Papers. USDA-SCS, Lincoln, NE.

Alsharhan, A.S., K.W. Glennie, G.L. Whittle, and C.G. St. C. Kendall (eds.). 1998. Quaternary deserts and climatic change. A.A. Balkema, Rotterdam, the Netherlands.

Amundson, R., O. Chadwick, C. Kendall, Y. Wang, and M. DeNiro. 1996. Isotopic evidence for shifts in atmospheric circulation patterns during the late Quaternary in mid-North America. Geology 24:23–26.

Amundson, R., O.A. Chadwick, J.M. Sowers, and H.E. Doner. 1988. Relationship between climate and vegetation and the stable carbon isotope chemistry of soils in the eastern Mojave Desert, Nevada. Quat. Res. 29:245–254.

Anderson, K., S. Wells, and R. Graham. 2002. Pedogenesis of vesicular horizons, Cima Volcanic Field, Mojave Desert, California. Soil Sci. Soc. Am. J. 66:878–887.

Antevs, E. 1955. Geologic-climatic dating in the West. Am. Antiq. 20:317–335.

Archibold, O.W. 1995. Ecology of world vegetation. Chapman and Hall, London, U.K.

Arkley, R.J. 1963. Calculation of carbonate and water movement in soil from climatic data. Soil Sci. 96:239–248.

Baldwin, M., C.E. Kellogg, and J. Thorp. 1938. Soil classification, p. 979–1001. USDA yearbook of agriculture. U.S. Government Printing Office, Washington, DC.

Bar-Ziv, J., and G.M. Goldberg. 1974. Simple siliceous pneumoconiosis in Negev Bedouins. Arch. Environ. Health 29:121–126.

Bhark, E.W., and E.E. Small. 2003. Association between plant canopies and the spatial patterns of infiltration in shrubland and grassland of the Chihuahuan Desert, New Mexico. Ecosystems 6:185–196.

Belnap, J. 2002. Nitrogen fixation in biological soil crusts from southeast Utah, USA. Biol. Fertil. Soils 35:128–135.

Belnap, J. 2006. The potential roles of biological soil crusts in dryland hydrologic cycles. Hydrol. Process. 20:3159–3178.

Belnap, J., and D.A. Gillette. 1998. Vulnerability of desert soil surfaces to wind erosion: The influences of crust development, soil texture, and disturbance. J. Arid Environ. 39:133–142.

Belnap, J., and O.L. Lange (eds.) 2001. Biological soil crusts: Structure, function, and management. Vol. 150. Springer-Verlag, Berlin, Germany.

Bestelmeyer, B.T., A.J. Tugel, G.L. Peacock, Jr., D.G. Robinett, P.L. Shaver, J.R. Brown, J.E. Herrick, H. Sanchez, and K.M. Havstad. 2009. State-and-transition models for heterogeneous landscapes: A strategy for development and application. Rangeland Ecol. Manage. 62:1–15.

Birkeland, P.W. 1984. Soils and geomorphology. Oxford University Press, New York.

Birkeland, P.W. 1999. Soils and geomorphology. Oxford University Press, New York.

Boettinger, J.L., and J.L. Richardson. 2001. Saline and wet soils of wetlands in dry climates, p. 383–390. *In* J.L. Richardson and M.J. Vepraskas (eds.) Wetland soils: Their genesis, hydrology, landscapes, and classification. Lewis Publishers, Boca Raton, FL.

Boettinger, J.L., and R.J. Southard. 1990. Micromorphology and mineralogy of a calcareous duripan formed in granitic residuum, Mojave Desert, California, USA. Dev. Soil Sci. 19:409–416.

Boettinger, J.L., and R.J. Southard. 1991. Silica and carbonate sources for Aridisols on a granitic pediment, western Mojave Desert. Soil Sci. Soc. Am. J. 55:1057–1067.

Boettinger, J.L., and R.J. Southard. 1995. Phyllosilicate distribution and origin in Aridisols on a granitic pediment, western Mojave Desert. Soil Sci. Soc. Am. J. 59:1189–1198.

Breecker, D.O., X.D. Sharp, and L.D. McFadden. 2009. Seasonal bias in the formation and stable isotopic composition of pedogenic carbonate in modern soils from central New Mexico, USA. Geol. Soc. Am. Bull. 121:630–640.

Broecker, W.S., and T. Liu. 2001. Rock varnish: Recorder of desert wetness? GSA Today 11:4–10.

Brown, I.C., and M. Drosdoff. 1940. Chemical and physical properties of soils and their colloids developed from granitic materials in the Mojave Desert. J. Agric. Res. 61:38–344.

Buck, B.J., K. Wolff, D.J. Merkler, and N.J. McMillan. 2006. Salt mineralogy of Las Vegas Wash, Nevada: Morphology and subsurface evaporation. Soil Sci. Soc. Am. J. 70:1639–1651.

Bull, W.B. 1991. Geomorphic response to climatic change. Oxford University Press, New York.

Buol, S.W., F.D. Hole, and R.J. McCracken. 1980. Soil genesis and classification. 2nd edn. Iowa State University Press, Ames, IA.

Buol, S.W., R.J. Southard, R.C. Graham, and P.A. McDaniel. 2002. Soil genesis and classification. 5th edn. Iowa State University Press, Ames, IA.

Buol, S.W., and M.S. Yesilsoy. 1964. A genesis study of a Mojave sandy-loam profile. Soil Sci. Soc. Am. Proc. 8:254–256.

Capo, R.C., C.E. Whipkey, J.R. Blachere, and O.A. Chadwick. 2000. Pedogenic origin of dolomite in a basaltic weathering profile, Kohala Peninsula, Hawaii. Geology 28:271–274.

Cerling, T.E. 1999. Paleorecords of C_4 plants and ecosystems, p. 445–469. *In* R.F. Sage and R.K. Monson (eds.) C_4 plant biology. Academic Press, San Diego, CA.

Cerling, T.E., and J. Quade. 1993. Stable carbon and oxygen isotopes in soil carbonates, p. 217–231. *In* P.K. Swart, K.C. Lohmann, J. McKenzie, and S. Savin (eds.) Climate change in continental isotopic records. American Geophysical Union Monograph 78, Washington, DC.

Chadwick, O.A., and J.O. Davis. 1990. Soil-forming intervals caused by eolian sediment pulses in the Lahontan basin, northwestern Nevada. Geology 18:243–246.

Chadwick, O.A., D.M. Hendricks, and W.D. Nettleton. 1987a. Silica in duric soils: II. Mineralogy. Soil Sci. Soc. Am. J. 982–985.

Chadwick, O.A., D.M. Hendricks, and W.D. Nettleton. 1987b. Silica in duric soils: I. A depositional model. Soil Sci. Soc. Am. J. 51:975–982.

Chadwick, O.A., and W.D. Nettleton. 1990. Micromorphologic evidence of adhesive and cohesive forces in soil cementation. Dev. Soil Sci. 19:207–212.

Cooke, R.U. 1970. Stone pavements in deserts. Ann. Assoc. Am. Geogr. 60:560–577.

Cooke, R., A. Warren, and A. Goudie. 1993. Desert geomorphology. UCL Press, London, U.K.

Dick-Peddie, W.A. 1993. New Mexico vegetation: Past, present, and future. University of New Mexico Press, Albuquerque, NM.

Donaldson, J., and R. Cates. 2004. Screening for anticancer agents from Sonoran desert plants: A chemical ecology approach. Pharm. Biol. 42:478–487.

Doner, H.E., and W.C. Lynn. 1989. Carbonate, halide, sulfate, and sulfide minerals, p. 279–330. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Dorn, R.I., and T.M. Oberlander. 1981. Microbial origin of desert varnish. Science 213:1245–1247.

Dregne, H.E. 1976. Soils of arid regions. Elsevier Scientific Publishers, Amsterdam, the Netherlands.

Dregne, H.E. 1983. Desertification of arid lands. Harwood Academic Press, New York.

Driessen, P.M., and R. Schoorl. 1973. Mineralogy and morphology of salt efflorescences on saline soils in the Great Konya Basin, Turkey. J. Soil Sci. 24:436–442.

Duniway, M., J.E. Herrick, and H.C. Monger. 2007. The high water-holding capacity of petrocalcic horizons. Soil Sci. Soc. Am. J. 71:812–819.

Durant, W. 1935. Our oriental heritage. Simon and Schuster, New York.

Eckert, R.E., Jr., M.K. Wood, W.H. Blackburn, F.F. Peterson, J.L. Stephens, and M.S. Meurisse. 1978. Effects of surface-soil morphology on improvement and management of some arid and semi-arid rangelands, p. 299–302. *In* D.N. Hyder (ed.) Proc. 1st. Int. Rangeland Congr. Am. Soc. Range Manage., Denver, CO.

Eghbal, M.K., and R.J. Southard. 1993a. Mineralogy of aridisols on dissected alluvial fans, western Mojave Desert, California. Soil Sci. Soc. Am. J. 57:538–544.

Eghbal, M.K., and R.J. Southard. 1993b. Stratigraphy and genesis of Durorthids and Haplargids on dissected alluvial fans, western Mojave Desert, California. Geoderma 59:151–174.

Eghbal, M.K., and R.J. Southard. 1993c. Micromorphological evidence of polygenesis of three aridisols, western Mojave Desert, California. Soil Sci. Soc. Am. J. 57:1041–1050.

Eghbal, M.K., R.J. Southard, and L.D. Whittig. 1989. Dynamics of evaporite distribution in soils on a fan-playa transect in the Carrizo Plain, California. Soil Sci. Soc. Am. J. 53:898–903.

Ehleringer, J.R. 1988. Carbon isotope ratios and physiological processes in aridland plants, p. 41–54. *In* P.W. Rundel, J.R. Ehleringer, and K.A. Nagy (eds.) Stable isotopes in ecological research. Springer, New York.

Ekart, D.D., T.E. Cerling, I.P. Montanez, and N.J. Tabor. 1999. A 400 million year carbon isotope record of pedogenic carbonate: Implications for paleo atmospheric carbon dioxide. Am. J. Sci. 299:805–827.

Eswaran and Zi-Tong. 1991. Properties, genesis, classification, and distribution of soils with gypsum, p. 89–119. *In* W.D. Nettleton (ed.) Occurrence, characteristics and genesis of carbonate, gypsum, and silica accumulations in soils. SSSA, Madison, WI.

Eswaran, H., P.F. Reich, J.M. Kimble, F.H. Beiroth, E. Radmanabhan, and P. Moncharoen. 2000. Global carbon stocks, p. 15–61. *In* R. Lal, J.M. Kimble, H. Eswaran, and B.A. Stewart (eds.) Global climate change and pedogenic carbonates. CRC Press, Boca Raton, FL.

Flach, K.W., W.D. Nettleton, L.H. Gile, and J.C. Cady. 1969. Pedocementation: Induration by silica, carbonates, and sesquioxides in the Quaternary. Soil Sci. 107:442–453.

Flach, K.W., W.D. Nettleton, and R.E. Nelson. 1973. The micromorphology of silica-cemented soil horizons in western North America, p. 714–729. *In* G.K. Rutherford (ed.) Proc. 4th Int. Work. Meet. Soil Micromorph. Soil microscopy. QueenÕs University, Kingston, Ontario, Canada.

Follett, R.F., J.M. Kimble, and R. Lal. 2001. The potential of U.S. Grazing lands to sequester carbon and mitigate the greenhouse effect. CRC Press, New York.

Franks, P.C., and A. Swineford. 1959. Character and genesis of massive opal in Kimball Member, Ogallala Formation, Scott County, Kansas. J. Sediment. Petrol. 29:186–196.

Fredrickson, E., K.M. Havstad, R. Estell, and P. Hyder. 1998. Perspectives on desertification: Southwestern United States. J. Arid Environ. 39:191–208.

Gardner, L.R. 1972. Origin of the Mormon Mesa caliche, Clark County, Nevada. Geol. Soc. Am. Bull. 83:143–156.

Garvie, L.A.J, D.M. Burt, and P.R. Buseck. 2008. Nanometer-scale complexity, growth, and diagenesis in desert varnish. Geology 36:215–218.

Gile, L.H. 1961. A classification of ca horizons in the soils of a desert region, Dona Ana County, New Mexico. Soil Sci. Soc. Am. Proc. 25:52–61.

Gile, L.H. 1966. Coppice dunes and the Rotura soil. Soil Sci. Soc. Am. Proc. 30:657–660.

Gile, L.H. 1975a. Holocene soils and soil-geomorphic relations in an arid region of southern New Mexico. Quat. Res. 5:321–360.

Gile, L.H. 1975b. Causes of soil boundaries in an arid region: I. Age and parent material. Soil Sci. Soc. Am. Proc. 39:316–323.

Gile, L.H. 1975c. Causes of soil boundaries in an arid region: II. Dissection, moisture, and faunal activity. Soil Sci. Soc. Am. Proc. 39:324–330.

Gile, L.H., R.J. Ahrens, and S.P. Anderson (eds.). 2003. Supplement to the desert project soil monograph: Soils and landscapes of a desert region Astride the Rio Grande Valley near Las Cruces, New Mexico. Vol. III. Soil Survey Investigations Report No. 44. USDA-NRCS. National Soil Survey Center, Lincoln, NE.

Gile, L.H., and R.B. Grossman. 1968. Morphology of the argillic horizon in desert soils of southern New Mexico. Soil Sci. 106:6–15.

Gile, L.H., and R.B. Grossman. 1979. The desert project soil monograph. USDA-SCS, Washington, DC.

Gile, L.H., R.B. Grossman, and J.W. Hawley. 1969. Effects of landscape dissection on soils near University Park, New Mexico. Soil Sci. 108:273–282.

Gile, L.H., and J.W. Hawley. 1972. The prediction of soil occurrence in certain desert regions of the southwestern United States. *Soil Science Society of America Proceedings*. 36:119–124.

Gile, L.H., J.W. Hawley, R.B. Grossman, H.C. Monger, C.E. Montoya, and G.H. Mack. 1995. Supplement to the desert project guidebook, with emphasis on soil micromorphology. New Mexico Bureau of Mines and Mineral Resources, Bulletin 142. New Mexico Bureau of the Mines and Mineral Resources, Socorro, NM.

Gile, L.H., F.F. Peterson, and R.B. Grossman. 1965. The K horizon: A master soil horizon of carbonate accumulation. Soil Sci. 99:74–82.

Gile, L.H., F.F. Peterson, and R.B. Grossman. 1966. Morphological and genetic sequences of carbonate accumulation in desert soils. Soil Sci. 101:347–360.

Goss, D.W., S.J. Smith, and B.A. Stewart. 1973. Movement of added clay through calcareous materials. Geoderma 9:97–103.

Goudie, A. 1973. Duricrusts in tropical and subtropical landscapes. Oxford University Press, London, U.K.

Graham, R.C., D.R. Hirmas, Y.A. Wood, and C. Amrhein. 2008. Large near-surface nitrate pools in soils capped by desert pavement in the Mojave Desert, California. Geology 36:259–262.

Grootes, P.M. 1993. Interpreting continental oxygen isotope records. p. 37–46. In P.K. Swart, K.C. Lohmann, J. McKenzie, and S. Savin (eds.) Climate change in continental isotopic records. American Geophysical Union Monograph 78, Washington, DC.

Gumuzzio, J., J. Batlle, and J. Casas. 1982. Mineralogical composition of salt efflorescences in a typic salorthid, Spain. Geoderma 28:39–51.

Guo, Y., R. Amundson, P. Gong, and Q. Yu. 2006. Quantity and spatial variability of soil carbon in the conterminous United States. Soil Sci. Soc. Am. J. 70:590–600.

Harden, J.W., E.M. Taylor, M.C. Reheis, and L.D. McFadden. 1991. Calcic, gypsic, and siliceous soil chronosequences in arid and semi-arid environments, p. 1–16. *In* W.D. Nettleton (ed.) Occurrence, characteristics and genesis of carbonate, gypsum, and silica accumulations in soils. SSSA, Madison, WI.

Havstad, K.M., D.P.C. Peters, R. Skaggs, J. Brown, B. Bestelmeyer, E. Fredrickson, J. Herrick, and J. Wright. 2007. Ecological services to and from rangelands of the United States. Ecol. Econ. 64:261–268.

Hawley, J.W. 2005. Five million years of landscape evolution in New Mexico: An overview based on two centuries of geomorphic conceptual-model development. p. 9–94. In S.G. Lucas, G.S. Morgan, and K.E. Zeigler (eds.), New Mexico's ice ages. Bulletin No. 28 New Mexico Museum of Natural History and Science, Albuquerque, NM.

Hawley, J.W., G.O. Bachman, and K. Manley. 1976. Quaternary stratigraphy in the Basin and Range and Great Plains provinces, New Mexico and western Texas, p. 235–274. *In* W.C. Mahaney (ed.) Quaternary stratigraphy of North America. Dowden Hutchinson, and Ross, Inc., Stroudsburg, PA.

He, J., E. Wijeratne, B. Bashyal, J. Zhan, C. Seliga, M. Liu, E. Pierson, L. Pierson, H. VanEtten, and A. Gunatilaka. 2004. Cytotoxic and other metabolites of *Aspergillus* inhabiting the rhizosphere of Sonoran desert plants. J. Nat. Prod. 67:1985–1991.

Herbel, C.H., and R.P. Gibbens. 1989. Matric potential of clay-loam soils on arid rangelands in southern New Mexico. J. Range Manage. 42:386–392.

Herbel, C.H., and L.H. Gile. 1973. Field moisture regimes and morphology of some arid-land soils in New Mexico, p. 119–152. *In* Field-soil water regime. SSSA Special Publication No. 5. SSSA, Madison, WI.

Herbel, C.H., L.H. Gile, E.L. Fredrickson, and R.P. Gibbens. 1994. Soil water and soils at the soil water sites, Jornada Experimental Range. In L.H. Gile and R.J. Ahrens (eds.) Supplement to the Desert Project Soil Monograph, Vol. I. Soil Survey Investigations Report No. 44. National Soil Survey Center, Lincoln, NE. 592 p.

Herrero, J., O. Artieda, and W.H. Hudnall. 2009. Gypsum, a tricky material. Soil Sci. Soc. Am. J. 73:1757–1763.

Herrero, J., and J. Porta. 2000. The terminology and the concepts of gypsum-rich soils. Geoderma 96:47–61.

Hipple, K.W., G.H. Logan, and M.A. Fosberg. 1990. Aridisols with cryic soil temperature regimes, p. 99–109. *In* J.M. Kimble and W.D. Nettleton (eds.) Proc. 4th Int. Soil Correl. Meet. (ISCOM): Charact. classif. util. aridisols. Part A: Papers. USDA-SCS, Lincoln, NE.

Hirmas, D.R., C. Amrhein, and R.C. Graham. 2009. Spatial and process-based modeling of soil inorganic carbon storage in an arid piedmont. Geoderma 154:486–494.

Huenneke, L.F., and I. Noble. 1996. Ecosystem function of biodiversity in arid ecosystems, p. 99–128. *In* H.A. Mooney, J.H. Cushman, E. Medina, O.E. Sala, and E.D. Schulze (eds.) Functional roles of biodiversity: A global perspective. John Wiley & Sons, New York.

Iler, R.K. 1979. The chemistry of silica. John Wiley & Sons, New York.

IUSS Working Group WRB. 2006. World reference base for soil resources 2006: A framework for international classification, correlation and communication. World Soil Resources Report 103. Rome, Italy.

Jackson, E.A. 1957. Soil features in arid regions with particular reference to Australia. J. Aust. Inst. Agric. Sci. 23:196–208.

Jenny, H., and C.D. Leonard. 1934. Functional relationships between soil properties and rainfall. Soil Sci. 38:363–381.

Johnston, R. 1997. Introduction to microbiotic crusts. USDA-NRCS, Soil Quality Institute and Grazing Lands Technology Institute, Fort Worth, TX.

Kapur, S., C. Saydam, E. Akca., V.S. Cavusgil, C. Karaman, I. Atalay, and T. Ozsoy. 2000. Carbonate pools in soils of the Mediterranean: A case study from Anatolia, p. 187–212. *In* R. Lal, J.M. Kimble, H. Eswaran, and B.A. Stewart (eds.) Global climate change and pedogenic carbonates. Lewis Publishers, Boca Raton, FL.

Kaufman, Y.J., I. Koren, L.A. Remer, D. Rosenfeld, and Y. Rudich. 2005. The effect of smoke, dust and pollution aerosol on shallow cloud development over the Atlantic Ocean. Proc. Natl. Acad. Sci. U. S. A. 102:11207–11212.

Kelly, E.F. 1989. A study of the influence of climate and vegetation on the stable isotope chemistry on soils in grassland ecosystems of the Great Plains. Ph.D. Dissertation. University of California. Berkeley, CA.

Kelly, E.F., R.G. Amundson, B.D. Marino, and M.J. DeNiro. 1991. Stable isotope ratios of carbon in phytoliths: A potential method of monitoring vegetation and climate change. Quat. Res. (NY) 35:222–233.

Kidron, G.J. 2007. Millimeter-scale microrelief affecting runoff yield over microbiotic crust in the Negev Desert. Catena 70:266–273.

Kidron, G.J., A. Yair, A. Vonshak, and A. Abeliovich. 2003. Microbiotic crust control of runoff generation on sand dunes in the Negev desert. Water Resour. Res. 39:1108–1112.

Kraimer, R.A., H.C. Monger, and R.L. Steiner. 2005. Mineralogical distinctions of carbonates in desert soils. Soil Sci. Soc. Am. J. 69:1773–1781.

Krumbein, W.E., and K. Jens. 1981. Biogenic rock varnishes of the Negev desert (Israel) an ecological study of iron and manganese transformation by cyanobacteria and fungi. Oecologia 50:25–38.

Labova, E. 1967. Soils of the desert zone of the USSR. Israel Program for Scientific Translation, Jerusalem, Israel.

Langbein, W.B., and S.A. Schumm. 1958. Yield of sediment in relation to mean annual precipitation. Am. Geophys. Union Trans. 39:1076–1084.

Levy, G.J. 2000. Sodicity, p. G27–G64. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Linacre, E., and B. Geerts. 1997. Climate and weather explained. Routledge, London, U.K.

Liu, X., H.C. Monger, and W.G. Whitford. 2007. Calcium carbonate in termite galleries—Biomineralization or upward transport? Biogeochemistry 82:241–250.

Liu, B., F. Phillips, and A.R. Campbell. 1996. Stable carbon and oxygen isotopes of pedogenic carbonates, Ajo Mountains, southern Arizona: Implication for paleoenvironmental change. Palaeogeogr. Palaeoclimatol. Palaeoecol. 124:233–246.

Lowenstam, H.A., and S. Weiner. 1989. On biomineralization. Oxford University Press, New York.

Ludwig, J.A. 1987. Primary productivity in arid lands: Myth and realities. J. Arid Environ. 13:1–7.

Ludwig, J.A., B.P. Wilcox, D.D. Breshears, D.J. Tongway, and A.C. Imeson. 2005. Vegetation patches and runoff-erosion as interacting ecohydrological processes in semiarid landscapes. Ecology 86:288–297.

Machette, M.N. 1985. Calcic soils of the southwestern United States, p. 1–21. *In* D.L. Weide (ed.) Soils and Quaternary geology of the southwestern United States. Geological Society of America. Special Paper 203. Boulder, CO.

Mack, G.H., 1992. Paleosols as an indicator of climatic change at the Early-Late Cretaceous boundary, southwestern New Mexico. J. Sediment. Petrol. 62:483–494.

Mack, G.H., W.C. McIntosh, M.R. Leeder, and H.C. Monger. 1996. Plio-Pleistocene pumice floods in the ancestral Rio Grande, southern Rio Grande rift, USA. Sediment. Geol. 103:1–8.

Malam Issa, O., L.J. Stal, C. Défarge, A. Couté, and J. Trichet. 2001. Nitrogen fixed by microbiotic crusts from desiccated Sahelian soils (Niger). Soil Biol. Biochem. 33:1425–1428.

Maliva, R.G., and R. Siever. 1988a. Mechanism and controls of silicification of fossils in limestone. J. Geol. 96:387–398.

Maliva, R.G., and R. Siever. 1988b. Diagenetic replacement controlled by force of crystallization. Geology 16:688–691.

Marbut, C.F. 1928. Subcommission II. Classification, nomenclature, and mapping of soils in the United States. Soil Sci. 25:61–71.

Marbut, C.F. 1935. Soils of the United States, atlas of American agriculture, Part III. USDA Bureau of Chemistry and Soils, Government Printing Office, Washington, DC.

Marion, G.M., W.H. Schlesinger, and P.J. Fonteyn. 1985. CALDEP: A regional model for soil $CaCO_3$ (caliche) deposition in southwestern deserts. Soil Sci. 139:468–481.

Massey, F.P., A.R. Ennos, and S.E. Hartley. 2006. Silica in grasses as a defence against insect herbivores: Contrasting effects on folivores and a phloem feeder. J. Anim. Ecol. 75:595–603.

Matmon, A., O. Simhai, R. Amit, I. Haviv, N. Porat, E. McDonald, L. Benedetti, and R. Finkel. 2009. Desert pavement—Coated surfaces in extreme deserts present the longest-lived landforms on Earth. GSA Bull. 121:688–697.

McFadden, L.D., R.G. Amundson, and O.A. Chadwick. 1991. Numerical modeling, chemical and isotopic studies of carbonate accumulation in soils of arid regions, p. 17–35. *In* W.D. Nettleton (ed.) Occurrence, characteristics and genesis of carbonate, gypsum, and silica accumulations in soils. SSSA, Madison, WI.

McFadden, L.D., and J.C. Tinsley. 1985. Rate and depth of pedogenic-carbonate accumulation in soils: Formation and testing of a compartment model, p. 23–42. *In* D.L. Weide (ed.) Soils and Quaternary geology of the southwestern United States. Geological Society of America Special Paper No. 203. Geological Society of America, Boulder, CO.

McFadden, L.D., S.G. Wells, and M.J. Jercinovic. 1987. Influences of eolian and pedogenic processes on the origin and evolution of desert pavements. Geology 15:504–508.

Medina-Holguín, A.L., F.O. Holguín, S. Micheletto, S. Goehle, J.A. Simon, and M.A. O'Connell. 2008. Chemotypic variation of essential oils in the medicinal plant, *Anemopsis californica*. Phytochemistry 69:919–927.

Medina-Holguín, A.L., C. Martin, S. Micheletto, F.O. Holguín, J. Rodriguez, and M.A. O'Connell. 2007. Environmental influences on essential oils in roots of *Anemopsis californica*. HortScience 42:1578–1583.

Meunier, J.D., F. Colin, and C. Alarcon. 1999. Biogenic silica storage in soil. Geology 27:835–838.

Miller, D.E. 1971. Formation of vesicular structure in soil. Soil Sci. Soc. Am. Proc. 35:635–637.

Milnes, A.R., M.J. Wright, and M. Thiry. 1991. Silica accumulations in saprolites and soils in South Australia, p. 121–149. *In* W.D. Nettleton (ed.) Occurrence, characteristics, and genesis of carbonate, gypsum, and silica accumulations in soils. SSSA Special Publication No. 26. SSSA, Madison, WI.

Ming, D.W., and J.L. Boettinger. 2001. Zeolites in soil environments. Rev. Miner. Geochem. 45:323–345.

Monger, H.C. 2003. Millennial-scale climate variability and ecosystem response at the Jornada LTER site, p. 341–369. *In* D. Greenland, D.G. Goodin, and R.C. Smith (eds.) Climate variability and ecosystem response at long-term ecological research sites. Oxford University Press, London, U.K.

Monger, H.C., and H.A. Adams. 1996. Micromorphology of calcite–silica deposits, Yucca Mountain, Nevada. Soil Sci. Soc. Am. J. 60:519–530.

Monger, H.C., and B.T. Bestelmeyer. 2006. The soil-geomorphic template and biotic change in arid and semiarid ecosystems. J. Arid Environ. 65:207–218.

Monger, H.C., D.R. Cole, B.J. Buck, and R.A. Gallegos. 2009. Scale and the isotopic record of C_4 plants in pedogenic carbonate: From the biome to the rhizosphere. Ecology 90:1498–1511.

Monger, H.C., and L.A. Daugherty. 1991a. Neoformation of palygorskite in a southern New Mexico aridisol. Soil Sci. Soc. Am. J. 55:1646–1650.

Monger, H.C., and L.A. Daugherty. 1991b. Pressure solution: Possible mechanism for silicate grain dissolution in a petrocalcic horizon. Soil Sci. Soc. Am. J. 55:1625–1629.

Monger, H.C., L.A. Daugherty, W.C. Lindemann, and C.M. Liddell. 1991. Microbial precipitation of pedogenic calcite. Geology 19:997–1000.

Monger, H.C., and R.A. Gallegos. 2000. Biotic and abiotic processes and rates of pedogenic carbonate accumulation in the Southwestern United States—Relationship to atmospheric CO_2 sequestration, p. 273–289. *In* R. Lal et al. (eds.) Global climate change and pedogenic carbonates. Lewis Publishers, New York.

Monger, H.C., and E.F. Kelly. 2002. Silica minerals, p. 611–636. *In* J.B. Dixon and D.G. Schluze (eds.) Soil mineralogy with environmental applications. SSSA Book Series No. 7. SSSA, Madison, WI.

Monger, H.C., and W.C. Lynn. 1996. Clay mineralogy of the Desert Project and Rincon soils. Soil Survey Investigations Report 44. USDA, Lincoln, NE.

Monger, H.C., and J.J. Martinez-Rios. 2001. Inorganic carbon sequestration in grazing lands, p. 87–118. *In* R.F. Follett et al. (eds.) The potential of U.S. grazing lands to sequester carbon and mitigate the greenhouse effect. Lewis Publishers, Boca Raton, FL.

Monger, H.C., and L.P. Wilding. 2002. Soil inorganic carbon: Composition and formation, p. 701–705. *In* R. Lal. (ed.) Encyclopedia of soil science. Marcel-Dekker, New York.

Mora, C.I., S.G. Driese, and L.A. Colarusso. 1996. Middle to late Paleozoic atmospheric CO_2 levels from soil carbonate and organic matter. Science 271:1105–1107.

Nettleton, W.D., K.W. Flach, and B.R. Brasher. 1969. Argillic horizons without clay skins. Soil Sci. Soc. Am. Proc. 33:121–125.

Nettleton, W.D., R.E. Nelson, B.R. Brasher, and P.S. Derr. 1982. Gypsiferous soils in the western United States, p. 147–168. *In* J.A. Kittrick, D.S. Fanning, and L.R. Hossner (eds.) Acid sulfate weathering. SSSA Special Publication No. 10. SSSA, Madison, WI.

Nettleton, W.D., and F.F. Peterson. 1983. Aridisols, p. 165–215. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy II. The soil orders. Elsevier, Amsterdam, the Netherlands.

Nettleton, W.D., J.E. Witty, R.E. Nelson, and J.W. Hawley. 1975. Genesis of argillic horizons in soils of desert areas of the southwestern United States. Soil Sci. Soc. Am. Proc. 39:919–926.

Nikiforoff, C.C. 1937. General trends of the desert type of soil formation. Soil Sci. 43:105–131.

Otterman, J., and V. Gornitz. 1983. Saltation vs. soil stabilization: Two processes determining the character of surfaces in arid regions. Catena 10:339–362.

Pal, D.K., P. Srivastava, and T. Bhattacharyya. 2003. Clay illuviation in calcareous soils of the semiarid part of the Indo-Gangetic Plains, India. Geoderma 115:177–192.

Pease, D.S. 1967. Opal phytoliths as an indicator of paleosols. M.S. Thesis. New Mexico State University. Las Cruces, NM.

Peterson, F.F. 1977. Dust infiltration as a soil forming process in deserts, p. 172. Agronomy Abstracts. ASA, Los Angeles, CA.

Peterson, F.F. 1980. Holocene desert soil formation under sodium salt influence in a playa-margin environment. Quat. Res. 13:172–186.

Philips, S.E., A.R. Milnes, and R.C. Forster. 1987. Calcified filaments: An example of biological influences in the formation of calcrete in South Australia. Aust. J. Soil Sci. 25:405–428.

Pillans, B. 2005. Geochronology of the Australian regolith, p. 41–61. *In* R.R. Anand and P. de Broekert (eds.) Regolith-landscape evolution across Australia. CRC LEME, Perth, Australia.

Prasse, R., and R. Bornkamm. 2000. Effect of microbiotic soil surface crusts on emergence of vascular plants. Plant Ecol. 150:65–75.

Pustovoytov, K.E. 2002. Pedogenic carbonate cutans on clasts in soils as a record of history of grassland ecosystems. Palaeogeogr. Palaeoclimatol. Palaeoecol. 177:199–214.

Quade, J., T.E. Cerling, and J.R. Bowman. 1989. Systematic variations in the carbon and oxygen isotopic composition of pedogenic carbonate along elevation transects in the southern Great Basin, United States. Geol. Soc. Am. Bull. 101:464–475.

Rabenhorst, M.C., and L.P. Wilding. 1986. Pedogenesis on the Edwards Plateau, Texas: II. Formation and occurrence of diagnostic horizons in a climosequence. Soil Sci. Soc. Am. J. 50:687–692.

Rango, A., S.L. Tartowski, A. Laliberte, J. Wainwright, and A. Parsons. 2006. Islands of hydrologically enhanced biotic productivity in natural and managed arid ecosystems. J. Arid Environ. 65:235–252.

Reheis, M.C. 1987. Gypsic soils on the Kane alluvial fans, Big Horn County, Wyoming. U.S. Geological Survey Bulletin 1590. U.S. Government Printing Office, Washington, DC.

Reheis, M.C. 1988. Pedogenic replacement of aluminosilicate grains by $CaCO_3$ in Ustollic Haplargids, south-central Montana, U.S.A. Geoderma 41:243–261.

Reheis, M.C., J.C. Goodmacher, J.W. Harden, L.D. McFadden, T.K. Rockwell, R.R. Shroba, J.M. Sowers, and E.M. Taylor. 1995. Quaternary soils and dust deposition in southern Nevada and California. Geol. Soc. Am. Bull. 107:1003–1022.

Reheis, M.C., J.M. Sowers, E.M. Taylor, L.D. McFadden, and J.W. Harden. 1992. Morphology and genesis of carbonate soils on the Kyle Canyon fan, Nevada, U.S.A. Geoderma 52:303–342.

Reid, D.A., R.C. Graham, R.J. Southard, and C. Amrhein. 1993. Slickspot soil genesis in the Carrizo Plain, California. Soil Sci. Soc. Am. J. 57:162–168.

Reynolds, R., J. Belnap, M. Reheis, P. Lamothe, and F. Luiszer. 2001. Aeolian dust in Colorado Plateau soils: Nutrient inputs and recent change in source. Proc. Natl. Acad. Sci. U. S. A. 98:7123–7127.

Reynolds, J.F., M.S. Smith, E.F. Lambin, B.L. Turner, II et al. 2007. Global desertification: Building a science for dryland development. Science 316:847–851.

Rode, A.A. 1962. Soil science (translated from Russian). Published for the National Science Foundation, Washington, DC by the Israel Program for Scientific Translation, Jerusalem.

Rosen, A.M., and S. Weiner. 1994. Identifying ancient irrigation: A new method using opaline phytoliths from emmer wheat. J. Archaeol. Sci. 21:125–132.

Royer, D.L. 1999. Depth to pedogenic carbonate horizon as a paleoprecipitation indicator? Geology 27:1123–1126.

Ruhe, R.V. 1962. Age of the Rio Grande valley in southern New Mexico. *Journal of Geology* 70:151–167.

Ruhe, R.V. 1967. Geomorphic surfaces and surficial deposits in southern New Mexico. Memoir 18. New Mexico Bureau of Mines and Mineral Resources, Socorro, NM.

Sage, R.F., D.A. Wedin, and M. Li. 1999. The biogeography of C_4 photosynthesis: Patterns and controlling factors, p. 313–373. *In* R.F. Sage and R.K. Monson (eds.) C_4 plant biology. Academic Press, San Diego, CA.

Schaetzl, R.J., and S. Anderson. 2005. Soils: Genesis and geomorphology. Cambridge University Press, Cambridge, U.K.

Schlesinger, W.H., and C.S. Jones. 1984. The comparative importance of overland runoff and mean annual rainfall to shrub communities of the Mojave desert. Bot. Gaz. 145:116–124.

Schlesinger, W.H., J.F. Reynolds, G.L. Cunningham, L.F. Huenneke, W.M. Jarrell, R.A. Virginia, and W.G. Whitford. 1990. Biological feedbacks in global desertification. Science 247:1043–1048.

Schlesinger, W.H., S.L. Tartowski, and S.M. Schmidt. 2006. Nutrient cycling within an arid ecosystem, p. 133–150. *In* K. Havstad, L.F. Huenneke, and W.H. Schlesinger (eds.) Structure and function of a Chihuahuan Desert ecosystem: The Jornada Basin long term ecological research site. Oxford University Press, Oxford, U.K.

Schlesinger, W.H., T.J. Ward, and J. Anderson. 2000. Nutrient losses in runoff from grassland and shrubland habitats in southern New Mexico: II. Field plots. Biogeochemistry 49:69–86.

Schmidt, R.H. 1979. A climatic delineation of the 'real' Chihuahuan Desert. J. Arid Environ. 2:243–250.

Serna-Perez, A., H.C. Monger, J.E. Herrick, and L. Murray. 2006. Carbon dioxide emissions from exhumed petrocalcic horizons. Soil Sci. Soc. Am. J. 70:795–805.

Shainberg, I., M.E. Sumner, W.P. Miller, M.P.W. Farina, M.A. Pavan, and M.V. Fey. 1989. Use of gypsum on soils: A review, p. 1–111. *In* B.A. Stewart (ed.) Advances in soil science. Vol. 13. Springer Verlag, New York.

Sibirtsev, N.M. 1898. Selected works. Vol. 1. Soil science. Issued in translation by the Israel program for scientific translation, Jerusalem, 1966, 354pp.

Simonson, R.W. 1959. Outline of a generalized theory of soil genesis. Soil Sci. Soc. Am. Proc. 23:152–156.

Singer, A. 1989. Palygorskite and sepiolite group minerals, p. 829–872. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Smale, D. 1973. Silcretes and associated silica diagenesis in southern Africa and Australia. J. Sediment. Petrol. 43:1077–1089.

Smith, G.D. 1986. The Guy Smith interviews: Rationale for concepts in soil taxonomy. Soil Management Support Services Technical Monograph No. 11. USDA, Washington, DC.

Smith, B.R., and S.W. Buol. 1968. Genesis and relative weathering studies in three semi-arid soils. Soil Sci. Soc. Am. Proc. 32:261–265.

Sobecki, T.M., D.L. Mofitt, J. Stone, C.D. Franks, and A.G. Mendenhall. 2001. A broad-scale perspective on the extent, distribution, and characteristics of U.S. grazing lands, p. 21–63. *In* R.F. Follett, J.M. Kimble, and R. Lal (eds.) The Potential of U.S. grazing lands to sequester carbon and mitigate the greenhouse effect. CRC Press, New York.

Sobecki, T.M., and L.P. Wilding. 1983. Formation of calcic and argillic horizons in selected soils of the Texas Coast Prairie. Soil Sci. Soc. Am. J. 47:707–715.

Soil Survey Division Staff. 1951. Soil survey manual. U.S. Department of agriculture handbook no. 18. USDA-SCS. U.S. Government Printing Office, Washington, DC.

Soil Survey Division Staff. 1993. Soil survey manual. USDA agriculture handbook no. 18. USDA-SCS. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1975. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. Agriculture handbook no. 436. USDA-SCS. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1996. Soil survey laboratory methods manual. Soil Survey Investigations Report No. 42. USDA-NRCS, National Soil Survey Center, Lincoln, NE.

Soil Survey Staff. 1999. Soil taxonomy—A basic system of soil classification for making and interpreting soil surveys. USDA agriculture handbook no. 436. USDA-SCS. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS. U.S. Government Printing Office, Washington, DC.

Southard, R.J. 2000. Aridisols, p. E321–E338. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Southard, R.J., J.L. Boettinger, and O.A. Chadwick. 1990. Identification, genesis, and classification of duripans, p. 45–60. *In* J.M. Kimble and W.D. Nettleton (eds.) Proc. 4th Int. Soil Correl. Meet. (ISCOM): Charact. classif. util. aridisols. Part A: Papers. USDA-SCS, Lincoln, NE.

Springer, M.E. 1958. Desert pavement and vesicular layer of some soils of the desert of the Lahontan Basin, Nevada. Soil Sci. Soc. Am. Proc. 22:63–66.

Strahler, A.N., and A.H. Strahler. 1987. Modern physical geography. Wiley, New York.

Summerfield, M.A. 1983. Silcrete, p. 59–91. *In* A.S. Goudie and K. Pye (eds.) Chemical sediments and geomorphology. Academic Press Ltd., New York.

Swineford, A., J.C. Frye, and A.B. Leonard. 1958. Petrology of the Pliocene pisolitic limestone in the Great Plains. Kansas Geol. Surv. Bull. 130:97–116.

Taylor, E.M. 1986. Impact of time and climate on Quaternary soils in the Yucca Mountain area of the Nevada Test Site. M.S. Thesis. University of Colorado. Boulder, CO.

Thorp, J. 1949. Effects of certain animals that live in soils. Sci. Monit. 68:180–190.

Tongway, D.J., C. Valentin, and J. Seghieri (eds.) 2001. Banded vegetation patterns in arid and semiarid environments: Ecological processes and consequences for management. Ecological Studies 149. Springer, New York.

Turk, J.K., and R. Graham. 2009. Vesicular pore formation: The effect of microbes and soil temperature. ASA–CSSA–SSSA annual meeting abstracts. ASA–CSSA–SSSA, Madison, WI.

USDA-NRCS. 1997a. Soils data bases. Survey Section, Statistical Laboratory, Iowa State University, Ames, IA.

USDA-NRCS. 1997b. Soil Survey Laboratory database. National Soil Survey Center, Lincoln, NE.

Valentine, G.A., and C.D. Harrington. 2006. Clast size controls and longevity of Pleistocene desert pavements at Lathrop Wells and Red Cone volcanoes, southern Nevada. Geology 34:533–536.

Vanden Heuvel, R.C. 1966. The occurrence of sepiolite and attapulgite in the calcareous zone of a soil near Las Cruces, New Mexico, p. 193–207. *In* Proc. 13th Natl. Conf. Clays Clay Miner. Clays and Clay Minerals. Pergamon Press, New York.

Verrecchia, E., A. Yair, G.J. Kidron, and K. Verrecchia. 1995. Physical properties of the psammophile cryptogamic crust and their consequences to the water balance of sandy soils, Northwestern Negev Desert, Israel. J. Arid. Environ. 29:427–437.

Verstraete, M.M., R.J. Scholes, and M.S. Smith. 2009. Climate and desertification: Looking at an old problem through new lenses. Front. Ecol. Environ. 7:421–428.

West, L.T., L.R. Drees, L.P. Wilding, and M.C. Rabenhorst. 1988. Differentiation of pedogenic and lithogenic carbonate forms in Texas. Geoderma 43:271–287.

Wilcox, B.P., and T.L. Thurow. 2006. Emerging issues in rangeland ecohydrology: Vegetation change and the water cycle. Rangeland Ecol. Manage. 59:220–224.

Wilding, L.P. 2000. Introduction: General characteristics of soil orders and global distribution, p. E175–E182. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Ration, FL.

Wilding, L.P., and L.R. Drees. 1974. Contributions of forest opal and associated crystalline phases to fine silt and clay fractions of soils. Clays Clay Miner. 22:295–306.

Wilding, L.P., N.E. Smeck, and L.R. Drees. 1977. Silica in soils: Quartz, cristobalite, tridymite, and opal, p. 471–552. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. SSSA, Madison, WI.

Withgott, J., and S. Brennan. 2009. Essential environment: The science behind the stories. Pearson, San Francisco, CA.

Witty, J.E. 1990. The orthids past, present and future classification, p. 93–98. *In* J.M. Kimble and W.D. Nettleton (eds.) Proc. 4th Int. Soil Correl. Meet. (ISCOM): Charact. classif. util. aridisols. Part A: Papers. USDA-SCS, Lincoln, NE.

Yaalon, D.H. 1983. Climate, time and soil development, p. 233–251. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy I. Concepts and interactions. Elsevier, Amsterdam, the Netherlands.

Yaalon, D.H., and E. Ganor. 1973. The influence of dust on soils during the Quaternary. Soil Sci. 116:146–155.

Zhi, D., X. Ding, W. Nan, and H. Li. 2009. Nematodes as an indicator of biological crust development in the Tengger Desert, China. Arid Land Res. Manage. 23:223–236.

33.11 Alfisols

C.T. Hallmark

D.P. Franzmeier

33.11.1 Introduction

The central concept of Alfisols embraces soils of the semiarid to humid climates with light-colored surfaces and subsoils moderately rich in basic cations, formed, at least in part, by movement of clay from overlying horizons into the subsoil. Most Alfisols developed under deciduous forest vegetation in humid climates and grass or savannah vegetation in drier regimes. In loamy and silty Alfisols of the humid regions, fragipans are not uncommon; in drier climates, incomplete leaching of salts and carbonates can result in concentrations of secondary carbonates, gypsum, amorphous silica, or enrichment of exchangeable Na. As the processes of formation of Alfisols require both time and energy, most are found on geomorphic surfaces that are of Pleistocene age or older. They form on a wide variety of parent materials.

In previous classification systems in the United States, most Alfisols were in the Gray–Brown Podzolic great group, but others were in the Low-Humic Gley, Noncalcic Brown, Reddish-Brown Lateritic, Reddish Brown, Reddish Chestnut, Gray Wooded and Planosol great groups. Further, natric great groups or subgroups of Alfisols also include salt-affected soils that were previously classified as Solonetz and Soloth soils (Soil Survey Staff, 1975).

33.11.2 Distribution

Alfisols occur on all continents except Antarctica (Figure 33.72). They occupy about 9.6% of the ice-free Earth's surface (Table 33.1). They are found between the latitudes of 65°N and 45°S and concentrated in the Northern Hemisphere, specifically North America, Europe, and east Central Asia between latitudes of 30°N and 65°N; also, significant occurrences are found in sub-Saharan Africa, India, South America, and Australia. Cryalfs are found within the colder reaches such as Northern Europe and into Western Russia. Ustalfs are common in sub-Saharan Africa, Eastern Brazil, the eastern portion of India, and Southeastern Australia. Udalfs are prominent in Central Europe and Eastern Australia, while Xeralfs are found in countries bordering the Mediterranean Sea.

Within the United States, Cryalfs are found in the north central states and at higher elevations in the western states. Udalfs are prominent in the eastern portion of the Midwest, along the Mississippi and Ohio River valleys while Ustalfs occur to the south and west of the Great Plains as well as the southern portion of the High Plains. Xeralfs are restricted to the west coast states.
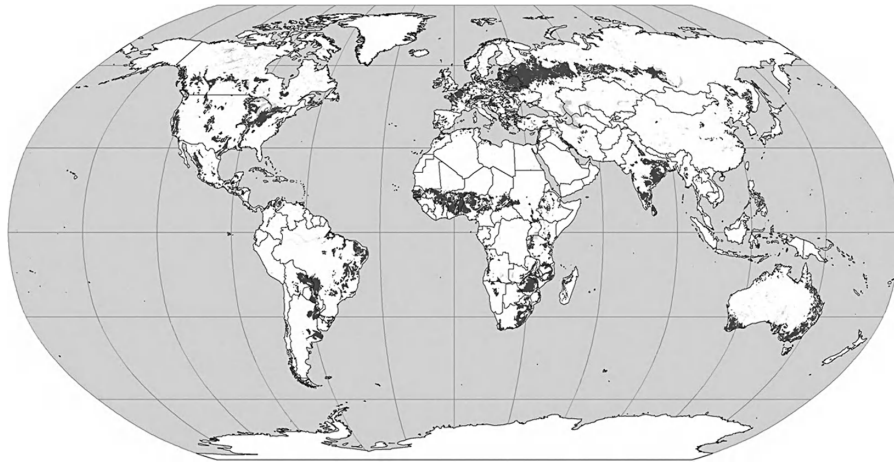
FIGURE 33.72 Global distribution of Alfisols. (Courtesy of USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC, 2010.)

33.11.3 Factors of Soil Formation

33.11.3.1 Parent Material

Alfisols form on a wide variety of parent materials, which includes glacial till (Borchardt et al., 1968; Smeck et al., 1968; Smith and Wilding, 1972; Ranney et al., 1975; Franzmeier et al., 1985; Weisenborn and Schaetzl, 2005), loess (Bouma et al., 1968; Grossman and Fehrenbacher, 1971), eolian sands (Davis, 1970; Allen et al., 1972; Miles and Franzmeier, 1981), residuum from limestone (Reynders, 1972), sandstone (Anderson et al., 1975; Stahnke et al., 1983; Chittleborough, 1989; Delgado et al., 1994), siltstones and shales (Gallez et al., 1975; Delgado et al., 1994), acid igneous and metamorphic rock (Goss and Allen, 1968; Ranney and Beatty, 1969; Yerima et al., 1989; Kooistra et al., 1990), mafic igneous and metamorphic rock (Gallez et al., 1975; Bhattacharyya et al., 1993; Juo and Wilding, 1996), lacustrine sediments (Borchardt et al., 1968; Ranney and Beatty, 1969), colluvium derived from acid shales and siltstones (Bailey and Avers, 1971), basalt (Weitkamp et al., 1996), coastal plain sediments (Vepraskas and Wilding, 1983; Juo and Wilding, 1996), and alluvium (Parsons et al., 1968; Ranney et al., 1975; Chittleborough et al., 1984; Klich et al., 1990; Delgado et al., 1994; Shaw et al., 2003; Calero et al., 2008). The literature is replete with examples of Alfisols developing from two or more parent materials, for example, loess over till (Allan and Hole, 1968; Foss and Rust, 1968; Ransom et al., 1987), loess over basalt (Blank and Fosberg, 1991), loess over shale (Ranney et al., 1975), loess over sandstone or limestone (Aide and Marshaus, 2002) and colluvium over residuum (Weitkamp et al., 1996). Uniformity of parent material also continues to be a question in evaluating the influence of lithogenic inherited vs. pedogenic and translocated clay in the argillic horizon (Ruhe, 1984a; Schaetzl, 1998). In eolian sands, the argillic (zone of clay accumulation) horizon usually consists of a sequence of thin bands called lamellae that contain somewhat more clay than the soil material between the bands.

Lotspeich and Coover (1962) stated that "texture of the parent material controls the texture of the soil because texture is a nearly permanent characteristic of soil," particularly in arid and semiarid regions. While this may be true in Alfisols for which the parent materials have undergone previous cycles of weathering and are rich in quartz and resistant minerals (Anderson et al., 1975), many examples are available where weathering gives rise to textures and a pedogenic mineral suite markedly different from the parent material (Paeth et al., 1971; Gallez et al., 1975).

33.11.3.2 Organisms

Most Alfisols formed primarily under hardwood forest vegetation (Soil Survey Staff, 1999), but some suborders and great groups support a significant component of grass vegetation. This is illustrated in Figure 33.73 where soil classes with abbreviated vegetative descriptions from the literature are arranged in a temperature–moisture regime matrix. Aqualfs and Udalfs support forest vegetation, although dominant species and composition of forests change in response to climatic conditions. Ustalfs and Xeralfs support moderate to dominant grassland vegetation, commonly described as savanna. The forest vegetation of moister Cryalfs yields to a greater composition of grasses as Cryalfs occupy dryer climates. It is hazardous to assume that present or presettlement vegetation (Figure 33.73) represents the native vegetation of the soil as both evolve together with climate changes over time modifying vegetative communities (Wells, 1970). Even within broad vegetative zones, individual species may influence soil properties over relatively short distances. Gersper and Holowaychuk (1970a, 1970b) showed precipitation stemflow down the trunk of American beech (*Fagus grandifolia*) trees thickened E (A2) horizons, increased low chroma mottling of B horizons, and decreased clay content, pH, exchangeable Ca, Mg, and K, CEC, base saturation, and free Fe oxides in the B horizons near the stem. Davis et al. (1995) believe available P that was correlated with microelevational changes on a cultivated Psammentic Paleustalf in Niger is related to location of bushes and trees in the savanna or to cultivation.

Fauna dwelling in Alfisols has also been credited with affecting soil formation. Nielsen and Hole (1964) found earthworm population masses of 2,218 kg ha^{-1} under a deciduous forest
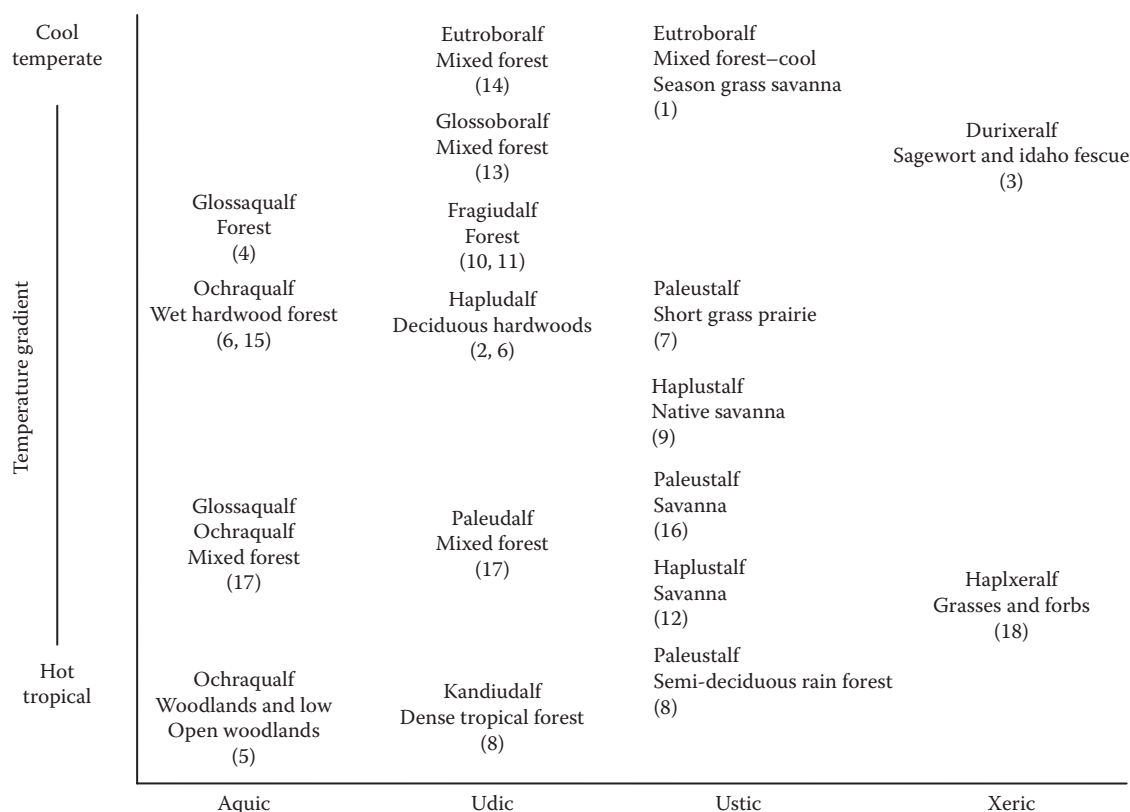
FIGURE 33.73 Generalized native vegetation for selected Alfisols as a function of temperature and moisture regime. Sources are as follows: (1) Anderson et al. (1975); (2) Bailey and Avers (1971); (3) Blank and Fosberg (1991); (4) Bouma and van Schuylenborgh (1969); (5) Coventry and Williams (1984); (6) Cremeens and Mokma (1986); (7) Davis (1970); (8) Gallez et al. (1975); (9) Goss and Allen (1968); (10) Lozet and Herbillon (1971); (11) Miller et al. (1971); (12) Moberg and Esu (1991); (13) Ranney and Beatty (1969); (14) Schaetzl (1996); (15) Smith and Wilding (1972); (16) Stahnke et al. (1983); (17) Vepraskas and Wilding (1983); (18) Weitkamp et al. (1996).

and earthworm middens numbering 274,000 ha⁻¹. Essentially, all annual leaf fall was moved to the middens by earthworms. Wiecek and Messenger (1972) described calcite spheroids in acid A horizons of Hapludalfs under forest vegetation that are produced in earthworm calciferous glands and were responsible for close-range variability in soil pH of A horizons. Hugie and Passey (1963) describe burrowing activity of western cicadas, crediting them with development of cylindrical structure as they burrowed through loamy soils. Their numbers were greatly reduced on coarse- and fine-textured materials. Ant, termite, and crustacean activity have also been described (Thorp, 1949). Crayfish activity in Ochraqualfs (formally) of the Texas coast has been so intense that up to 75% of the subsoil volume comprise their krotovinas prompting the establishment of a new great group (Vermaqualfs) to accommodate these and similar soils highly modified by fauna (Vepraskas and Wilding, 1983). Arkley and Brown (1954) observed sufficient rodent activity to suggest that mima mounds in western states are the result of pocket gophers.

33.11.3.3 Climate

In the process of clay eluviation, percolating water has mobilized silicate clays in upper horizons, carried them downward in the pedon with final deposition in subsoil horizons. This occurs most readily in climates in which precipitation exceeds evapotranspiration for at least several months each year. Consequently, Alfisols occur in semiarid to humid climates. In temperate and cool climates, Alfisols tend to form a belt between the Mollisols of the grasslands and the Spodosols and Inceptisols of very humid climates (Soil Survey Staff, 1999). In warmer climates, they often lie between the Aridisols of arid regions and the Inceptisols, Ultisols, and Oxisols of warm, humid climates.

Within the Alfisol areas, several soil properties vary with climate. Early works of Jenny (1935) and Jenny and Leonard (1934) serve to broadly illustrate soil changes over long transects where gradients of temperature and precipitation exist. When crystalline rocks of similar composition and geomorphic age weather across a temperature gradient, soil clay content in the upper solum increases with mean annual temperature. Along the 11°C annual isotherm from eastern Colorado to western Missouri, they found that with increased annual precipitation, depth to free carbonates, N and clay contents increased, and pH decreased. Partitioning climatic influence over long distances is difficult at best. Climate implies both temporal and spatial fluxes in a number of parameters including temperature, precipitation, radiation, humidity, and wind speed and direction. Locally, relief, slope, and aspect can provide a microclimate somewhat

dissimilar to the broader regional climate. Cooper (1960) illustrated the role of slope aspect on microclimate in Michigan Udalfs where he found south-facing slopes had higher light intensities, maximum air temperature, evaporation rate, and soil temperature and lower soil moisture than north-facing slopes. The B horizons of soils on the southern aspect were redder and finer in texture, and sola were thinner than those of the northern exposure. Torrent and Nettleton (1979) proposed a textural index (based upon fine silt:total silt ratios) to assess chemical weathering. Using soils, primarily Udalfs and Aqualfs developed from loess along the Mississippi Valley from Minnesota to Mississippi, they related the textural index describing weathering of silts to clays with increasing mean annual temperature. Subsequently, Ruhe (1984a), utilizing a smaller data set of Hapludalfs along a similar Minnesota to Mississippi transect, interpreted the trends as sedimentation dependent. In a companion study on the same Hapludalfs, Ruhe (1984b) related the more intensive and deeper leaching of bases in the southern sector to both historic and paleoclimate differences in annual and effective precipitation.

33.11.3.4 Relief

Alfisols occur mostly on hillslopes that are of moderate gradient and are linear or convex in cross section. On steeper slopes, clay translocation is limited because the soil is very shallow or stony or because material moves downslope, resulting in Inceptisols or Entisols. In depressional areas, downward water movement might be limited, or soil organic matter (SOM) content of upper horizons is high enough for the soils to be Mollisols.

Relief is responsible for the distribution of Aqualfs in many settings where water moves to lower topographic positions either overland or by throughflow. In these lower positions, epiaquic (perched water table) or endoaquic conditions may occur. Cremeens and Mokma (1986) described soil hydrosequences in Michigan where well- and moderately well-drained Hapludalfs yielded to somewhat poorly drained Ochraqualfs (later reclassified to Endoaqualfs) and poorly drained Haplaquepts and Argiaquolls. Argillans (clay films) and parameters related to movement of clay from eluvial to illuvial zones were greatest in the better drained soils. Working with seasonally saturated Alfisols in a toposequence on the Texas Coastal Plain, Vepraskas and Wilding (1983) found soil pH generally increased as did base saturation and exchangeable Na levels from upper to lower topographic positions. Also, smectite was the predominant clay mineral in the upper sola of the soils in the lower positions while kaolinite was dominant in higher positions. Cremeens and Mokma (1986) also found smectite in the wettest sites of their study with its absence in the better drained soils. Goss and Allen (1968) found appreciable accumulation of exchangeable Na downslope in a Natrustalf and evidence of accumulation of smectite due to relief.

33.11.3.5 Time

Alfisols in areas with moderate and cool temperatures are mostly on late Pleistocene surfaces as most of these surfaces were affected by glaciation. In warmer areas, they may be on older surfaces. Usually, there has not been enough clay translocation for the soils to be classified as Alfisols on surfaces younger than late Pleistocene. On old surfaces, leaching of bases commonly has progressed to the point where the soils are classified as Ultisols instead of Alfisols.

The length of time for development of soil and specific soil properties has long been of scientific interest (Jenny, 1941). In recent years, interest in soil age has been increasing as evidenced by the growing volume of literature (Bilzi and Ciolkosz, 1977; Bockheim, 1980; Little and Ward, 1981; Meixner and Singer, 1981; Harden, 1982; Muhs, 1982; Chittleborough et al., 1984; Catt, 1991; Markewich and Pavich, 1991; Delgado et al., 1994; Shaw et al., 2003; Calero et al., 2008). Part of this renewed interest is due to application of soil interpretation to subjects of current interest such as climate change, C sequestration rates, landscape stability, and archaeology. Improved and new methodologies to directly determine soil age such as radiocarbon assay, luminescence, amino acid analysis, paleomagnetic measurement, and U–Th disequilibrium series have augmented dendrochronology and recorded historic events (mudflows, volcanic eruptions, etc.) and indirect methods of correlation of stratigraphic units. As soil (and landscape) age increases, so does the concern for significant changes in other factors of state, particularly climate and its effect on vegetation. The time required for the development of Alfisols is incumbent upon the formation of argillic and kandic horizons. Bilzi and Ciolkosz (1977) believe that 2,000–12,000 years are required to develop fragipan-like features and argillic horizons in alluvium in Pennsylvania under udic soil moisture and mesic soil temperature regimes. In northeast Iowa on earthen mounds of varying age built by Indians, Parsons et al. (1962) found that soil genesis had taken place with A horizon properties remaining relatively constant with increasing age beyond 1000 years, E horizon platy structure developing within 1000 years (maximum expression required >2500 years) and the blocky structure plus clay translocation sufficient to meet argillic horizon definitions in the B horizon occurring between 1000 and 2500 year BP. In comparison, the nearby Fayette soil (Typic Hapludalf) had developed over a 14,000 year period. In preconditioned material used by Indians to construct mounds and ridges in Louisiana, argillic horizons in Ultisols and Alfisols had developed in 5000–5400 years (Saunders et al., 1997). Shaw et al. (2003) found Alfisols on stream terrace surfaces dating to 5500 year BP in Alabama.

In the Midwest, most Alfisols are found on surfaces affected by late Wisconsinian glaciation (12,000–25,000 year BP). In the Upper Midwest, many Alfisols formed in late Wisconsinian till deposited by glacier ice, and in outwash and lacustrine deposits from the glacier when it melted. Also, much of the Midwest and the lower Mississippi Valley was blanketed with late Wisconsinian age loess that was calcareous when it was deposited. Soils formed in it are relatively high in basic cations. This loess also covered surfaces on which soils had formed previously and had been eroded. These buried soils (paleosols) were enriched with basic cations (especially Ca and Mg) that were leached from the loess and translocated downward.

TABLE 33.49   Ages of Ustalfs and Xeralfs Included in Chronosequence Studies

Soil Class	Age, Years BP	MAP (mm)	Location	Source
Typic Haploxeralf	250,000	310	California	Meixner and Singer (1981)
Typic Haploxeralf	40,000	410		Harden (1982)
Typic Haploxeralf	130,000			
Typic Haploxeralf	600,000			
Typic Natrixeralf	60–107,000	170		Muhs (1982)
Typic Natrixeralf	120–134,000			
Typic Natrixeralf	375–460,000			
Ultic Haplustalf	3,740	420	New South Wales, Australia	Chittleborough et al. (1984)
Rhodic Paleustalf	9,220			
Ultic Paleustalf	26,700			
Ultic Paleustalf	27,040			
Ultic Paleustalf	29,000			
Paleustalf	42,500	—	Victoria, Australia	Little and Ward (1981)
Paleustalf	210,000			
Paleustalf	227,000			
Paleustalf	763,000			

All soils developed from alluvium. Dating methods in the studies include amino acid, U-series, uplift rates, and stratigraphic correlations.

Alfisols of ustic and xeric soil moisture regimes and developed from alluvium have been included in a number of chronosequence studies (Table 33.49). The annual precipitation ranged from 170 to 420 mm. The youngest Alfisol (Ultic Haplustalf) was 3,740 years BP and the oldest (Paleustalf) was 763,000 years BP, while the youngest Paleustalf was 9,220 years BP. Much older Haploxeralfs were noted in the California studies, but in the xeric moisture regime sufficient development of the argillic horizon to permit classification as a Palexeralf had not occurred.

33.11.4 Morphology

The morphology of Alfisols ranges greatly across the broad spectrum of conditions under which they occur. In all Alfisols, however, there is an increase in clay content from the A and E horizons near the surface to the B horizons below. In most Alfisols, this clay increase meets the concept of an argillic horizon where the illuvial Bt (t from ton Gr, clay) horizon shows evidence of oriented clays. In older soils dominated by low activity clays (low CEC of the clay fraction), evidence of oriented clay may be absent but clay increase is sufficient for a kandic horizon.

Where some of the clay in the Bt horizon is oriented, that is, stacked like a pile of papers or playing cards, the oriented clay may be on ped surfaces where it forms clay films (thin layers of oriented clay) that often cover and may be a different color from sand grains in the ped (see Chapter 30 for electron micrograph of a clay film). In sandy soils, the oriented clay forms bridges between adjacent sand grains and may line root channels. Often, this clay mentioned above actually contains a fair amount of silt. In many Alfisols, the transitional EB or BE horizons have silt coatings instead of clay on the outside of peds. In these, and some other Alfisols, especially those formed in calcareous parent materials, the best developed clay films are in the lower Bt horizon. Apparently, clay is stripped from the ped coatings in the upper B or BE horizon and is deposited on ped surfaces deeper in the profile, illustrating that clay translocation is a dynamic process, and that the argillic horizon is moving down in the profile.

Before a soil can be classified in *Soil Taxonomy*, germane diagnostic horizons and soil properties (defined in Chapter 31) must be identified. Several are essential or common in Alfisols, and their occurrence in the various suborders of Alfisols is shown in Table 33.50.

TABLE 33.50   Occurrence of Soil Horizons or Properties in Soils of the Suborders of Alfisols

	Soil Horizon or Property									
Suborder	Argillic Horizon	Natric Horizon	Kandic Horizon	Glossic Horizon	Fragipan	Duripan	Plinthite	Very Red Argillic	Very Deep Argillic	Other
Aqualfs	X	X	X	X	X	X	X			Biotic activity, very cold[a]
Cryalfs	X			X					X	
Udalfs	X	X	X	X	X			X	X	Fe nodules
Ustalfs	X	X	X			X	X	X	X	Calcic/Petrocalcic
Xeralfs	X	X			X	X	X	X	X	Calcic/Petrocalcic

[a] Also: abrupt clay increase at upper boundary of the argillic horizon, and perched water table.

33.11.5 Suborders: Their Morphology, Distribution, and Use

The Alfisol order is subdivided into five suborders, one based on temperature and four based on soil moisture regime. Cryalfs are the Alfisols in cold climates. Of the four suborders based on soil moisture regime, three (Udalfs, Ustalfs, and Xeralfs) follow broad geographical climatic belts. The distribution of Aqualfs depends on local landscape and parent material factors. A complete listing of the great groups and subgroups in each suborder appears in Table 33.51.

33.11.5.1 Aqualfs

These are the wet Alfisols in which the water table is at or near the surface (50 cm) for at least several weeks each year, but may be very deep during the dry season. The water table may be continuous to the aquifer, or it may be held up by a layer with low permeability in or just beneath the soil surface. Their location in the landscape is controlled mainly by local conditions of parent material and relief. They occur as small areas in larger areas of Cryalfs, Udalfs, Ustalfs, Xeralfs or other soil suborders, but they are more common in zones of Udalfs than Ustalfs. They may have several kinds of diagnostic horizons (Table 33.50). Nearly all Aqualfs had forest vegetation at some time in the past (Soil Survey Staff, 1999). Most are cultivated after they were drained by underground tiles or ditches. Rice is a common crop in warmer regions.

Aqualfs display quantifiable redoximorphic features (gray colors) immediately below a depth of 25 cm from the mineral soil surface. In the absence of quantifiable redoximorphic features, a positive reaction to active Fe^{2+} (α,α-dipyridyl) can be used. In this process (gleying), soil horizons are saturated with water, microbes deplete available O_2, causing Fe(III) in brownish and reddish Fe oxide minerals to be reduced. The resulting Fe^{2+} is released into the soil solution, much of which is leached from the soil. The result is that the horizon, or much of it, takes on the gray color of the silicate minerals. Gleyed horizons are designated with a g as Btg.

Gleying in Aqualfs may be noted in the upper 25 cm of the soil. Gleyed soils are generally found on lower topographic positions receiving additional water, or on broad, level interfluves where surface drainage may be very slow. The A horizon, commonly overlain by O horizon material in varying stages of decomposition, is usually either too thin or too low in organic C to meet requirements of a mollic epipedon. A few may qualify as an umbric or mollic epipedon, but have insufficient bases in the subsoil to be classified as Mollisols. Most Aqualfs have well-developed E horizons, although the underlying Bt and Btg horizons often show only minimally expressed clay films. Boundaries between E and Bt and Btg horizons often show evidence of mixing by fauna (i.e., crayfish) or tongues and stringers of E horizon material between structural units of Bt and Btg material (glossic material).

33.11.5.2 Cryalfs

Cryalfs are restricted to frigid and cryic temperature regimes and generally occur in more mountainous and sloping landscapes that are well drained. They formed in North America,

eastern Europe, and Asia above 49°N latitude and in some high mountains of lower latitude where they occur at lower altitudes than Spodosols or Inceptisols. Most have been under coniferous forest because of the cool, short growing season. Commonly, parent materials are residuum or colluvium from local bedrock. Most are not cultivated, and thus, have thin to moderately thick (2–10 cm) O horizons in varying stages of organic matter decomposition. Many lack A horizons, but when present, they are thin (5 cm) with granular or subangular blocky structure. Most have E horizons of platy or subangular blocky structure that are relatively thick (15–30 cm). The Bt horizons range from thin to thick (15–135 cm), display a wide range of clay film features, and commonly have angular or subangular blocky structure. Many have lithic or paralithic contacts, and most contain significant quantities of coarse fragments throughout the solum.

33.11.5.3 Udalfs

These are the freely drained Alfisols of humid regions, extensive in the United States and Europe. They formed under forest vegetation at some time during their development. The forests were mainly deciduous, but in colder regions, may have been mixed coniferous and deciduous. Many Udalfs have been cleared and intensively farmed; in some that have been severely eroded, the Bt horizon is immediately below or may be incorporated into the Ap horizon (Soil Survey Staff, 1999). Many have fragipans, and some have natric and other diagnostic horizons (Table 33.50). Corn and soybeans are common crops on Udalfs.

In well-drained, temperate humid environs under forest, uncultivated Udalfs have thin (2–5 cm) O horizons with organic matter in various stages of decomposition. Most also have a thin (10–15 cm) A horizon commonly of granular structure darkened by organic matter and underlain by a thin (10–15 cm), light-colored E horizon with platy structure, a zone of maximum eluviation of silicate clays. Because many Alfisols are cultivated, the A and E horizons have been mixed by plowing producing a light color. The intense eluviation in E horizons results in a concentration of sand and silt, and in extreme cases, all or most coloring compounds have been removed leaving skeletal grains whose color dominates the horizon. Often there is a transitional EB or BE horizon between the E and Bt horizons. As compared with the thinner A and E horizons, the Bt horizon is thick, possesses subangular or angular blocky structure, and contains significantly more silicate clay. At least a portion of the silicate clay occurs as clay films on structural or pore surfaces. In the Midwest where most of the parent material is calcareous and associated with the last glaciation, complete leaching of carbonates from the solum has occurred while C horizons remain calcareous.

33.11.5.4 Ustalfs

These are the Alfisols of subhumid to semiarid regions. Moisture moves through these soils to deeper layers only in occasional years (Soil Survey Staff, 1999). If there are carbonates in the parent material or in dust, the soil may have a carbonate-enriched horizon (calcic) below the argillic. Original vegetation may have been xerophytic trees, savanna, or grassland. Dryland crops for

TABLE 33.51 Listing of Suborders, Great Groups, and Subgroups in the Alfisols Order (Ref. Soil Survey Staff, 2010)

Suborder	Great Group	Subgroup
Aqualfs	Cryaqualfs	Typic
	Plinthaqualfs	Typic
	Duraqualfs	Typic
	Natraqualfs	Vertic, Vermic, Albic Glossic, Albic, Glossic, Mollic, Typic
	Fragiaqualfs	Vermic, Aeric, Plinthic, Humic, Typic
	Kandiaqualfs	Arenic, Grossarenic, Plinthic, Aeric Umbric, Aeric, Umbric, Typic
	Vermaqualfs	Natric, Typic
	Albaqualfs	Arenic, Aeric Vertic, Chromic Vertic, Vertic, Udollic, Aeric, Aquandic, Mollic, Umbric, Typic
	Glossaqualfs	Histic, Arenic, Aeric Fragic, Fragic, Aeric, Mollic, Typic
	Epiaqualfs	Aeric Chromic Vertic, Aeric Vertic, Chromic Vertic, Vertic, Aquandic, Aeric Fragic, Fragic, Arenic, Grossarenic, Aeric Umbric, Udollic, Aeric, Mollic, Umbric, Typic
	Endoaqualfs	Aquandic, Chromic Vertic, Vertic, Aeric Fragic, Fragic, Arenic, Grossarenic, Udollic, Aeric Umbric, Aeric, Mollic, Umbric, Typic
Cryalfs	Palecryalfs	Andic, Vitrandic, Aquic, Oxyaquic, Xeric, Ustic, Mollic, Umbric, Typic
	Glossocryalfs	Lithic, Vertic, Andic, Vitrandic, Aquic, Oxyaquic, Fragic, Xerollic, Umbric Xeric, Ustollic, Xeric, Ustic, Mollic, Umbric, Eutric, Typic
	Haplocryalfs	Lithic, Vertic, Andic, Vitrandic, Aquic, Oxyaquic, Lamellic, Psammentic, Inceptic, Xerollic, Umbric Xeric, Ustollic, Xeric, Ustic, Mollic, Umbric, Eutric, Typic
Ustalfs	Durustalfs	Typic
	Plinthustalfs	Typic
	Natrustalfs	Salidic, Leptic Torrertic, Torrertic, Aquertic, Aridic Leptic, Vertic, Aquic Arenic, Aquic, Arenic, Petrocalcic, Leptic, Haplargidic, Aridic Glossic, Aridic, Mollic, Typic
	Kandiustalfs	Grossarenic, Aquic Arenic, Plinthic, Aquic, Arenic Aridic, Arenic, Aridic, Udic, Rhodic, Typic
	Kanhaplustalfs	Lithic, Aquic, Aridic, Udic, Rhodic, Typic
	Paleustalfs	Aquertic, Oxyaquic Vertic, Udertic, Vertic, Aquic Arenic, Aquic, Oxyaquic, Lamellic, Psammentic, Arenic Aridic, Grossarenic, Arenic, Plinthic, Petrocalcic, Calcidic, Aridic, Kandic, Rhodic, Ultic, Udic, Typic
	Rhodustalfs	Lithic, Kanhaplic, Udic, Typic
	Haplustalfs	Lithic, Aquertic, Oxyaquic Vertic, Torrertic, Udertic, Vertic, Aquic Arenic, Aquultic, Aquic, Oxyaquic, Vitrandic, Lamellic, Psammentic, Arenic Aridic, Arenic, Calcidic, Aridic, Kanhaplic, Inceptic, Calcic Udic, Ultic, Calcic, Udic, Typic
Xeralfs	Durixeralfs	Natric, Vertic, Aquic, Abruptic Haplic, Abruptic, Haplic, Typic
	Natrixeralfs	Vertic, Aquic, Typic
	Fragixeralfs	Andic, Vitrandic, Mollic, Aquic, Inceptic, Typic
	Plinthoxeralfs	Typic
	Rhodoxeralfs	Lithic, Vertic, Petrocalcic, Calcic, Inceptic, Typic
	Palexeralfs	Vertic, Aquandic, Andic, Vitrandic, Fragiaquic, Aquic, Petrocalcic, Lamellic, Psammentic, Arenic, Natric, Fragic, Calcic, Plinthic, Ultic, Haplic, Mollic, Typic
	Haploxeralfs	Lithic Mollic, Lithic Ruptic-Inceptic, Lithic, Vertic, Aquandic, Andic, Vitrandic, Fragiaquic, Aquultic, Aquic, Natric, Fragic, Lamellic, Psammentic, Plinthic, Calcic, Inceptic, Ultic, Mollic, Typic
Udalfs	Natrudalfs	Vertic, Glossaquic, Aquic, Typic
	Ferrudalfs	Aquic, Typic
	Fraglossudalfs	Andic, Vitrandic, Aquic, Oxyaquic, Typic
	Fragiudalfs	Andic, Vitrandic, Aquic, Oxyaquic, Typic
	Kandiudalfs	Plinthaquic, Aquic, Oxyaquic, Arenic Plinthic, Grossarenic Plinthic, Arenic, Grossarenic, Plinthic, Rhodic, Mollic, Typic
	Kanhapludalfs	Lithic, Aquic, Oxyaquic, Rhodic, Typic
	Paleudalfs	Vertic, Andic, Vitrandic, Anthraquic, Fragiaquic, Plinthaquic, Glossaquic, Albaquic, Aquic, Oxyaquic, Fragic, Arenic Plinthic, Grossarenic Plinthic, Lamellic, Psammentic, Arenic, Grossarenic, Plinthic, Glossic, Rhodic, Mollic, Typic
	Rhodudalfs	Typic
	Glossudalfs	Aquertic, Oxyaquic Vertic, Vertic, Aquandic, Andic, Vitrandic, Fragiaquic, Aquic Arenic, Aquic, Arenic Oxyaquic, Oxyaquic, Fragic, Arenic, Haplic, Typic
	Hapludalfs	Lithic, Aquertic Chromic, Aquertic, Oxyaquic Vertic, Chromic Vertic

Ustalfs include drought tolerant selections such as sorghum, wheat, cotton, and millet. Ustalfs, which are common in the southern Great Plains, do not have fragipans, but may have several other kinds of diagnostic horizons Table 33.50.

In the United States, Ustalfs tend to be on older landforms than Udalfs and have thicker sola; many are in the Paleustalf great group because of either an abrupt textural change between E and Bt horizons or a thick Bt horizon enriched in clay throughout. There is less tendency for E horizons to display platy structure, while with less rainfall, E horizons are often absent. The Bt horizons often have prismatic parting to angular or subangular blocky structure. Secondary carbonates commonly accumulate to form calcic horizons in the lower portion of the argillic horizon or just below it in the dryer Ustalfs.

33.11.5.5 Xeralfs

These occur in regions that have a Mediterranean climate, where precipitation occurs during the cool season and the summers are hot and dry. In some winters, water moves through the entire soil profile to deeper layers. Most Xeralfs border the Mediterranean Sea or lie to the east of an ocean, such as those in the western United States and Australia. Where there is no irrigation, grains are common crops but with irrigation, a variety of crops can be grown. Grapes and olives are grown in the warmer areas (Soil Survey Staff, 1999). Occurrence of E horizons is uncommon, and A horizons with weak subangular blocky structure to massive condition tend to range from 15 to 25 cm thick. Most argillic horizons are relatively thin (15–75 cm) with significant quantities of coarse fragments throughout the pedon.

33.11.6 Soil Formation Processes

33.11.6.1 Clay Translocation

Alfisols and Ultisols form mainly by eluviation of silicate clay from A and E to Bt horizons. In Alfisols, however, the base saturation of deep subsoil layers is greater than it is in Ultisols. Alfisols are similar to the Gray–Brown Podzolic soils of previous soil classification systems. The name Alfisol was derived from the term Pedalfer, a coined word used in previous classifications derived from ped, as in pedology (the science of natural soils), Al, aluminum, and Fe, iron. The implication is that in Pedalfers, Al and Fe accumulate in the subsoil, in contrast to Pedocals, in which Ca accumulates.

The concept of Gray–Brown Podzolic soils originated at a time when soil colloids (clays) were thought to be mainly amorphous. Then, soil scientists determined total chemical composition of the colloidal fraction and interpreted the results by examining SiO_2/Al_2O_3 and $SiO_2/(Fe_2O_3 + Al_2O_3)$ ratios of the various horizons. These ratios showed that both Podzols (Spodosols) and Gray–Brown Podzolic soils had B horizons enriched in Fe and Al, but that Podzol B horizons were also enriched in organic C. This led Marbut (1935) to conclude that the differences in the processes leading to the two kinds of soil were of degree, not kind, and that Gray–Brown Podzolic soils were less developed than Podzols. The word Podzolic is a descriptor previously used for many Alfisols.

Hendricks and Fry (1930) reported two key discoveries that laid the groundwork for the current understanding of the processes by which Alfisols form. First, they showed that soil colloids are mainly crystalline silicate minerals, and that when clay suspensions are allowed to dry, the plate-shaped minerals lay flat to form masses that have a regular crystallographic, or at least optical, arrangement and are large enough to study with a polarizing microscope. This discovery led to an understanding that in Alfisols, the material translocated is mainly silicate clay minerals (with Fe oxides on their surfaces), but in Spodosols, it is mainly amorphous complexes of organic matter with Al and usually Fe. Second, they provided the basis for interpreting what are now called clay films.

These ideas were gradually applied to studies of soil genesis. Jenny and Smith (1935) produced early stages of argillic horizon formation in columns of pure quartz sand by first coating the sand grains with electropositive iron hydroxide and then passing dispersed clay through the column. Brown and Thorp (1942) noted that the B horizons in the Miami soil (Hapludalf) from Indiana and similar soils contained much more clay than did the A and E horizons. They concluded, however, that the clay increase in the B horizons was only partly due to illuviation. Although the types of clay were identified by x-ray diffraction and thermal methods, soil formation processes were still discussed mainly in terms of $SiO_2/(Fe_2O_3 + Al_2O_3)$ ratios. Frei and Cline (1949) concluded that the apparent loss of clay from A and E horizons and the apparent gain in clay in B horizons of a New York soil could be accounted for in at least three possible ways: (1) clay migrates downward as a sol in percolating water; (2) clay is synthesized in the B horizon from soluble weathering products that were released there or moved down in the profile; or (3) the apparent increase in clay was really due to loss of other materials, such as $CaCO_3$. They showed, too, that there was a strong concentration of clays with a high degree of optical continuity on peds in the B horizon of a Gray–Brown Podzolic soil, but in the upper B horizon, these coatings were highly degraded. According to their interpretation, clay that was mobilized in A and E horizons and stripped from ped surfaces in the upper B horizon, moved down the profile in suspension, and was deposited on ped surfaces as films of oriented clay particles. Thorp et al. (1959), in another study of the Miami soil of Indiana, showed that the small particle size of clays, their association with organic compounds, and shrink/swell cycles in upper horizons all enhanced their mobility.

In summary, the movement of clay from A and E horizons to Bt horizons can be viewed as the net result of three subprocesses: dispersion, transport, and deposition (Jenny, 1980). Dispersion, the release of individual clay particles from aggregates, is favored by the replacement of exchangeable Ca^{2+} with Na^+, by adsorption of humus molecules on clay particles, and by mechanical disruption of aggregates, which may be caused by wetting and drying, freezing and thawing, or by mechanical disruption by faunal activity. Small particles move more readily than large particles, so that illuvial clay, as in clay films, contains more fine than immobile clay, such as that in ped interiors. Little clay is dispersed in calcareous soil materials because Ca^{2+} ions on exchange sites favor clay flocculation. Clay that is mobilized in A and E horizons

moves as a percolating sol or as slurry creep along ped surfaces. Deposition is favored when Na$^+$ is replaced by Ca^{2+}, resulting in clay flocculation by an increase in pH, or by reaction of the silicate clay with oxides such as Fe oxide. When the water of a slurry enters into a ped, the clay it contains is deposited on the ped surface, like passing the slurry through filter paper. In some cases, the slurry moves down to the wetting front in the subsoil, and clay is deposited as water evaporates or is taken up by plants. Texture discontinuities may also arrest the percolation of suspensions.

Because many Alfisols formed in calcareous materials, CaCO$_3$ must be dissolved and the weathering products leached from the soil before much clay can move. Usually H$^+$ replaces a portion of the Ca^{2+} in these Alfisols, causing leached horizons to become acidic. The shallowest depth at which carbonate minerals are found, as detected by dilute HCl, serves as an index of the intensity and length of soil formation.

It should be stressed that although illuviation of clays is diagnostic for the argillic horizon, much or even most clay in the Alfisol Bt horizons originates from other sources to include inheritance from parent material, clay neoformation, in situ weathering of lithogenic minerals, and dissolution of silt- and sand-sized carbonates (Smeck and Wilding, 1980).

33.11.6.2 Natric Horizons

Natric horizons are argillic horizons high in exchangeable Na$^+$. A natric horizon in an Alfisol seems to be an oxymoron; if rainfall has been high enough to move sufficient clay for the horizon to qualify as an argillic horizon, Na$^+$ should have long since been leached from the soil. However, natric horizons do exist in Alfisols indicating that there are circumstances in which they form. Some might be caused by a change in climate where the argillic horizon was formed during a previous wetter period, but rainfall is now less, and the leaching of Na$^+$ is reduced. Other soils with natric horizons occur in small areas where Na$^+$ has accumulated. Although the original source of much of the Na is feldspar minerals, Na$^+$ moves about in the landscape once it is released into solution. Often, soils with natric horizons are low in the landscape because solutions containing Na$^+$ move downslope either on the soil surface or through subsurface horizons, thus concentrating Na$^+$ in the low areas. In some soils, Na$^+$ is retained in the soil because the hydraulic conductivity of B horizons is very low, and leaching of Na$^+$ is retarded. This seems to be a self-perpetuating process; Na$^+$ accumulates because permeability is low, while Na$^+$ promotes dispersion, further slowing permeability. Furthermore, some soils with natric horizons were formerly saline soils. When soluble salts are leached from a soil, Na$^+$ may remain behind on the exchange sites.

33.11.6.3 Fragipan Formation

Fragipans occur in many Alfisols of the humid temperate regions, especially those of silty and loamy textures. These subsurface horizons restrict the entry of water and roots into the soil matrix (prisms) and often underlie argillic horizons. Commonly, the fragipan has a relatively low content of organic matter and high bulk density relative to horizons above. Most fragipans consist

1. Root
2. Humic matter
3. Non-bleached earth
4. Bleached earth
5. Iron precipitation
6. Clay accumulation, diffuse or compact (black)
7. Platy or lenticular structure
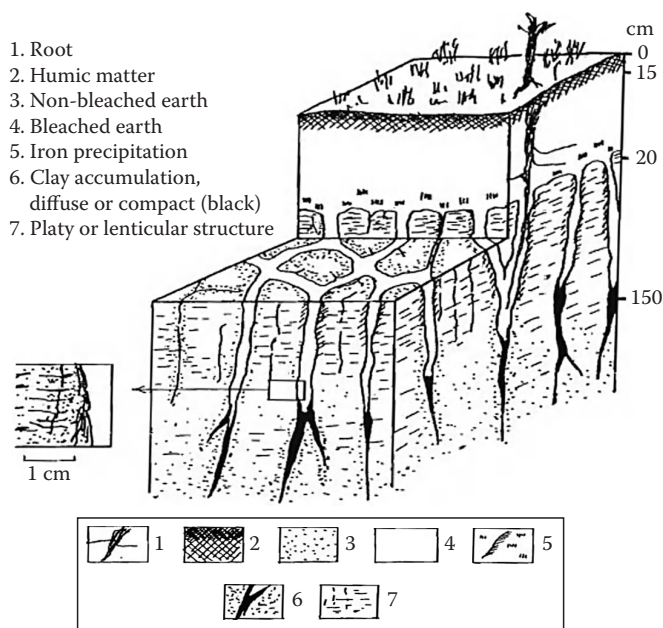
FIGURE 33.74 Schematic of a fragipan in an imperfectly drained silty soil. (Reprinted from van Vliet, B., and R. Langohr. 1981. Correlation between fragipans and permafrost with special reference to silty deposits in Belgium and northern France. Catena 8:137–154. Copyright with permission from Elsevier Science.)

of prisms more than 10 cm across with light-colored coatings on prism faces. These light-colored coatings form a polygonal pattern as shown in Figure 33.74. Material within prisms has brittle failure, a tendency for a clod to rupture suddenly rather than undergo slow deformation when pressure is applied. Fragments of the fragipan slake (fall apart) when submersed in water.

There are many theories on how fragipans form, summarized in two reviews (Smalley and Davin, 1982; Smeck and Ciolkosz, 1989). One reason for the diversity in ideas is that there may be several kinds of horizons included under the name fragipan, each with its own set of processes of formation. Most soil scientists believe that they have formed, or are forming, under present day conditions. The main evidence that fragipans are genetic horizons related to the current land surface is that they are roughly parallel to the soil surface, and the upper boundary has a relatively narrow range of about 50–100 cm depth. Some scientists, however, believe that they formed much earlier. The following are some of the various theories of fragipan formation.

33.11.6.3.1 Paleosols

According to this theory, fragipans formed in earlier times and were buried by younger deposits in which fragipans are not forming. In western Tennessee, Buntley et al. (1977) observed that fragipans are mainly in older loess deposits that were buried by younger Peorian loess in which no fragipans form.

33.11.6.3.2 Rapid Initial Development

Some workers believe that fragipans formed quickly in their life history. Fragipan formation has been related to periglacial

conditions (Payton, 1992) and to permafrost (FitzPatrick, 1956; van Vliet and Langohr, 1981). The polygonal pattern formed by coatings on the coarse prismatic structure is similar to that found in periglacial regions. Others present evidence that the structure of loess collapses when it first dries after deposition (Bryant, 1989), and that this denser material becomes the fragipan.

Much glacial till is very dense because it was compacted by the mass of glacier ice. Subsoil horizons that form in dense till may inherit some of their strength from the parent material (Lindbo and Veneman, 1989), but they might also develop fragipan properties through soil-forming processes. In these soils, which are common from eastern Ohio to New England, it is difficult to decide whether the dense material should be considered a fragipan or a slightly modified parent material. The great strength of these horizons is due in part to their high density, but it may also be due to cementation by Si or carbonates (McBurnett and Franzmeier, 1997). The Cd horizon was introduced, in part, to describe little altered, unconsolidated, dense till which exclude roots much like fragipans (Soil Survey Staff, 2006).

33.11.6.3.3 Chemical Cementation

Fragipans could also be cemented by Fe-, Al-, or Si-rich materials. Iron-rich materials cause the cementation of ortstein layers, and silica bonds particles together to form duripans. Although duripans form mainly in aridic and xeric climates, in more humid climates, they grade into fragipans (Soil Survey Staff, 1999). The idea that fragipans are cemented by silica was first proposed in the 1930s and 1940s but was discredited and ignored for about 40 years before revitalized in the late 1970s. The proposal requires certain interactions of parent materials and climate (Franzmeier et al., 1989). In addition to undergoing periodic leaching, soils must also become seasonally dry, which commonly occurs in areas with a udic moisture regime under hardwood forest. Also, downward percolation of soil solutions must be arrested, rather than passing quickly through subsoil layers. This could be caused by the depth of penetration of the wetting front, by discontinuities in parent materials (Smeck et al., 1989), or by slowly permeable, deeper subsoil layers. Silicate minerals weather in upper horizons, and the silica released moves down, usually in the winter and early spring, to the subsoil where it remains for a time. When the trees leaf out, water taken up by roots concentrates silica in solution eventually causing it to precipitate. Compared with grasses, hardwood trees take up relatively little Si, and thus have the potential to concentrate Si in the soil. The precipitated silica bridges between clay particles, or more likely between Fe oxide minerals adsorbed on silicate clay surfaces, causing weak cementation. The greater Si uptake in grasses may be the reason that fragipans do not form under prairie vegetation in humid climates (Franzmeier et al., 1989). The precipitation of Si must be somewhat irreversible, or else any cementation that formed in the summer would be destroyed in the winter, and no net fragipan formation would take place. Fragipans do not form readily on steep slopes, in deep loess, or in loess over more permeable materials, such as outwash. In these soils, the soil solution moves laterally

downslope or through the profile, preventing the accumulation of Si-rich soil solutions.

Most of these theories of fragipan formation individually have deficiencies. Perhaps, various combinations of them can explain how a particular soil forms.

33.11.7 Selected Physical, Chemical, and Mineralogical Properties

An increase in clay content between the E and/or A horizon(s) and underlying Bt horizons occurs in all Alfisols; however, this textural contrast ranges from little to great. This is illustrated (Figure 33.75) by the change in clay content with depth for a Psammentic Haploxeralf from California (Torrent et al., 1980), a Typic Hapludalf from Illinois (Grossman and Fehrenbacher, 1971) and a Udertic Paleustalf from Texas (Klich et al., 1990). The Psammentic Haploxeralf has a minimal argillic horizon, with a clay increase of about 4% above that of eluvial horizons. By contrast, the Paleustalf shows a clay increase of more than 40% between the surface horizon and the finest textured portion of the argillic horizon. The Hapludalf is intermediate in textural contrast.

Alfisols with Bt horizons dominated by fine clay (<0.0002 mm, usually smectitic) display significant volume change between wet and dry states. This volume change is reported as the coefficient of linear extensibility (COLE) and can be determined from bulk density values associated with moist and dry, natural clods
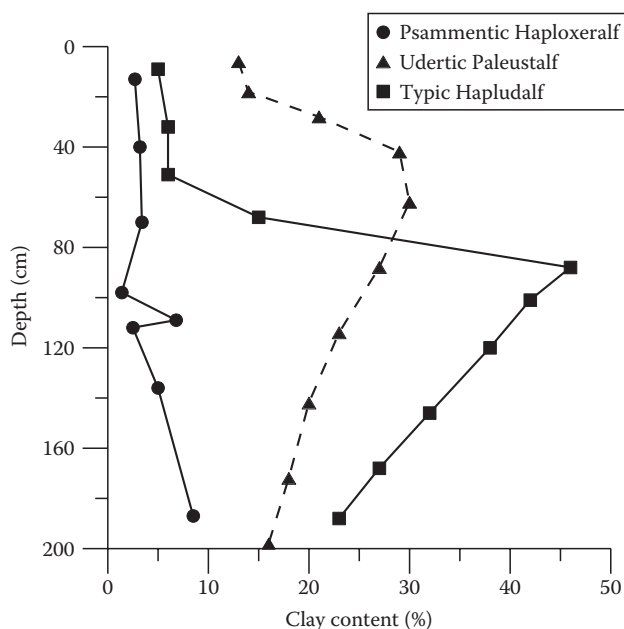
FIGURE 33.75 Clay distribution in three Alfisols of contrasting degrees of argillic horizon development. (From Torrent, J., W.D. Nettleton, and G. Borst. 1980. Clay illuviation and lamella formation in a Psammentic Haploxeralf in southern California. Soil Sci. Soc. Am. J. 44:363–369; Grossman, R.B., and J.B. Fehrenbacher. 1971. Distribution of moved clay in four loess-derived soils that occur in southern Illinois. Soil Sci. Soc. Am. Proc. 35:948–951; Klich, I., L.P. Wilding, and A.A. Pfordresher. 1990. Close-interval spatial variability of udertic paleustalfs in East Central Texas. Soil Sci. Soc. Am. J. 54:489–494.

TABLE 33.52 Mineralogy of the Argillic and Kandic Horizons of Selected Alfisols

			Minerals and/or Silt-Fraction[a]																
			Clay Fraction[a] (<0.002 mm)								Sand and/or Silt-Fraction[b]								
Classification	Location	Parent Material	Ch	Ve	ln	Sm	Mi	Ka	Qz	GH	Ps	Kf	Pl	Mi	Qz	Vg	AP	Ze	Source
Glossoboralfs	Wisconsin	Lacustrine sediments	1	2		2	2	1				12	18	8	55				Ranney and Beatty (1969)
Eutroboralf	New Mexico	Sandstone				2	2	2				18			76	6			Anderson et al. (1975)
Hapludalf	Wisconsin	Thin loess over till	1	2		2	2	2	1			13	14	8	61				Borchardt et al. (1968)
Hapludalf	Oregon	Greenish tuff and breccia	1			3					22		52	3	1	t	6	11	Paeth et al. (1971)
Haploxeralf	Spain	Calcareou salluvium			1	2	3	1				-a-		t	t				Delgado et al. (1994)
Rhodoxeralf		Red sandstone			1	1	3	1				-a-		c					
Haplustalf	Texas	Granite				2	2	1											Goss and Allen (1968)
Natrustalf						2	2	1											
Paleudalf	Texas	Coastal plains sediments				1	1	3+		1									Vepraskas and Wilding
Haplustalf?	India	Schist/quartzite/phyllite				t	2	3	1	1		31			65				Sahu et al. (1990)
Kandiudalf (reclassified)	Nigeria	Olivine basalt						3	t	1									Gallez et al. (1975)
Plinthustalf		Basement complex				t		1	3	1									
Paleustalf	Niger	Eolian sands						3	1										West et al. (1984)

t, trace; Ch, chlorite; Ve, vermiculite; ln, interstratified 2:1 minerals; Sm, smectite; Mi, hydrous mica or illite; Ka, kaolinite; Qz, quartz; GH, goethite and/or hematite; Ps, pseudomorphs of clay minerals; Kf, K feldspar; Pl, plagioclase feldspar; Mi, mica; Qz, quartz; Vg, volcanic glass; AP, amphibole and/or pyroxene; Ze, zeolite; a, abundant; c, common; t, trace; +, chloritized vermiculite also noted.

[a] Relative amounts 3 = >50%; 2 = 10%–50%; 1 = <10%.

[b] Values are for a major fraction of sand or slit.

(Grossman et al., 1968). The Paleustalf in Figure 33.75 has COLE values >0.07 (7% volume change between moist and dry states) for horizons containing 30% or more clay. Because such soils are intergrades to Vertisols and possess similar features such as high subsoil clay content and shrink/swell potential, and slickensides, they present similar problems for use as construction materials (i.e., unstable subgrades for roads and houses).

Texture and clay content affect water-holding and water transmission properties of soils since water-related properties are dependent on void size and continuity. When the size of voids change greatly over a short distance, water movement is restricted. Alfisols with lamella type argillic horizons may exhibit slower drainage and better soil–plant relations due to greater available water retention than similar textured soils without lamella. Under more extreme conditions, strongly contrasting textures of eluvial and argillic horizons may result in water perched above the argillic horizons for several weeks as in Natraqualfs and Albaqualfs. In Alfisols with compact, clayey argillic horizons, root growth may be limited primarily to structural planes due to the high bulk density of ped interiors. Jones (1983) showed that root exclusion due to soil bulk density is a function of clay content.

Although organic C is greatest in the surface horizons, Alfisols tend to have greater CEC in the argillic horizon where clay and smectite contents are greatest. Generally, Alfisols must have a base saturation of 35% or more at the shallower of the following depths: (1) 1.25 m below the top of the argillic or kandic horizon, (2) 1.8 m below the soil surface, (3) 75 cm below the top of a fragipan, or (4) immediately above a densic, petroferric, lithic, or paralithic contact (just above continuous hard or soft bedrock; see Soil Survey Staff, 2010, for details). Most Alfisols have base saturation percentages well above 35%; those that approach low base status are recognized as intergrades to Ultisols by classification in Ultic subgroups. Generally, base saturation (and pH) decreases initially with depth in Alfisols, often reaching a minimum in the E horizons or upper portions of the argillic horizon, then increases again with depth. Alfisols under deciduous forest vegetation have relatively high base saturations in the surface horizon due to cycling of basic cations from depth to the surface in leaf fall. The relationship between soil pH and base saturation (Beery and Wilding, 1971; Blosser and Jenny, 1971; Ranney et al., 1974) is often used in field operations to separate Alfisols from Ultisols, but care must be exercised as this relationship changes with mineralogy.

The mineralogy from a range of Alfisols illustrating variations in parent materials, climates, and degree of development is presented in Table 33.52. Most Alfisols contain substantial quantities of 2:1 clay minerals such as smectite and hydrous mica, and perhaps some vermiculite, chlorite, and kaolinite. However, for more intensely weathered Alfisols (Kandiudalf, Paleudalf, Paleustalf, and Plinthustalf), kaolinite is the dominant clay mineral in the argillic or kandic horizon. Similarly, the most common mineral dominating the sand and silt fraction is quartz (Table 33.52), which comprised >90% of the sand fraction in the Paleustalf and Paleudalf. The Udalf from Oregon, which formed from tuff and breccia, was dominated by plagioclase feldspars.

Most Alfisols contain significant quantities of weatherable minerals including K feldspars, plagioclase, and mica in the sand and silt fractions.

References

Aide, M., and A. Marshaus. 2002. Fragipan genesis in two Alfisols in east central Missouri. Soil Sci. 167:453–464.

Allan, R.J., and F.D. Hole. 1968. Clay accumulation in some Hapludalfs as related to calcareous till and incorporated loess on drumlins in Wisconsin. Soil Sci. Soc. Am. Proc. 32:403–408.

Allen, B.L., B.L. Harris, K.R. Davis, and G.B. Miller. 1972. The mineralogy and chemistry of High Plains playa lake soils and sediments. Texas Technical University Water Resources Center WRC 72-7. Texas Tech University, Lubbock, TX.

Anderson, J.U., O.F. Bailey, and D. Rai. 1975. Effects of parent material on genesis of Borolls and Boralfs in south-central New Mexico mountains. Soil Sci. Soc. Am. Proc. 39:901–904.

Arkley, R.J., and H.C. Brown. 1954. The origin of mima mound (hogwallow) microrelief in the far western states. Soil Sci. Soc. Am. Proc. 18:195–199.

Bailey, H.H., and P.E. Avers. 1971. Classification and composition of soils from mountain colluvium in eastern Kentucky and their importance in forestry. Soil Sci. 111:244–251.

Beery, M., and L.P. Wilding. 1971. The relationship between soil pH and base-saturation percentage for surface and subsoil horizons of selected Mollisols, Alfisols, and Ultisols in Ohio. Ohio J. Sci. 71:43–55.

Bhattacharyya, T., D.K. Pal, and S.B. Deshpande. 1993. Genesis and transformation of minerals in the formation of red (Alfisols) and black (Inceptisols and Vertisols) soils on Deccan basalt in the Western Ghats, India. J. Soil Sci. 44:159–171.

Bilzi, A.F., and E.J. Ciolkosz. 1977. Time as a factor in the genesis of four soils developed in recent alluvium in Pennsylvania. Soil Sci. Soc. Am. J. 41:122–127.

Blank, R.R., and M.A. Fosberg. 1991. Duripans of the Owyhee Plateau region of Idaho: Genesis of opal and sepiolite. Soil Sci. 152:116–133.

Blosser, D.L., and H. Jenny. 1971. Correlations of soil pH and percent base saturation as influenced by soil-forming factors. Soil Sci. Soc. Am. Proc. 35:1017–1018.

Bockheim, J.G. 1980. Solution and use of chronofunctions in studying soil development. Geoderma 24:71–85.

Borchardt, G.A., F.D. Hole, and M.L. Jackson. 1968. Genesis of layer silicates in representative soils in a glacial landscape of southeastern Wisconsin. Soil Sci. Soc. Am. Proc. 32:399–402.

Bouma, J., L.J. Pons, and J. van Schuylenborgh. 1968. On soil genesis in temperature humid climate. VI. The formation of a Glossudalf in loess (silt loam). Neth. J. Agric. Sci. 16:58–70.

Bouma, J., and J. van Schuylenborgh. 1969. On soil genesis in temperate humid climate. VII. The formation of a Glossaqualf in a silt loam terrace deposit. Neth. J. Agric. Sci. 17:261–271.

Brown, I.C., and J. Thorp. 1942. Morphology and composition of some soils of the Miami family and the Miami catena. USDA Technical Bulletin No. 834. USDA, Washington, DC.

Bryant, R.B. 1989. Physical processes of fragipan formation, p. 141–150. *In* N.E. Smeck and E.J. Ciolkosz (eds.) Fragipans: Their occurrence, classification, and genesis. SSSA Special Publication No. 24. SSSA, Madison, WI.

Buntley, G.J., R.B. Daniels, E.E. Gamble, and W.T. Brown. 1977. Fragipan horizons in soils of the Memphis–Loring–Grenada sequence in west Tennessee. Soil Sci. Soc. Am. J. 41:400–407.

Calero, J., R. Delgado, G. Delgado, and J.M. Martin-Garcia. 2008. Transformation of categorical field soil morphological properties into numerical properties for the study of chronosequences. Geoderma 145:278–287.

Catt, J.A. 1991. Soils as indicators of Quaternary climatic change in mid-latitude regions. Geoderma 51:167–187.

Chittleborough, D.L. 1989. Genesis of a xeralf on feldspathic sandstone, South Australia. J. Soil Sci. 40:235–250.

Chittleborough, D.J., P.H. Walker, and J.M. Oades. 1984. Textural differentiation in chronosequences from eastern Australia. I. Description, chemical properties and micromorphologies of soils. Geoderma 32:181–202.

Cooper, A.W. 1960. An example of the role of microclimate in soil genesis. Soil Sci. 90:109–120.

Coventry, R.J., and J. Williams. 1984. Quantitative relationships between morphology and current soil hydrology in some Alfisols in semi-arid tropical Australia. Geoderma 33:191–218.

Cremeens, D.L., and D.L. Mokma. 1986. Argillic horizon expression and classification in the soils of two Michigan hydrosequences. Soil Sci. Soc. Am. J. 50:1002–1007.

Davis, K.R. 1970. Mineralogy of a playa soil and underlying sediments from the High Plains medium-textured soil zone. M.S. Thesis. Texas Tech University. Lubbock, TX.

Davis, J.G., L.R. Hossner, L.P. Wilding, and A. Manu. 1995. Variability of soil chemical properties in two sandy, dunal soils of Niger. Soil Sci. 159:321–330.

Delgado, R., J. Aguilar, and G. Delgado. 1994. Use of numerical estimators and multivariate analysis to characterize the genesis and pedogenic evolution of xeralfs from southern Spain. Catena 23:309–325.

FitzPatrick, E.A. 1956. An indurated horizon formed in permafrost. J. Soil Sci. 7:248–254.

Foss, J.E., and R.H. Rust. 1968. Soil genesis study of a lithologic discontinuity in glacial drift in western Wisconsin. Soil Sci. Soc. Am. Proc. 32:393–398.

Franzmeier, D.P., R.B. Bryant, and G.C. Steinhardt. 1985. Characteristics of Wisconsin glacial tills in Indiana and their influence on argillic horizon development. Soil Sci. Soc. Am. J. 49:1481–1486.

Franzmeier, D.P., L.D. Norton, and G.C. Steinhardt. 1989. Fragipan formation in loess of the midwestern United States, p. 69–98. *In* N.E. Smeck and E.J. Ciolkosz (eds.) Fragipans: Their occurrence, classification, and genesis. SSSA Special Publication No. 24. SSSA, Madison, WI.

Frei, E., and M.G. Cline. 1949. Profile studies of normal soils of New York: II. Micro-morphological studies of the gray-brown podzolic soil sequence. Soil Sci. 68:333–344.

Gallez, A., A.S.R. Juo, A.J. Herbillon, and F.R. Moormann. 1975. Clay mineralogy of selected soils in southern Nigeria. Soil Sci. Soc. Am. Proc. 39:577–585.

Gersper, P.L., and N. Holowaychuk. 1970a. Effect of stemflow water on a Miami soil under a beech tree. I. Morphological and physical properties. Soil Sci. Soc. Am. Proc. 34:779–786.

Gersper, P.L., and N. Holowaychuk. 1970b. Effects of stemflow water on a Miami soil under a beech tree: II. Chemical properties. Soil Sci. Soc. Am. Proc. 34:786–794.

Goss, D.W., and B.L. Allen. 1968. A genetic study of two soils developed on granite in Llano County, Texas. Soil Sci. Soc. Am. Proc. 32:409–413.

Grossman, R.B., B.R. Brasher, D.P. Franzmeier, and J.L. Walker. 1968. Linear extensibility as calculated from natural-clod bulk density measurements. Soil Sci. Soc. Am. Proc. 32:570–573.

Grossman, R.B., and J.B. Fehrenbacher. 1971. Distribution of moved clay in four loess-derived soils that occur in southern Illinois. Soil Sci. Soc. Am. Proc. 35:948–951.

Harden, J.W. 1982. A quantitative index of soil development from field descriptions: Examples from a chronosequence in central California. Geoderma 28:1–28.

Hendricks, S.B., and W.H. Fry. 1930. The results of X-ray and microscopical examinations of soil colloids. Soil Sci. 29:457–479.

Hugie, V.K., and H.B. Passey. 1963. Cicadas and their effect upon soil genesis in certain soils in southern Idaho, northern Utah, and northeastern Nevada. Soil Sci. Soc. Am. Proc. 27:78–82.

Jenny, H. 1935. The clay content of the soil as related to climatic factors, particularly temperature. Soil Sci. 40:111–128.

Jenny, H. 1941. Factors of soil formation. McGraw-Hill, New York.

Jenny, H. 1980. The soil resource. Springer-Verlag, New York.

Jenny, H., and C.D. Leonard. 1934. Functional relationships between soil properties and rainfall. Soil Sci. 38:363–381.

Jenny, H., and G.D. Smith. 1935. Colloid chemical aspects of clay pan formation in soil profiles. Soil Sci. 39:377.

Jones, C.A. 1983. Effect of soil texture on critical bulk densities for root growth. Soil Sci. Soc. Am. J. 47:1208–1211.

Juo, A.S.R., and L.P. Wilding. 1996. Soils of the lowland forests of West and Central Africa. Proc. R. Soc. Edinb. 104B:15–29.

Klich, I., L.P. Wilding, and A.A. Pfordresher. 1990. Close-interval spatial variability of udertic paleustalfs in East Central Texas. Soil Sci. Soc. Am. J. 54:489–494.

Kooistra, M.J., A.S.R. Juo, and D. Schoonderbeek. 1990. Soil degradation in cultivated Alfisols under different management systems in southwestern Nigeria, p. 61–69. *In* L.A. Douglas (ed.) Soil micromorphology: A basic and applied science. Elsevier, New York.

Lindbo, D.L., and P.L.M. Veneman. 1989. Fragipans in the northeastern United States, p. 11–31. *In* N.E. Smeck and E.J. Ciolkosz (eds.) Fragipans: Their occurrence, classification, and genesis. SSSA Special Publication No. 24. SSSA, Madison, WI.

Little, I.P., and W.T. Ward. 1981. Chemical and mineralogical trends in a chronosequence developed on alluvium in eastern Victoria, Australia. Geoderma 25:173–188.

Lotspeich, F.B., and J.R. Coover. 1962. Soil forming factors on the Llano Estacado: Parent material, time and topography. Tex. J. Sci. 14:7–17.

Lozet, J.M., and A.J. Herbillon. 1971. Fragipan soils of Controz (Belgium): Mineralogical, chemical and physical aspects in relation to their genesis. Geoderma 5:325–343.

Marbut, C.F. 1935. Atlas of American agriculture. Part III: Soils of the United States. U.S. Government Printing Office, Washington, DC.

Markewich, H.W., and M.J. Pavich. 1991. Soil chronosequence studies in temperature to subtropical, low-latitude, low-relief terrain with data from the eastern United States. Geoderma 51:213–239.

McBurnett, S.L., and D.P. Franzmeier. 1997. Pedogenesis and cementation in calcareous till in Indiana. Soil Sci. Soc. Am. J. 61:1098–1104.

Meixner, R.E., and M.J. Singer. 1981. Use of a field morphology rating system to evaluate soil formation and discontinuities. Soil Sci. 131:114–123.

Miles, R.J., and D.P. Franzmeier. 1981. A lithochronosequence of soils formed in dune sand. Soil Sci. Soc. Am. J. 45:362–367.

Miller, F.P., N. Holowaychuk, and L.P. Wilding. 1971. Canfield silt loam, a fragiudalf: I. Macromorphological, physical and chemical properties. Soil Sci. Soc. Am. Proc. 35:319–324.

Moberg, J.P., and I.E. Esu. 1991. Characteristics and composition of some savanna soils in Nigeria. Geoderma 48:113–129.

Muhs, D.R. 1982. A soil chronosequence on Quaternary marine terraces, San Clemente Island, California. Geoderma 28:257–283.

Nielsen, G.A., and F.D. Hole. 1964. Earthworms and the development of coprogenous A1 horizons in forest soils of Wisconsin. Soil Sci. Soc. Am. Proc. 28:426–430.

Paeth, M.E., M.E. Harward, E.G. Knox, and C.T. Dryness. 1971. Factors affecting mass movement of four soils in the Western Cascades of Oregon. Soil Sci. Soc. Am. Proc. 35:943–947.

Parsons, R.B., W.H. Scholtes, and F.F. Riecken. 1962. Soils of Indian mounds in northeast Iowa as benchmarks of soil genesis. Soil Sci. Soc. Am. Proc. 26:491–496.

Parsons, R.B., G.H. Simonson, and C.A. Balster. 1968. Pedogenic and geomorphic relationships of associated aqualfs, albolls, and xerolls in Western Oregon. Soil Sci. Soc. Am. Proc. 32:556–563.

Payton, R.W. 1992. Fragipan formation in argillic brown earths (Fragiudalfs) of the Milfield Plain, northeast England. I. Evidence for a periglacial stage of development. J. Soil Sci. 43:621–644.

Ranney, R.W., and M.T. Beatty. 1969. Clay translocation and albic tongue formation in two glossoboralfs of west-central Wisconsin. Soil Sci. Soc. Am. Proc. 33:768–775.

Ranney, R.W., E.J. Ciolkosz, R.L. Cunningham, G.W. Petersen, and R.P. Matelski. 1975. Fragipans in Pennsylvania soils: Properties of bleached prism face materials. Soil Sci. Soc. Am. Proc. 39:695–698.

Ranney, R.W., E.J. Ciolkosz, G.W. Peterson, R.P. Matelski, L.J. Johnson, and R.L. Cunningham. 1974. The pH and base-saturation relationships in B and C horizons of Pennsylvania soils. Soil Sci. 118:247–253.

Ransom, M.D., N.E. Smeck, and J.M. Bigham. 1987. Stratigraphy and genesis of polygenetic soils on the Illinoisan till plain of Southwestern Ohio. Soil Sci. Soc. Am. J. 51:135–141.

Reynders, J.J. 1972. A study of argillic horizons in some soils in Morocco. Geoderma 8:267–271.

Ruhe, R.V. 1984a. Loess-derived soils, Mississippi Valley Region: I. Soil-sedimentation system. Soil Sci. Soc. Am. J. 48:859–863.

Ruhe, R.V. 1984b. Loess-derived soils, Mississippi Valley Region: II. Soil-climate system. Soil Sci. Soc. Am. J. 48:864–867.

Sahu, G.C., S.N. Patnaik, and P.K. Das. 1990. Morphology, genesis, mineralogy and classification of soils of Northern Plateau Zone of Orissa. J. Indian Soc. Soil Sci. 38:116–121.

Saunders, J.W., R.D. Mandel, R.T. Saucier, E.T. Allen, C.T. Hallmark, J.K. Johnson, E.H. Jackson et al. 1997. A mound complex in Louisiana at 5400–5000 years before the present. Science 277:1796–1799.

Schaetzl, R.J. 1996. Spodosol–Alfisol intergrades: Bisequal soils in N. E. Michigan, USA. Geoderma 74:23–47.

Schaetzl, R.J. 1998. Lithologic discontinuities in some soils on drumlins: Theory, detection, and application. Soil Sci. 163:570–590.

Shaw, J.N., J.W. Odom, and B.F. Hajek. 2003. Soils on Quarternary terraces of the Tallapoosa River, central Alabama. Soil Sci. 168:707–717.

Smalley, I.J., and J.E. Davin. 1982. Fragipan horizons in soils: A bibliographic study and review of some of the hard layers in loess and other materials. New Zealand Soil Bureau Bibliographic Report 30. Department of Scientific and Industrial Research, Wellington, New Zealand.

Smeck, N.E., and E.J. Ciolkosz (eds.). 1989. Fragipans: Their occurrence, classification, and genesis. SSSA Special Publication No. 24. SSSA, Madison, WI.

Smeck, N.E., M.L. Thompson, L.D. Norton, and M.J. Shipitalo. 1989. Weathering discontinuities: A key to fragipan formation, p. 99–112. *In* N.E. Smeck and E.J. Ciolkosz (eds.) Fragipans: Their occurrence, classification, and genesis. SSSA Special Publication No. 24. SSSA, Madison, WI.

Smeck, N.E., and L.P. Wilding. 1980. Quantitative evaluation of pedon formation in calcareous glacial deposits in Ohio. Geoderma 24:1–16.

Smeck, N.E., L.P. Wilding, and N. Holowaychuk. 1968. Genesis of argillic horizons in Celina and Morley soils of western Ohio. Soil Sci. Soc. Am. Proc. 32:550–556.

Smith, H., and L.P. Wilding. 1972. Genesis of argillic horizons on ochraqualfs derived from fine-textured till deposits of northwestern Ohio and southeastern Michigan. Soil. Sci. Soc. Am. Proc. 36:808–815.

Soil Survey Staff. 1975. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. USDA agriculture handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd edn. USDA agriculture handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS, U.S. Government Printing Office, Washington, DC.

Stahnke, C.R., L.P. Wilding, J.D. Moore, and L.R. Drees. 1983. Genesis and properties of Paleustalfs of North Central Texas: Morphological, physical, and chemical properties. Soil Sci. Soc. Am. J. 47:728–733.

Thorp, J. 1949. Effects of certain animals that live in soils. Sci. Mon. 68:180–191.

Thorp, J., J.G. Cady, and E.E. Gamble. 1959. Genesis of Miami silt loam. Soil Sci. Soc. Am. Proc. 23:156–161.

Torrent, J., and W.D. Nettleton. 1979. A simple textural index for assessing chemical weathering in soils. Soil Sci. Soc. Am. J. 43:373–377.

Torrent, J., W.D. Nettleton, and G. Borst. 1980. Clay illuviation and lamella formation in a Psammentic Haploxeralf in southern California. Soil Sci. Soc. Am. J. 44:363–369.

van Vliet, B., and R. Langohr. 1981. Correlation between fragipans and permafrost with special reference to silty deposits in Belgium and northern France. Catena 8:137–154.

Vepraskas, M.J., and L.P. Wilding. 1983. Deeply weathered soils in the Texas Coastal Plain. Soil Sci. Soc. Am. J. 47:293–300.

Weisenborn, B.N., and R.J. Schaetzl. 2005. Range of fragipan expression in some Michigan soils: II. A model for fragipan evolution. Soil Sci. Soc. Am. J. 69:178–187.

Weitkamp, W.A., R.C. Graham, and M.A. Anderson. 1996. Pedogenesis of a vernal pool Entisol–Alfisol–Vertisol catena in southern California. Soil. Sci. Soc. Am. J. 60:316–323.

Wells, P.V. 1970. Postglacial vegetational history of the Great Plains. Science 167:1574–1582.

West, L.T., L.P. Wilding, J.K. Landeck, and F.G. Calhoun. 1984. Soil survey of the ICRISAT Sahelian Center, Niger, West Africa. Texas A&M University, College Station, TX.

Wiecek, C.S., and A.S. Messenger. 1972. Calcite contributions by earthworms to forest soils in northern Illinois. Soil Sci. Soc. Am. Proc. 36:478–480.

Yerima, B.P.K., L.P. Wilding, C.T. Hallmark, and F.G. Calhoun. 1989. Statistical relationships among selected properties of northern Cameroon Vertisols and associated Alfisols. Soil Sci. Soc. Am. J. 53:1758–1763.

33.12 Ultisols

Larry T. West

Joey N. Shaw

Fredrich H. Beinroth

33.12.1 Introduction

Ultisols are a group of soils with an argillic or kandic horizon and low base saturation in lower subsoil horizons. These soils may be found on a variety of parent materials and under a range of climatic conditions, though most have developed under forest vegetation in humid climates. Ultisols are generally considered to be less productive than soils in many of the other orders, but in many regions of the world, they are the most productive soils available. With proper management of organic residues, fallow periods, and/or chemical inputs, the productivity of this resource can be enhanced and maintained. This section presents a brief overview of Ultisol genesis and properties. For a more comprehensive review, the reader is referred to Miller (1983) and West et al. (1997).

33.12.2 Classification: Historical and Current

As soil classification has evolved in the United States, soils currently considered to be Ultisols have been included with the Red soils, Yellow soils, or Lateritic soils (Marbut, 1928), the Yellow Podzolic or Red Podzolic great soil groups (Baldwin et al., 1938), and the Red–Yellow podzolic great soil group (Kellogg, 1949). The order Ultisols was introduced in the *7th Approximation* (Soil Survey Staff, 1960) to include soils with an argillic horizon and base saturation <35% in lower subsoil horizons. The 35% base saturation criteria was placed to separate soils where basic cations are maintained in the root zone through biocycling, and soil amendments are commonly needed for sustained productivity (Ultisols), from soils that replenish a significant amount of basic cations through weathering (Alfisols) (Forbes, 1986). Since first publication of *Soil Taxonomy* (Soil Survey Staff, 1975), the classification of Ultisols has been revised to include kandic horizons as criteria for placement in the order and to include soils in any temperature regime. For a comprehensive definition of Ultisols, the reader is referred to the *Keys to Soil Taxonomy* (Soil Survey Staff, 2010). Table 33.53 lists current orders, suborders, and great groups in Ultisols (Soil Survey Staff, 2010).

33.12.3 Geographic Distribution

Worldwide, Ultisols are currently estimated to cover about 11,054,000 km² which is 8.1% of the ice-free global landmass (Table 33.53; Figure 33.76). About 80% of Ultisols are in tropical regions, and about 18% of the tropics is covered by Ultisols (Eswaran, 1993). About 50% of Ultisols are Udults, 35% Ustults, and 12% Aquults with Humults and Xerults comprising the

TABLE 33.53   Listing of Suborders, Great Groups and Subgroups in the Ultisols Order

Suborder	Great Group	Subgroup
Aquults	Plinthaquults	Kandic, Typic
	Fragiaquults	Aerie Plinthic, Umbric, Typic
	Albaquults	Vertic, Kandic, Aerie, Typic
	Kandiaquults	Acraquoxic, Arenic Plinthic, Arenic Umbric, Arenic, Grossarenic, Plinthic, Aerie, Umbric, Typic
	Kanhaplaquults	Aquandic, Plinthic, Aerie Umbric, Aerie, Umbric, Typic
	Paleaquults	Vertic, arenic plinthic, arenic umbric. Arenic, grossarenic, plinthic, aerie, umbric, typic
	Umbraquults	Plinthic, Typic
	Epiaquults	Vertic, Aerie Fragic, Fragic, Arenic, Grossarenic, Aerie, Typic
	Endoaquults	Arenic, Grossarenic, Aerie, Typic
Humults	Sombrihumults	Typic
	Plinthohumults	Typic
	Kandihumults	Andic Ombroaquic, Ustandic, Andic, Aquic, Ombroaquic, Plinthic, Ustic, Xeric, Anthropic, Typic
	Kanhaplohumults	Lithic, Ustandic, Andic, Aquic, Ombroaqic, Ustic, Xeric, Anthropic, Typic
	Palehumults	Aquandic, Andic, Aquic, Plinthic, Oxyaquic, Ustic, Xeric, Typic
	Haplohumults	Lithic, Aquandic, Aquic, Andic, Plinthic, Oxyaquic, Ustic, Xeric, Typic
Udults	Plinthudults	Typic
	Fragiudults	Arenic, Plinthaquic, Glossaquic, Aquic, Plinthic, Glossic, Humic, Typic
	Kandiudults	Arenic Plinthaquic, Aquic Arenic, Arenic Plinthic, Arenic Rhodic, Arenic, Grossarenic Plinthic, Grossarenic, Acrudoxic Plinthic, Acrudoxic, Plinthaquic, Aquandic, Andic, Aquic, Plinthic, Ombroaquic, Oxyaquic, Sombric, Rhodic, Typic
	Kanhapludults	Iithic, Fragiaquic, Arenic Plinthic, Arenic, Acrudoxic, Plinthaquic, Andic, Aquic, Ombroaquic, Oxyaquic, Fragic, Plinthic, Rhodic, Typic
	Paleudults	Vertic, Spodic, Arenic Plinthaquic, Aquic Arenic, Plinthaquic, Fragiaquic, Aquic, Anthraquic, Oxyaquic, Lamellic, Arenic, Plinthic, Psammentic, Grossarenic Plinthic, Plinthic, Arenic Rhodic, Arenic, Grossarenic, Fragic, Rhodic, Typic
	Rhoduldults	Lithic, Psammentic, Typic
	Hapludults	Lithic Ruptic-Entic, Lithic, Vertic, Fragiaquic, Aquic Arenic, Aquic, Fragic, Oxyaquic, Lamellic, Psammentic, Arenic, Grossarenic, Inceptic, Humic, Typic
Ustults	Plinthustults	Haplic, Typic
	Kandiustults	Acrustoxic, Aquic, Arenic Plinthic, Arenic, Udandic, Andic, Plinthic, Aridic, Udic, Rhodic, Typic
	Kanhaplustults	Lithic, Acrustoxic, Aquic, Arenic, Udandic, Andic, Plinthic, Ombroaquic, Aridic, Udic, Rhodic, Typic
	Paleustults	Typic
	Rhodustults	Lithic, Psammentic, Typic
	Haplustults	Lithic, Petroferric, Aquic, Arenic, Ombroaquic, Plinthic, Kanhaplic, Typic
Xerults	Palexerults	Aquandic, Aquic, Andic, Typic
	Haploxerults	Lithic Ruptic-Lnceptic, Lithic, Aquic, Andic, Lamellic, Psammentic, Arenic, Grossarenic, Typic

Source: Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS. US Government Printing Office, Washington, DC.
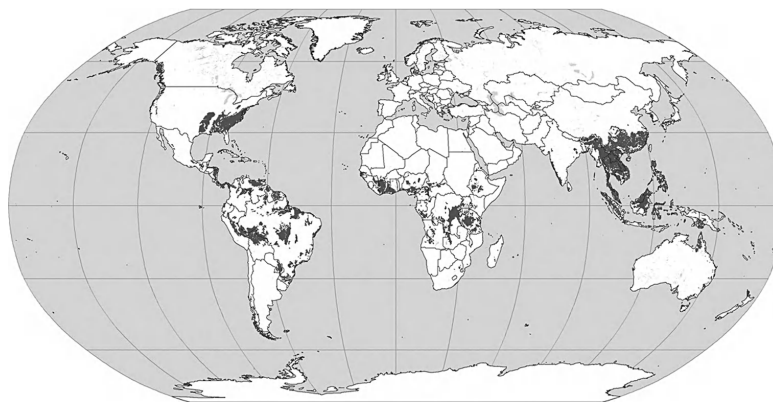

FIGURE 33.76   Global distribution of Ultisols. (Courtesy of USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC, 1998.)

TABLE 33.54 Distribution of Ultisols in the Conterminous United States and Puerto Rico

State	Area (km²)	Percentage of Total
Alabama	97,965	11.4
Arkansas	68,644	8.0
California	8,418	1.0
Delaware	3,421	0.4
Florida	28,106	3.3
Georgia	114,802	13.3
Idaho	17	0.0
Indiana	3,158	0.4
Illinois	44	0.0
Kansas	79	0.0
Kentucky	27,402	3.2
Louisiana	23,395	2.7
Massachusetts	22	0.0
Maryland	16,004	1.9
Missouri	28,534	3.3
Mississippi	49,826	5.8
North Carolina	85,057	9.9
New Jersey	7,617	0.9
New York	579	0.1
Oregon	11,195	1.3
Ohio	7,137	0.8
Oklahoma	16,784	1.9
Pennsylvania	39,375	4.6
Puerto Rico	1,784	0.2
South Carolina	54,720	6.3
Tennessee	44,317	5.1
Texas	24,967	2.9
Virginia	69,910	8.1
Washington	3,574	0.4
West Virginia	25,435	2.9
Total	862,288	100.0

Source: Courtesy of USDA-NRCS, Soil Survey Division, National Soil Survey Center, Washington, DC.

remaining 3% (Table 33.53). In the United States, Ultisols cover about 860,000 km² (Table 33.54) and occur in 30 states and Puerto Rico (Figure 33.77). The largest concentration of Ultisols in the United States is in the east central, southeast, and south central parts of the country. Ultisols are also extensive on the older islands of Hawaii, along the west coast, and in unglaciated regions of the northeast and north central parts of the country.

33.12.4 Genesis of Ultisols

Soil as an independent natural body whose properties are a function of five soil-forming factors: local climate, parent materials, organisms, relief, and age (Dokuchaev, 1883; Jenny, 1941), is a unifying philosophy in pedology. As such, Ultisol genesis will be briefly discussed in terms of these factors.

33.12.4.1 Climate

The definition of Ultisols implies two rainfall/evapotranspiration conditions. First, there must be some time during the year when evapotranspiration exceeds precipitation, as this appears to be a prerequisite for the formation of an argillic horizon (van Wambeke, 1991). Second, precipitation must exceed the capacity of the soil to retain water during some time of the year, so that water percolates through the solum to remove basic cations and maintain low base saturation (<35% by sum of cations [pH 8.2] which corresponds roughly to 50% base saturation measured at pH 7.0) in the subsoil (Miller, 1983). Most Ultisols occur in regions with mean annual air temperatures above 6°C. Thus, the climate conducive to the formation of Ultisols is typically humid tropical or warm temperate. As many Ultisols have evolved over long time periods, effects of paleoclimate on Ultisol development cannot be ignored.

33.12.4.2 Parent Material

The definition of Ultisols requires the presence of either an argillic or a kandic horizon, both of which contain silicate clays. Therefore, the parent material of Ultisols must be one that contains either phyllosilicates or primary minerals that can weather to produce them. As the vast majority of rocks meet one or both of these criteria, a wide variety of geologic formations are capable of producing parent materials suitable for Ultisol formation. Residual parent materials that have weathered in situ are usually saprolite. In contrast, transported sediments that form Ultisol parent materials have frequently been preweathered and may have gone through more than one weathering cycle (Fiskell and Perkins, 1970; Allen and Fanning, 1983). In sedimentary landscapes, parent material has a major influence on Ultisol property distribution (Shaw et al., 2004).

33.12.4.3 Biota

Although Ultisols occur within many tropical and subtropical managed ecosystems, forests are the dominant vegetation on most natural Ultisol landscapes. The processes in forest soils appear to accelerate the formation of argillic horizons through the desiccating effect of tree roots that absorb water, but not colloids in suspension. In Africa and South America, vast areas of Ultisols are under savanna vegetation. It cannot be ascertained, however, whether the soils formed under the present vegetation or under forests in one or more humid paleoclimates.

33.12.4.4 Relief

Ultisols occur on almost all landforms. The one characteristic that Ultisol landforms have in common, however, is geomorphic stability over long periods of time. Geomorphic studies of tropical landscapes invariably show that Ultisols occupy geomorphic positions that are younger and less stable than the surfaces where Oxisols occur, but older and more stable than those of soils in other Orders with which they are geographically associated (Beinroth et al., 1974; Lepsch and Buol, 1974; Beinroth, 1981). In temperate regions where Oxisols are absent, Ultisols
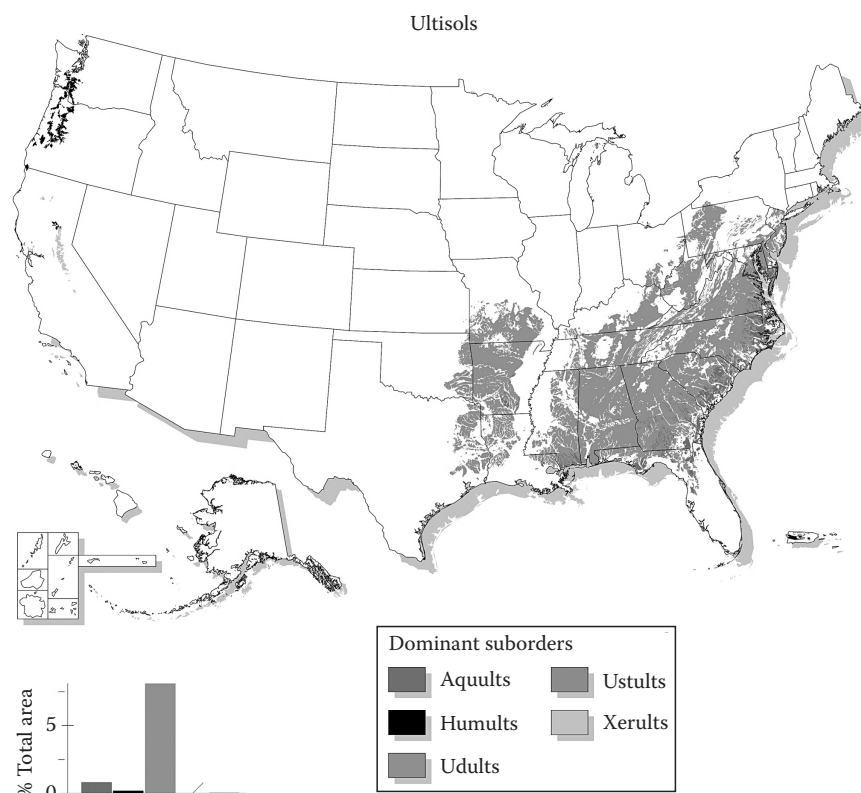
Ultisols

FIGURE 33.77 Distribution of Ultisols in the United States. (Courtesy of USDA-NRCS, Soil Survey Division, National Soil Survey Center, Washington, DC, 2010.)

have developed on the oldest, most stable and highly weathered landscapes such as the central Missouri Ozarks and the coastal plain of the southeastern United States (Scrivner et al., 1966; Daniels et al., 1971).

33.12.4.5 Time

The time required for the formation of Ultisols depends on the other pedogenic factors, notably weatherability, composition, and permeability of the parent materials, landscape stability, and climate fluctuations over time. There is much room for variability in these conditions, and consequently, for the age of Ultisols. The Soil Survey Staff (1999) concluded that formation of the argillic horizon ordinarily requires a few thousand years. This relatively short time may suffice to produce a minimal expression of an argillic horizon. The development of the thick argillic horizons that typify the Pale and Kandi great groups of Ultisols, however, would certainly take longer. As such, age of Ultisols has been reported to range from <12,000 years for soils on preweathered sediments to >15 million years on marine sediments in the coastal plain of the southeast United States (Gamble et al., 1970; Daniels et al., 1971; Miller, 1983).

33.12.5 Morphological Properties

33.12.5.1 Argillic and Kandic Horizons

Other than low base saturation, the one property that distinguishes Ultisols from all but one other soil order (Alfisols,

Section 33.11) is the requirement of a clay increase between surface and subsoil horizons in the form of either an argillic or kandic horizon. This clay increase essentially determines the morphology of Ultisols and influences many chemical and physical properties. The depth, thickness, and clay content of the argillic or kandic horizon, however, can vary widely, and this variation has considerable impact on management and interpretation of these soils. Figure 33.78 illustrates the range in depth and clay content of argillic or kandic horizons that can occur in Ultisols.

33.12.5.2 Soil Color

Color of surface horizons in Ultisols is commonly brown, yellowish brown, dark brown, or red as would be expected for soils with relatively low amounts of soil organic C (SOC). However, in cool climates or poorly drained conditions with slow organic matter decomposition, thick, dark, acidic surface horizons (umbric epipedons) are common. Many well-drained Ultisols have reddish Bt horizons because of the presence of hematite as a pigmenting agent. Yellowish-brown subsoil horizons, also common in Ultisols, lack hematite because either the parent material lacked mineral precursors for hematite, environmental conditions precluded hematite formation (Schwertmann and Taylor, 1989), or hematite was preferentially reduced and lost under short periods of saturation leaving behind more stable goethite (Macedo and Bryant, 1989; Dobos et al., 1990). Under extended periods of saturation and reduction, Ultisols will develop redox

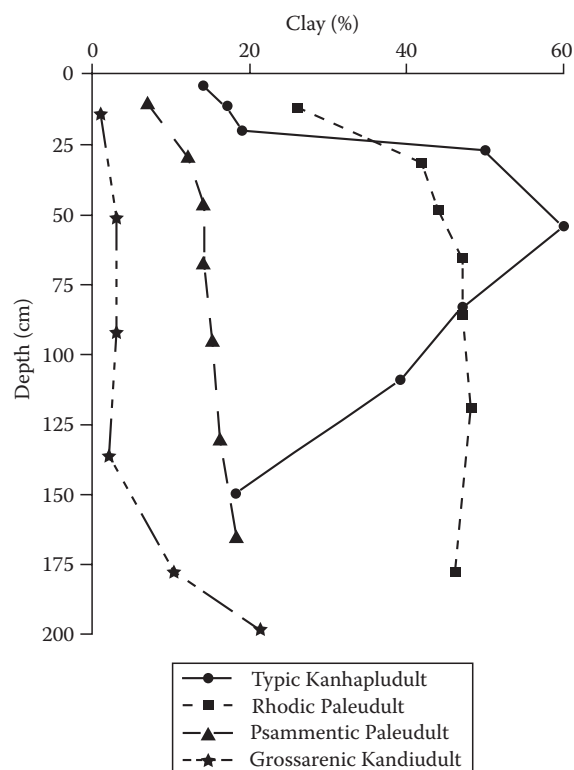FIGURE 33.78 Clay distribution with depth for four subgroups of Ultisols.

concentrations and depletions or gray matrices as Fe is reduced and mobilized.

33.12.5.3 Plinthite

Plinthite is defined in *Soil Taxonomy* as an Fe-rich, humus-poor mixture of clay with quartz and other diluents (Soil Survey Staff, 2010). This material most often occurs in lower subsoil horizons, is weakly indurated, and is normally low in basic cations and primary silicate minerals other than quartz. Kaolinite is commonly the only clay mineral present with abundance. The reader is referred to van Wambeke (1991) for a complete discussion of the genesis and kinds of plinthite and petroplinthite (or ironstone).

From a pedological perspective, plinthite's relationship with contemporary soil saturation is not well understood, but it does have significance in terms of movement of water through soils. Daniels et al. (1978) suggested that 10% platy plinthite would perch water. A similar amount of nodular plinthite was not considered to be water restrictive, but horizons subjacent to the plinthic horizon did restrict water movement. Movement of water and solutes through horizons containing plinthite is primarily restricted to areas of redox depletions (Carlan et al., 1985; Blume et al., 1987; Shaw et al., 1998). Because of water perching by plinthic or subjacent horizons, plinthite in soils suggests that water and solutes are moving from high to low landscape positions as shallow subsurface flow (Hubbard and Sheridan, 1983; Shirmohammadi et al., 1984).

33.12.6 Mineralogical Properties

33.12.6.1 Clay Minerals

The clay fraction of subsoil horizons of most Ultisols is dominated by kaolinite because of the stability of kaolinite when compared with other phyllosilicates (Jackson et al., 1948; Allen and Fanning, 1983). In addition, gibbsite, K mica, K feldspar, and amorphous silica may transform directly to kaolinite without passing through intermediates (Garrels and Christ, 1965). As such, kaolinite in soils has been shown to be an early weathering product of biotite and K feldspar in acid gneiss and schist parent materials (Robertus and Buol, 1985; Robertus et al., 1986; Norfleet and Smith, 1989).

In addition to kaolinite, hydroxy Al-interlayered vermiculite is often abundant in A and E horizons of Ultisols and may be a major component of upper Bt horizons of these soils (Bryant and Dixon, 1964; Fiskell and Perkins, 1970; Carlisle and Zelazny, 1973; Harris et al., 1980; Karathanasis et al., 1983). Studies in the Piedmont and Coastal Plain of Alabama (southeastern United States) indicate that hydroxy Al-interlayered vermiculite is more abundant than kaolinite in the upper solum of many Ultisols (Shaw et al., 2010). Appreciable amounts of mica, smectite, and other 2:1 minerals have also been reported to occur in Ultisols (Nash, 1979; Harris et al., 1984; Karathanasis et al., 1986; Nash et al., 1988). The 2:1 minerals in these Ultisols were inherited from the parent materials instead of forming in situ from primary minerals or other phyllosilicates.

A common but minor mineral component of many Ultisols is gibbsite (Carlisle and Zelazny, 1973; Hajek and Zelazny, 1985). Because this mineral is an end product of silicate mineral weathering (Jackson et al., 1948), minor amounts should be expected to be found in Ultisols. Clay fractions of Ultisols developed from acid gneiss and schist in the southeastern United States, however, have been reported to contain as much as 70% gibbsite formed as a product of K feldspar weathering at the rock–saprolite interface (Losche et al., 1970; McCracken et al., 1971; Norfleet and Smith, 1989). Minor amounts of halloysite and Ti and Mn oxides are commonly present in Ultisols (Figure 33.79).

Iron oxides and oxyhydroxides are also common, but minor components in the clay fraction of Ultisols. Goethite is the most common Fe oxide mineral in most Ultisols, but many well-drained Ultisols have appreciable amounts of hematite. Analyses of 13 horizons from kandic and/or argillic horizons of Ultisols sampled within the Alabama (southeastern United States) Piedmont and Coastal Plain indicated that hematite composed 0%–11.4% of the clay (<2 μm) fraction, while goethite composed 0%–13.7% (Shaw, 2001). Hematite formation is favored by warm temperatures and distinct dry seasons (Schwertmann and Taylor, 1989), which are environmental conditions commonly associated with Ultisols.

33.12.6.2 Sand and Silt Minerals

The sand and silt fraction of Ultisols commonly is dominated by quartz, and many Ultisols have >90% quartz and other resistant minerals in their sand and silt separates. However, weakly to moderately developed Ultisols may have appreciable quantities of mica, feldspars, and other weatherable minerals in their

(a)   (b)
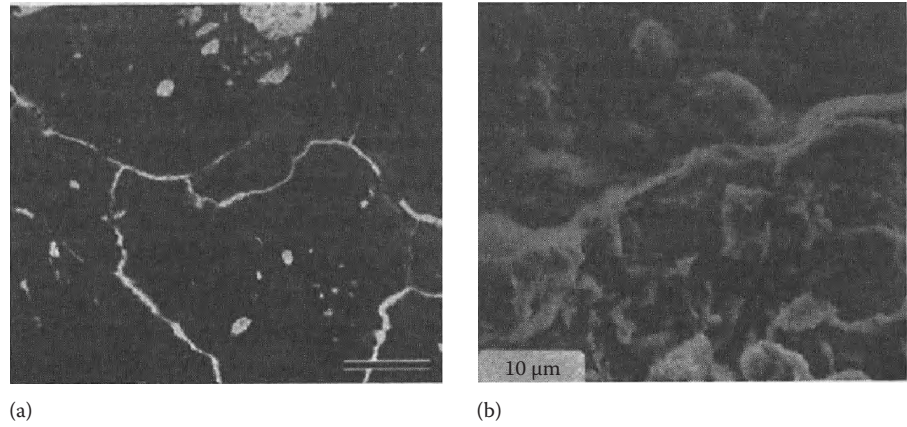
FIGURE 33.79   (a) Thin section of micrograph of clay coatings in Btv2 horizon of a Plinthic Kandiudult. Plane-polarized light. Bar length = 0.5 mm; (b) scanning electron micrograph of clay coating in 2Btx1 horizon of a Typic Fragiudult. (Original micrograph from Michael Thompson, Iowa State University, Ames, IW. With permission.)

coarse fraction (Robertus et al., 1986; Norfleet and Smith, 1989; Bockheim et al., 1996).

33.12.7 Physical Properties

33.12.7.1 Bulk Density, Coefficient of Linear Extensibility, Water Retention, and Hydraulic Conductivity (*K*)

Bulk density, COLE, water retention, and hydraulic conductivity of Ultisols vary considerably depending on soil and horizon properties including particle size distribution, SOC content, mineralogy, macroporosity, presence of fragipans, parent material, presence of ironstone and plinthite, and, for surface horizons, management.

A horizons of uncultivated Ultisols often have lower bulk density than subsoil horizons, but with tillage and the accompanying loss of organic matter, compaction, and aggregate breakdown, the A horizon may become as dense or more dense than underlying Bt horizons (Table 33.55; Hubbard et al., 1985). Shrink/swell potential for kaolinitic Ultisols commonly is low reflecting the abundance of kaolinite and other low activity minerals (Table 33.53; Hubbard et al., 1985; Southard and Buol, 1988a), but COLE values up to 0.157 have been reported for Bt horizons from Ultisols with smectitic mineralogy (Karathanasis and Hajek, 1985; Griffin and Buol, 1988). However, smectitic Ultisols with high shrink/swell potential often lack wide cracks common to Vertisols and Vertic subgroups of other orders (Karathanasis and Hajek, 1985).

TABLE 33.55   Ranges in Selected Physical Properties of Ultisols

Property	Value	References
Bulk density (Mg m^{-3})		
A horizon	0.54–1.85	Bruce et al. (1983)
Bt horizon	1.05–1.70	Hubbard et al. (1985)
		Quisenberry et al. (1987)
		Southard and Buol (1988a)
		Norfleet and Smith (1989)
COLE (Bt horizons)	0.006–0.157	Karathanasis and Hajek (1985)
		Griffin and Buol (1988)
		Southard and Buol (1988a)
Water retention (cm^3 cm^{-3})		
A horizons		Bruce et al. (1983)
−10 kPa	0.13–0.21	Hubbard et al. (1985)
−30 kPa	0.09–0.18	Quisenberry et al. (1987)
−1500 kPa	0.05–0.09	
Bt horizons		
−10 kPa	0.20–0.42	
−30 kPa	0.17–0.40	
−1500 kPa	0.09–0.33	
Saturated hydraulic conductivity (μm s^{-1})		
Bt horizons	0.01–13.1	Hubbard et al. (1985)
Btv horizons	0.3–1.3	Southard and Buol (1988b)

TABLE 33.56 Mean Pore Area for Dyed and Undyed Areas after Breakthrough of Methylene-Blue for Ultisols from the Piedmont and Coastal Plain of Georgia

Horizon	Mean Pore Area[a] (%)	
	Dyed	Undyed
Typic Kanhapludult (Piedmont)		
A	19.0	3.0
BA	17.7	5.0
Bt	15.8	2.4
Bt	14.4	2.8
Plinthic Kandiudult (Coastal Plain)		
Bt	5.0	1.1
Btv	5.9	0.8
BC	NA[b]	0.9

Sources: Shaw et al., 1997; Franklin, D.H., L.T. West, D.E. Radcliffe, and P.F. Hendrix. 2007. Characteristics and genesis of preferential flow paths in a piedmont ultisol. Soil Sci. Soc. Am. J. 71:752–758.

[a] Area of pores >0.05 mm equivalent circular diameter determined from impregnated polished blocks with image analysis.

[b] The BC horizon had insufficient dyed area for evaluation of pore area.

Ultisols in fine-loamy or fine particle size families would generally be considered capable of retaining and supplying sufficient moisture for plant growth in humid climates (Table 33.55). However, water often becomes a limiting factor for crop growth and yield for Ultisols in sandy or coarse-loamy particle size families, lithic subgroups, or that have thick sandy epipedons (Arenic and Grossarenic subgroups).

Most studies of hydraulic properties of Ultisols have found macroporosity to be the best predictor of saturated hydraulic conductivity (Ks) (O'Brien and Buol, 1984; Southard and Buol, 1988b). Preferential flow paths in Ultisols in the Piedmont and Coastal Plain of Georgia have textures similar to the matrix of the horizon, but the preferential flow zones have a significantly greater number and area of pores >0.05 mm diameter (Table 33.56; Shaw et al., 1998; Franklin et al., 2007). Formation of coarse pores in Ultisols is most often attributed to microfabric alteration and channel formation by roots and burrowing organisms (O'Brien and Buol, 1984; Southard and Buol, 1988b; Shaw et al., 1998; Franklin et al., 2007).

33.12.7.2 Infiltration and Surface Crusting

In many Ultisols, one factor affecting infiltration of water is formation of a surface crust or seal. For 25 Ultisols from the southeastern United States tested under 50–90 mm h^{-1} simulated rainfall, infiltration rate decreased rapidly during the first 10–20 mm of the rainfall event, and the infiltration rate at the end of rainfall was less than 10 mm h^{-1} for all of the soils (Miller and Radcliffe, 1992). In general, Ultisols in this region with sandy loam texture and high amounts of water dispersible clay are most prone to crusting. Soils higher in clay tend to take longer to form a surface crust, and sands and loamy sands form only weakly expressed crusts (Miller and Bahruddin, 1986; Radcliffe et al., 1991; Miller and Radcliffe, 1992; Chiang et al., 1993).

Because energy inputs from rainfall are needed to form surface crusts, crusting is minimal if the soils have residue or vegetative cover and organic matter is maintained in the upper part of the soil. For Ultisols in the Piedmont of Georgia, infiltration rates after 1 h of simulated rainfall were about 40% higher for soils that had been under no-tillage for 5 years than for soils that had been conventionally tilled (Table 33.57; Bruce et al., 1992).

33.12.8 Chemical Properties

The requirement that Ultisols have base saturation <35% in the lower part of the subsoil carries with it certain associated properties including low pH, potentially high Al saturation, appreciable weathering and associated kaolinitic mineralogy, and in many cases, relatively high contents of Fe and Al oxides and oxyhydroxides. This suite of properties has significant impact on the chemical properties of Ultisols and how these properties affect use and management of this soil order.

33.12.8.1 Cation and Anion Exchange Capacity

The CEC of Ultisol horizons depends on the amount and type of clay, Fe, Al, and Mn oxide content, organic matter content,

TABLE 33.57 Comparison of Organic C, Aggregate Stability, Infiltration Rate, and Percentage Moist Days under Conventional and No-Till Cropping Systems

Cropping System[a]	Organic C (g kg^{-1})	Aggregate Stability (g kg^{-1})	Infiltration Rate[a]		Moist Days,[b] % of Growing Season
			Residue (mm h^{-1})	Residue Removed (mm h^{-1})	
CTG	10.4	580	36	22	29
NTG	23.3	890	50	46	49

Sources: West, L.T., W.P. Miller, G.W. Langdale, R.R. Bruce, J.M. Laflen, and A.W. Thomas. 1991. Cropping system effects on interrill soil loss in the Georgia Piedmont. Soil Sci. Soc. Am. J. 55:460–466; Bruce, R.R., G.W. Langdale, L.T. West, and W.P. Miller. 1992. Soil surface modification by biomass inputs affecting rainfall infiltration. Soil Sci. Soc. Am. J. 56:1614–1620.

CTG, conventional-tilled grain sorghum (*Sorghum bicolor* (L.) Moench); NTG, no-till grain sorghum with clover winter cover.

[a] Infiltration rate at the end of a 1 h of rainfall simulation at approximately 64 mm h^{-1}.

[b] Moist days are defined as the days with soil moisture tension >0.1 MPa.

soil solution electrolyte concentration, and pH of the horizon. Reported CEC values for Ultisol horizons range from <3 to >20 cmol$_c$ kg^{-1} (Sanchez, 1976; Carlisle et al., 1985; Karathanasis et al., 1986). Generally, Ultisols with kaolinitic mineralogy have mostly pH-dependent charge and low CEC, while those with mixed or smectitic mineralogy have a greater proportion of fixed charge and higher CEC values.

Even when the net charge is negative, many Ultisols have an appreciable positive charge and anion exchange capacity (AEC). Gillman and Sumner (1987), in a study of four Ultisols from the Piedmont of Georgia, reported AEC values at the pH of the soil that ranged from 0.1 to 1.2 cmol$_c$ kg^{-1}, which was attributed to Fe and Al oxides in these soils. For Ultisols from Georgia and South Africa, Grove et al. (1982) reported AEC values ranging from 0.03 to 1.91 cmol$_c$ kg^{-1}. Measurable AEC has also been reported for Ultisols from tropical Australia and Peru (Gillman and Sumpter, 1986; Gillman and Sinclair, 1987).

33.12.8.2 Acidity and Exchangeable Aluminum

Below a pH of about 5.5, Al released from weathering of primary and secondary minerals is present as part of the exchange complex. An Al saturation of greater than 60% (measured as a percentage of the unbuffered CEC or ECEC) has been reported as the threshold that results in soil solution Al concentrations >1 mg L^{-1}, which may reduce yield in Al-sensitive crops (Nye et al., 1961; Kamprath, 1970; Sanchez, 1976). Farina and Channon (1991), however, observed yield reductions in maize for Al saturations >25% of the ECEC. Yields are also reduced at low pH due to Mn toxicity, and Ca, Mg, and Mo deficiency (Adams, 1984).

33.12.8.3 Phosphorus

Phosphate is specifically adsorbed by Fe, Al, and Mn oxides and amorphous or poorly crystalline aluminosilicates. Because of eluviation and weathering, surface horizons of most Ultisols have low contents of these components. Thus, P sorption is generally lower in Ultisols than in Oxisols and Andisols (Sanchez, 1976; van Wambeke, 1991). However, differences in crystallinity of Fe oxides among soils may confound this relationship within and among orders (Pratt et al., 1969; Fox et al., 1971).

33.12.9 Biological Properties

33.12.9.1 Organic Matter

Ultisols are commonly a fragile soil resource with low inherent soil quality (Norfleet et al., 2003). However, because these soils often exist in regions with long productive growing seasons and have favorable soil physical properties, Ultisols are often utilized for food, fiber, and timber production. Conservation management of these soils is critical for maximizing and sustaining productivity of agronomic systems. Numerous studies have documented the relationships between management, soil quality, productivity, and SOC in these soils (Reeves, 1997). A common misconception is that Ultisols have low contents of SOC compared with other soil orders. While Ultisols generally have lower amounts of SOC than Mollisols, Oxisols, and Andisols,

SOC levels in Ultisols are similar to those in Alfisols both within and across geographic areas (Sanchez, 1976). There is little difference in SOC content between Ultisols in temperate and tropical climates (Buol, 1973; Sanchez, 1976).

Cultivation of Ultisols has decreased SOC levels compared with relatively undisturbed ecosystems, and SOC contents decrease when these soils are converted from forest or pasture to row crop lands. Studies of Georgia (southeastern United States) Ultisols have found that organic matter content in surface horizons decreased from 20.5 to 16.6 g kg^{-1} 2 years after conversion from forest to intensive cultivation (Giddens, 1957), while soils in relatively undisturbed longleaf pine (*Pinus palustris* Miller) ecosystems in the Coastal Plain had 64% more SOC (relative) (0–30 cm) than conventional row crop lands (Levi et al., 2010). The use of conservation tillage, cover crops, and residue management (conservation systems) in row crop production can increase Ultisol SOC levels compared with conventional systems (Reeves, 1997). Bruce et al. (1992) reported that conservation systems increased SOC in the upper 15 mm of degraded Ultisols from 10.4 to 23.3 g kg^{-1} over a 5 year period. An investigation of 87 sites consisting mostly of Ultisols in the southeastern United States found SOC (0–20 cm) to be highest in pastures (38.9 Mg ha^{-1}), followed by conservation row crop systems (27.9 Mg ha^{-1}), and lastly, conventional row crop systems (22.2 Mg ha^{-1}) (Causarano et al., 2008). In addition, conservation system use in southeastern U.S. cotton (*Gossypium hirsutum* L.) production has been found to increase SOC quantities by 0.48 ± 0.56 Mg C ha^{-1} year^{-1} compared with conventional systems (Schwab et al., 2002; Causarano et al., 2006). Thus, Ultisols can play a significant role in managing SOC sequestration.

Increasing SOC in Ultisols significantly increases water stable aggregates and decreases runoff and interrill soil loss (Table 33.58; West et al., 1991, 1992; Bruce et al., 1992). Similarly, decreased runoff (Truman et al., 2005) and water dispersible clay (Shaw et al., 2003) were found for Alabama (southeastern United States) Coastal Plain Ultisols under conservation management with higher SOC quantities.

33.12.9.2 Biologic Populations and Processes

Populations of microorganisms in Ultisols are not appreciably different from soils in other orders. Surface horizons of most

TABLE 33.58 Interrill Soil Loss and Runoff from Three Ultisols with Varying Tillage and Surface Cover under 50 mm h^{-1} Rainfall Intensity for 1 h on 5%–9% Slopes

Tillage	Surface Condition	Interrill Soil Loss (g m^{-2})			Total Runoff (mm)		
		Soil 1	Soil 2	Soil 3	Soil 1	Soil 2	Soil 3
Conventional	Crusted	0.46	0.56	0.33	16	23	17
	Tilled	0.72	0.79	0.48	18	23	16
No-till	Bare	0.14	0.19	0.07	3	3	6
	Residue	0.07	0.03	0.02	1	1	1

Source: West, L.T., W.P. Miller, G.W. Langdale, R.R. Bruce, J.M. Laflen, and A.W. Thomas. 1991. Cropping system effects on interrill soil loss in the Georgia Piedmont. Soil Sci. Soc. Am. J. 55:460–466.

Ultisols have acid pH values. Thus, fungi may comprise a greater proportion of the microbial biomass in these soils because of the greater ability of fungi to survive under acid conditions than other microorganisms (Alexander, 1977). Low pH has been reported to reduce rates of nitrification and denitrification in pure cultures (Alexander, 1977). Natural adaptation by microorganisms to local field conditions, however, probably alters the effects of pH and other environmental conditions observed in the laboratory.

33.12.10 Conclusions

Worldwide, Ultisols are a widespread soil resource, which because of their extensive weathering, low base saturation, and associated properties, are often considered to be unproductive. It is true that Ultisols, in general, have lower native fertility than Mollisols and Alfisols, and in areas where these orders are abundant, Ultisols are the less desirable soils. In many regions of the world, however, Ultisols are dominant or are the most productive soils available. Only with a thorough understanding of the genesis, properties, and response to management of these soils can the productivity of this valuable resource be maintained and enhanced.

References

Adams, F. 1984. Crop response to lime in the Southern United States, p. 211–266. *In* F. Adams (ed.) Soil acidity and liming. 2nd edn. Agronomy Number 12. ASA, Madison, WI.

Alexander, M.A. 1977. Introduction to soil microbiology. 2nd edn. John Wiley & Sons, New York.

Allen, B.L., and D.S. Fanning. 1983. Composition and soil genesis, p. 141–192. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy I. Concepts and interactions. Elsevier, Amsterdam, the Netherlands.

Baldwin, M., C.E. Kellogg, and J. Thorp. 1938. Soil classification, p. 979–1001. *In* Soils and men. USDA yearbook of agriculture. U.S. Government Printing Office, Washington, DC.

Beinroth, F.H. 1981. Some highly weathered soils of Puerto Rico, 1. Morphology, formation and classification. Geoderma 27:1–73.

Beinroth, F.H., G. Uehara, and H. Ikawa. 1974. Geomorphic relationships of oxisols and ultisols on Kauai, Hawaii. Soil Sci. Soc. Am. Proc. 38:128–131.

Blume, L.J., H.F. Perkins, and R.K. Hubbard. 1987. Subsurface water movement in an upland coastal plain soil as influenced by plinthite. Soil Sci. Soc. Am. J. 51:774–779.

Bockheim, J.G., J.G. Marshall, and H.M. Kelsey. 1996. Soilforming processes and rates on uplifted marine terraces in southwestern Oregon, USA. Geoderma 73:39–62.

Bruce, R.R., J.J. Dane, V.L. Quisenberry, N.L. Powell, and A.W. Thomas. 1983. Physical characteristics of soils in the southern region: Cecil. Southern Cooperative Series Bulletin No. 267.

Bruce, R.R., G.W. Langdale, L.T. West, and W.P. Miller. 1992. Soil surface modification by biomass inputs affecting rainfall infiltration. Soil Sci. Soc. Am. J. 56:1614–1620.

Bryant, J.P., and J.B. Dixon. 1964. Clay mineralogy and weathering of red–yellow podzolic soil from quartz mica schist in the Alabama Piedmont. Clays Clay Miner. 12:509–521.

Buol, S.W. 1973. Soil genesis, morphology, and classification, p. 1–38. *In* P.A. Sanchez (ed.) A review of soils research in tropical Latin America. North Carolina Agriculture Experimental Station Technical Bulletin No. 219.

Carlan, W.L., H.F. Perkins, and R.A. Leonard. 1985. Movement of water in a Plinthic Paleudult using a bromide tracer. Soil Sci. 139:62–66.

Carlisle, V.W., M.E. Collins, F. Sodek, III, and L.C. Hammond. 1985. Characterization data for selected Florida soils. Soil Science Research Report 85-1. University of Florida, Gainesville, FL.

Carlisle, V.W., and L.W. Zelazny. 1973. Mineralogy of selected Florida Paleudults. Soil Sci. Soc. Fla. Proc. 33:136–139.

Causarano, H.J., A.J. Franzluebbers, D.W. Reeves, and J.N. Shaw. 2006. Soil organic carbon sequestration in cotton production systems of the Southeast USA: A review. J. Environ. Qual. 35:1374–1383.

Causarano, H.J., A.J. Franzluebbers, J.N. Shaw, D.W. Reeves, R.L. Raper, and C.W. Wood. 2008. Soil organic carbon fractions and aggregation in the Southern Piedmont and Coastal Plain. Soil Sci. Soc. Am. J. 72:221–230.

Chiang, S.C., D.E. Radcliffe, and W.P. Miller. 1993. Hydraulic properties of surface seals in Georgia soils. Soil Sci. Soc. Am. J. 57:1418–1426.

Daniels, R.B., E.E. Gamble, and J.G. Cady. 1971. The relation between geomorphology and soil morphology and genesis. Adv. Agron. 23:51–88.

Daniels, R.B., H.F. Perkins, and E.E. Gamble. 1978. Morphology of discontinuous phase plinthite and criteria for its field identification in the southeastern United States. Soil Sci. Soc. Am. J. 42:944–949.

Dobos, R.R., E.J. Ciolkosz, and W.J. Waltman. 1990. The effect of organic carbon, temperature, time and redox conditions on soil color. Soil Sci. 150:506–512.

Dokuchaev, V.V. 1883. Russian Chernozem (Russkii chernozem). [Translated from Russian by N. Kaner.] Israel Program for Scientific Translations, 1967. Available from U.S. Department Commerce, Springfield, VA.

Eswaran, H. 1993. Assessment of global resources: Current status and future needs. Pedologie 43:19–39.

Farina, M.P.W., and P. Channon. 1991. A field comparison of lime requirement indices for maize. Plant Soil 134:127–135.

Fiskell, J.G.A., and H.F. Perkins. 1970. Selected coastal plain soil properties. University of Florida Southern Cooperative Series Bulletin No. 148. University of Florida, Gainesville, FL.

Forbes, T.R. (ed.). 1986. The Guy Smith interviews: Rationale for concepts in soil taxonomy. Soil Management Support Services, Technical Monograph No. 11. USDA-Soil Conservation Service and Cornell university, Ithaca, NY.

Fox, R.L., S.M. Hasan, and R.C. Jones. 1971. Phosphate and sulfate sorption by latosols. Proc. Int. Symp. Soil Fert. Eval. 1:857–864.

Franklin, D.H., L.T. West, D.E. Radcliffe, and P.F. Hendrix. 2007. Characteristics and genesis of preferential flow paths in a piedmont ultisol. Soil Sci. Soc. Am. J. 71:752–758.

Gamble, E.E., R.B. Daniels, and W.D. Nettleton. 1970. Geomorphic surfaces and soils in the Black Creek Valley, Johnston County, North Carolina. Soil Sci. Soc. Am. Proc. 34:276–281.

Garrels, R.M., and C.L. Christ. 1965. Solutions, minerals and equilibria. Harper and Row, New York.

Giddens, J.E. 1957. Rate of loss of carbon from Georgia soils. Soil Sci. Soc. Am. Proc. 21:513–515.

Gillman, G.P., and D.F. Sinclair. 1987. The grouping of soils with similar charge properties as a basis for agrotechnology transfer. Aust. J. Soil Res. 25:275–285.

Gillman, G.P., and M.E. Sumner. 1987. Surface charge characterization and soil solution composition of four soils from the Southern Piedmont in Georgia. Soil Sci. Soc. Am. J. 51:589–594.

Gillman, G.P., and A.S. Sumpter. 1986. Surface charge characteristics and lime requirements of soils derived from basaltic, granitic, and metamorphic rocks in high rainfall tropical Queensland. Aust. J. Soil Res. 24:173–192.

Griffin, R.W., and S.W. Buol. 1988. Soil and saprolite characteristics of vertic and aquic hapludults derived from Triassic Basin sandstones. Soil Sci. Soc. Am. J. 52:1094–1099.

Grove, J.H., C.S. Fowler, and M.E. Sumner. 1982. Determination of the charge character of selected acid soils. Soil Sci. Soc. Am. J. 46:32–38.

Hajek, B.F., and L.W. Zelazny. 1985. Problems associated with soil taxonomy mineralogy placement in the nonglaciated humid region, p. 87–93. *In* J.A. Kittrick (ed.) Mineral classification of soils. SSSA Special Publication No. 16. SSSA, Madison, WI.

Harris, W.G., S.S. Iyengar, L.W. Zelazny, J.C. Parker, D.A. Lietzke, and W.J. Edmonds. 1980. Mineralogy of a chronosequence formed in New River alluvium. Soil Sci. Soc. Am. J. 44:862–868.

Harris, W.G., L.W. Zelazny, and J.C. Baker. 1984. Depth and particle size distribution of talc in a Virginia Piedmont Ultisol. Clays Clay Miner. 32:227–230.

Hubbard, R.K., C.R. Berdanier, H.F. Perkins, and R.A. Leonard. 1985. Characteristics of selected upland soils of the Georgia Coastal Plain. USDA-ARS, ARS-37. U.S. Government Printing Office, Washington, DC.

Hubbard, R.K., and J.M. Sheridan. 1983. Water and nitrate-nitrogen losses from a small upland, coastal plain watershed. J. Environ. Qual. 12:291–295.

Jackson, M.L., S.A. Tyler, A.L. Willis, B.A. Bourbeau, and R.P. Pennington. 1948. Weathering sequence of clay-size minerals. I. Fundamental generalizations. J. Phys. Colloid Chem. 52:1237–1260.

Jenny, H. 1941. Factors of soil formation. McGraw-Hill Book Co., New York.

Kamprath, E.J. 1970. Exchangeable aluminum as a criterion for liming leached mineral soils. Soil Sci. Soc. Am. Proc. 34:252–254.

Karathanasis, A.D., F. Adams, and B.F. Hajek. 1983. Stability relationships in kaolinite, gibbsite, and Al-hydroxy interlayered vermiculite soil systems. Soil Sci. Soc. Am. J. 47:1247–1251.

Karathanasis, A.D., and B.F. Hajek. 1985. Shrink–swell potential of montmorillonitic soils in udic moisture regimes. Soil Sci. Soc. Am. J. 49:159–166.

Karathanasis, A.D., G.W. Hurt, and B.F. Hajek. 1986. Properties and classification of montmorillonite-rich Hapludults in the Alabama Coastal Plains. Soil Sci. 14:76–82.

Kellogg, C.E. 1949. Preliminary suggestions for the classification and nomenclature of great soil groups in tropical and equatorial regions. Commonw. Bur. Soil Sci. Tech. Commun. 46:76–85.

Lepsch, I.F., and S.W. Buol. 1974. Investigation in an Oxisol–Ultisol toposequence in S. Paulo State, Brazil. Soil Sci. Am. Proc. 38:491–496.

Levi, M.R., J.N. Shaw, C.W. Wood, S.M. Hermann, E.A. Carter, and Y. Feng. 2010. Land management effects on near-surface soil properties in Southeastern U.S. Coastal Plain ecosystems. Soil Sci. Soc. Am. J. 74:258–271.

Losche, C.K., R.J. McCracken, and C.B. Davey. 1970. Soils of steeply sloping landscapes in the southern Appalachian Mountains. Soil Sci. Soc. Am. Proc. 34:473–478.

Macedo, J., and R.B. Bryant. 1989. Preferential microbial reduction of hematite over goethite in a Brazilian oxisol. Soil Sci. Soc. Am. J. 51:690–698.

Marbut, C.F. 1928. A scheme for soil classification. Proc. 1st Int. Congr. Soil Sci. 4:1–31.

McCracken, R.J., E.J. Penderson, L.E. Aull, C.I. Rich, and T.C. Peele. 1971. Soils of the Hayesville, Cecil and Pacolet series in the southern Appalachian and Piedmont regions. North Carolina State University Southern Cooperative Series Bulletin No. 157.

Miller, B.J. 1983. Ultisols, p. 283–323. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy II. The soil orders. Elsevier, Amsterdam, the Netherlands.

Miller, W.P., and M.K. Bahruddin. 1986. Relationship of soil dispersibility to infiltration and erosion of Southeastern soils. Soil Sci. 142:235–240.

Miller, W.P., and D.E. Radcliffe. 1992. Soil crusting in the southeastern United States, p. 233–266. *In* M.E. Sumner and B.A. Stewart (eds.) Soil crusting: Chemical and physical processes. Lewis Publishers, Boca Raton, FL.

Nash, V.E. 1979. Mineralogy of soils developed on Pliocene–Pleistocene terraces of the Tombigbee River in Mississippi. Soil Sci. Soc. Am. J. 43:616–623.

Nash, V.E., D.E. Pettry, and M.N. Sudin. 1988. Mineralogy and chemical properties of two ultisols formed in glauconitic sediments. Soil Sci. 145:270–277.

Norfleet, M.L., C.A. Ditzler, R.B. Grossman, J.N. Shaw, and W.E. Puckett. 2003. Soil quality and its relationship to pedology. Soil Sci. 168:149–155.

Norfleet, M.L., and B.R. Smith. 1989. Weathering and mineralogical classification of selected soils in the Blue Ridge Mountains of South Carolina. Soil Sci. Soc. Am. J. 53:1771–1778.

Nye, P., D. Craig, N.T. Coleman, and J.L. Ragland. 1961. Ion exchange equilibrium involving aluminum. Soil Sci. Soc. Am. Proc. 25:14–17.

O'Brien, E.L., and S.W. Buol. 1984. Physical transformations in a vertical soil-saprolite sequence. Soil Sci. Soc. Am. J. 48:354–357.

Pratt, P.F., F.F. Peterson, and C.S. Holzley. 1969. Qualitative mineralogy and chemical properties of a few soils from Sao Paulo, Brazil. Turrialba 19:491–496.

Quisenberry, V.H., D.K. Cassel, J.H. Dane, and J.C. Parker. 1987. Physical characteristics of soils of the Southern Region: Norfolk, Dothan, Wagram, and Goldsboro. Clemson University Southern Cooperative Series Bulletin No. 263.

Radcliffe, D.E., L.T. West, R.K. Hubbard, and L.E. Asmussen. 1991. Surface sealing in Coastal Plains loamy sands. Soil Sci. Soc. Am. J. 55:223–227.

Reeves, D.W. 1997. The role of soil organic matter in maintaining soil quality in continuous cropping systems. Soil Tillage Res. 43:131–167.

Robertus, R.A., and S.W. Buol. 1985. Intermittency of illuviation in Dystrochrepts and Hapludults from the Blue Ridge and Piedmont provinces of North Carolina. Geoderma 36:277–291.

Robertus, R.A., S.B. Weed, and S.W. Buol. 1986. Transformations of biotite to kaolinite during saprolite-soil weathering. Soil Sci. Soc. Am. J. 50:810–819.

Sanchez, P.A. 1976. Properties and management of tropical soils. John Wiley & Sons, New York.

Schwab, E.B., D.W. Reeves, C.H. Burmester, and R.L. Raper. 2002. Conservation tillage systems for cotton in the Tennessee Valley. Soil Sci. Soc. Am. J. 66:569–577.

Schwertmann, U., and R.M. Taylor. 1989. Iron oxides, p. 379–438. *In* J.B. Dixon and S.B. Weed (eds.) Minerals in soil environments. 2nd edn. SSSA, Madison, WI.

Scrivner, C.L., J.C. Baker, and B.J. Miller. 1966. Soils of Missouri: A guide to their identification and interpretation. C823, Extension Division. University of Missouri, Colombia, MO.

Shaw, J.N. 2001. Iron and aluminum oxide characterization for highly weathered Alabama Ultisols. Commun. Soil Sci. Plant Anal. 32:49–64.

Shaw, J.N., B.F. Hajek, and J.M. Beck. 2010. Highly weathered mineralogy of select soils from Southeastern U.S. Coastal Plain and Piedmont landscapes. Geoderma 154:447–456.

Shaw, J.N., D.W. Reeves, and C.C. Truman. 2003. Clay mineralogy and dispersibility of soil and sediment derived from Rhodic Paleudults. Soil Sci. 168:209–217.

Shaw, J.N., L.T. West, D.D. Bosch, C.C. Truman, and D.S. Leigh. 2004. Parent material influence on soil distribution and genesis in a Paleudult and Kandiudult (Southeastern USA) complex. Catena 57:157–174.

Shaw, J.N., L.T. West, and C.C. Truman. 1998. Hydraulic properties of soils with water restrictive horizons in the Georgia Coastal Plain. Soil Sci.162:875–885.

Shirmohammadi, A., W.G. Knisel, and J.M. Sheridan. 1984. An approximate method for partitioning daily streamflow data. J. Hydrol. 74:335–354.

Soil Survey Staff. 1960. Soil classification, a comprehensive system, 7th approximation. USDA. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1999. Soil taxonomy. A basic system of soil classification for making and interpreting soil surveys. 2nd edn. USDA-SCS. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS. US Government Printing Office, Washington, DC.

Southard, R.J., and S.W. Buol. 1988a. Subsoil blocky structure formation in some North Carolina Paleudults and Paleaquults. Soil Sci. Soc. Am. J. 52:1069–1076.

Southard, R.J., and S.W. Buol. 1988b. Subsoil saturated hydraulic conductivity in relation to soil properties in the North Carolina Coastal Plain. Soil Sci. Soc. Am. J. 52:1091–1094.

Truman, C.C., J.N. Shaw, and D.W. Reeves. 2005. Tillage effects on rainfall partitioning and sediment yield from an Ultisol in Central Alabama. J. Soil Water Conserv. 60:89–98.

van Wambeke, A. 1991. Soils of the tropics: Properties and appraisal. McGraw-Hill, New York.

West, L.T., F.H. Beinroth, M.E. Sumner, and B.T. Kang. 1997. Ultisols: Characteristics and impacts on society. Adv. Agron. 63:179–236.

West, L.T., W.P. Miller, R.R. Bruce, G.W. Langdale, J.M. Laflen, and A.W. Thomas. 1992. Cropping system and consolidation effects on rill erosion in the Georgia Piedmont. Soil Sci. Soc. Am. J. 56:1238–1243.

West, L.T., W.P. Miller, G.W. Langdale, R.R. Bruce, J.M. Laflen, and A.W. Thomas. 1991. Cropping system effects on interrill soil loss in the Georgia Piedmont. Soil Sci. Soc. Am. J. 55:460–466.

33.13 Oxisols

Friedrich H. Beinroth

Hari Eswaran

Goro Uehara

Christopher W. Smith

Paul F. Reich

33.13.1 Introduction

The deep, red and highly weathered soils of the tropics have long fascinated pedologists, particularly those from the temperate region. The uniqueness of these soils lies not only in their properties, but also in their geographic distribution being confined almost exclusively to the tropics. As this account is intentionally

concise, ample references are provided for those seeking additional or more detailed information.

33.13.2 Historical Background

In the early literature, the highly weathered soils of the tropics were identified as red soils, red loams, or red earths. In an historical report on travels in South India, Buchanan (1807) described soil material used for construction and called it laterite (Buchanan, 1807). The term was adopted by the pedologic community and is still used today with a wide variety of meanings (Alexander and Cady, 1962; Maignen, 1966). In 1949, a group of scientists proposed the term Latosols (Cline, 1975) that soon became popular and, by the midfifties, the concept of Latosols as highly weathered soils with a low negative charge was firmly established. Several soil classification systems developed during the 1950s and 1960s recognized such soils in a separate class with the nomenclature shown in Table 33.59.

The term Oxisol originated around 1954 (Smith, 1963, 1965) during the development of *Soil Taxonomy* (Soil Survey Staff, 1975). In 1978, an International Committee on the Classification of Oxisols (ICOMOX) was formed (Buol and Eswaran, 1988) to initiate discussions that would lead to an improved classification. Through circular letters and four international soil classification workshops held in Brazil (1976), Malaysia and Thailand (1978), Rwanda (1981), and Brazil (1986), the current state of knowledge of these soils was assembled and documented (Beinroth and Panichapong, 1978; Beinroth and Paramanathan, 1978; Camargo and Beinroth, 1978; Beinroth et al., 1983, 1986).

33.13.3 Concept and Classification of Oxisols

33.13.3.1 Perceptions

The rationale of the taxon of Oxisols is based largely on the concept of Latosols, which was developed to designate all zonal soils having their dominant characteristics associated with low silica:sesquioxide ratios, low base cation exchange capacity, low activity clays, and low content of weatherable minerals (Kellogg, 1949; Cline, 1975). These characteristics were selected because they were thought to manifest the effects of advanced pedogenesis under tropical conditions.

TABLE 33.59 Synonyms for Oxisols in Other Classification Systems

Synonym	Classification System	Sources
Sols ferrallitiques	French	Aubert (1958)
Latosols	Brazilian	Bennema et al. (1959)
Solos ferraliticos	Portuguese	Botelho da Costa (1954, 1959)
Oxisols	Ghana	Charter (1958)
Ferralsols		FAO–UNESCO (1971–1976)
Rotlehm	German	Harrassowitz (1930)
Laterites		Maignen (1966)
Red earths	British	Robinson (1951)
Kaolisols	Belgian	Tavernier and Sys (1965)

33.13.3.2 Definition

Simplified from the *Keys to Soil Taxonomy* (Soil Survey Staff, 2010), Oxisols may be defined as those soils that fail to meet the criteria definitive for Gelisols, Histosols, Spodosols, and Andisols but have either (1) an oxic horizon within 150 cm of the mineral soil surface and no kandic horizon within this depth, or (2) 40% or more clay in the surface horizon and a kandic horizon within 100 cm of the mineral soil surface that meets the weatherable mineral properties of the oxic horizon.

The definition of Oxisols has undergone significant changes since it was first published in *Soil Taxonomy* (Soil Survey Staff, 1975). Consequently, more soils now qualify for Oxisols; this should be kept in mind when consulting the earlier literature. It should also be mentioned that not all intensively weathered soils are Oxisols. Highly weathered soils with a marked clay increase below a coarse-textured surface horizon are by definition excluded from the Oxisols and are classified as kandic great groups of Ultisols and Alfisols (van Wambeke, 1992; Soil Survey Staff, 1998).

33.13.3.3 Diagnostic Criteria

Key to the identification of Oxisols is the presence of either an oxic horizon or a kandic horizon underlying a surface horizon with 40% or more clay. The key properties of the oxic horizon are its charge characteristics and negligible amounts of weatherable minerals. The charge characteristics are defined by the magnitude of the charge (CEC by NH_4OAc at pH 7.0 of <16 cmol$_c$ kg^{-1} clay) and the charge estimated by the Effective Cation Exchange Capacity (ECEC < 12 cmol$_c$ kg^{-1} clay). The low weatherable mineral requirement of <5% ensures that there are few weatherable minerals that could release plant nutrients. The particle size class of the oxic horizon is sandy loam or finer and its upper boundary is diffuse. To be diagnostic, the oxic horizon must have a minimum thickness of 30 cm and occur within 150 cm of the mineral soil surface.

The kandic horizon (Moormann, 1985) shares some of the properties of the oxic and argillic horizons. Like the argillic horizon, clay increases with depth but the charge characteristics are those of the oxic horizon. To be diagnostic for Oxisols, the kandic horizon must meet the weatherable mineral requirements of the oxic horizon, occur beneath a surface horizon that has 40% or more clay, and have its upper boundary within 100 cm of the soil surface.

Suborders of Oxisols are differentiated on the basis of soil moisture regimes, aquic conditions, the presence of a histic epipedon, and soil color. Defining criteria for great groups are the kandic and sombric horizons, plinthite, acric properties, and base saturation. The 14 differentiae used to establish subgroups are andic-like properties, Al + Fe percentage, aquic conditions, CEC, ΔpH ($pH_{KCl} - pH_{water}$), histic epipedon, kandic horizon, lithic contact, organic C content, petroferric contact, plinthite, soil color, soil depth, and the sombric horizon.

At the family level, criteria include particle size class, mineralogy class, and soil temperature regimes. Soil mineralogy is an important factor, and eight mineralogical classes are recognized (Table 33.60). When the soil has >18% Fe_2O_3 or gibbsite in the fine earth fraction (<2 mm), the mineralogical class is determined by

TABLE 33.60 Mineralogy Classes of Oxisols

Mineralogy Class	Fe$_2$O$_3$ (%)	Gibbsite (%)	Kaolinite (%)	Halloysite (%)
Ferritic	>40			
Gibbsitic	0–40	>40		
Sesquic	18–40	18–40		
Ferruginous	18–40	0–18		
Allitic	0–18	18–40		
Kaolinitic	0–18	0–18	>50	
Halloysitic	0–18	0–18	<50	>50
Mixed	0–18	0–18	<50	<50

Source: Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS. U.S. Government Printing Office, Washington, DC.

the amounts of each of these. If there is <18% Fe$_2$O$_3$ (12.6% Fe) or gibbsite, the composition of the clay fraction is considered.

33.13.3.4 Taxa

The class of Oxisols has 5 suborders, 22 great groups, 213 subgroups, an estimated 400 families, and perhaps as many as 1000 soil series. The great groups and possible subgroups of the Oxisols are presented in Table 33.61.

33.13.4 Geography of Oxisols

33.13.4.1 Global Extent and Geographic Distribution

The Oxisols of the world comprise about 9.8 million ha (Table 33.1), approximately equivalent to 7.5% of the global and 25% of the tropical land area. The distributions of Oxisol suborders in South America, Africa, and globally are illustrated in Figures 33.80 through 33.82. Over 95% of the Oxisols occur in the

TABLE 33.61 Listing of Suborders, Great Groups, and Subgroups in the Oxisols Order

Suborder	Great Group	Subgroups
Aquox	Acraquox	Plinthic, Aerie, Typic
	Plinthaquox	Aerie, Typic
	Eutraquox	Histic, Plinthic, Aerie, Humic, Typic
	Haplaquox	Histic, Plinthic, Aerie, Humic, Typic
Torrox	Acrotorrox	Petroferric, Lithic, Typic
	Eutrotorrox	Petroferric, Lithic, Typic
	Haplotorrox	Petroferric, Lithic, Typic
Ustox	Sombriustox	Petroferric, Lithic, Humic, Typic
	Acrustox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Anionic Aquic, Anionic, Plinthic, Aquic, Eutric, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic
	Eutrustox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Plinthaquic, Plinthic, Aquic, Kandiustalfic, Humic Inceptic, Inceptic, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic
	Kandiustox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Plinthaquic, Plinthic, Aquic, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic
	Haplustox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Plinthaquic, Plinthic, Aqueptic, Aquic, Oxyaquic, Inceptic, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic
Perox	Sombriperox	Petroferric, Lithic, Humic, Typic
	Acroperox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Anionic, Plinthic, Aquic, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic
	Eutroperox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Plinthaquic, Plinthic, Aquic, Kandiudalfic, Humic Inceptic, Inceptic, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic
	Kandiperox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Plinthaquic, Plinthic, Aquic, Andic, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic
	Haploperox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Plinthaquic, Plinthic, Aquic, Andic, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic
Udox	Sombriudox	Petroferric, Lithic, Humic, Typic
	Acrudox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Anion Aquic, Anionic, Plinthic, Aquic, Eutric, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic
	Eutrudox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Plinthaquic, Plinthic, Aquic, Kandiudalfic, Humic Inceptic, Inceptic, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic
	Kandiudox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Plinthaquic, Plinthic, Aquic, Andic, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic
	Hapludox	Aquic Petroferric, Petroferric, Aquic Lithic, Lithic, Plinthaquic, Plinthic, Aquic, Inceptic, Andic, Humic Rhodic, Humic Xanthic, Humic, Rhodic, Xanthic, Typic

Source: Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS. U.S. Government Printing Office, Washington, DC.

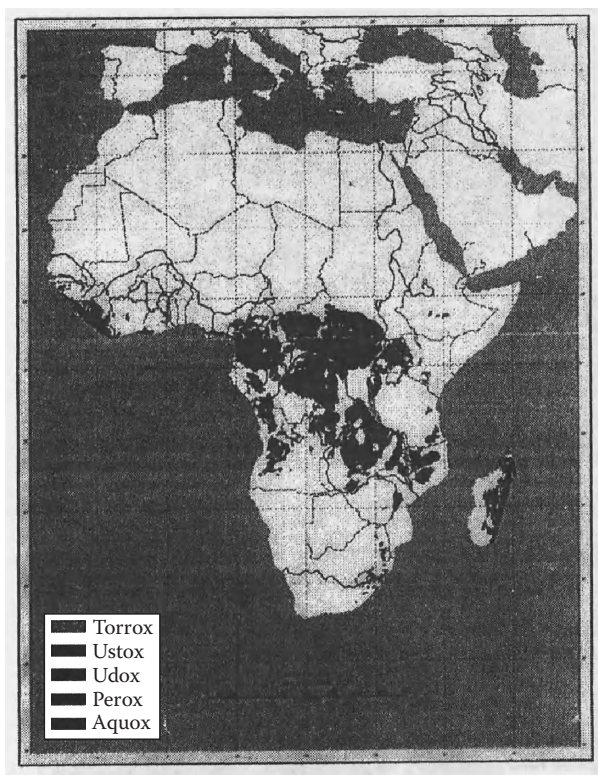**FIGURE 33.80**   Distribution of Oxisols in South America.

FIGURE 33.81   Distribution of Oxisols in Africa.

tropics (Table 33.2) where udox occupy 13.2% and ustox 7.9% of the land area. South America has the largest extent of Oxisols and is home to 57% of the world's Oxisols, most of them occurring in Brazil. In Africa, Zaire probably has the largest areas of Oxisols. In Southeast Asia, Oxisols only occur in small isolated areas with the largest presumed to be in Borneo (Kalimantan). In Oceania, there are small areas of Oxisols in Australia and on the islands of the Pacific and Caribbean Basins.

33.13.4.2 Extent and Geographic Distribution in the United States

Oxisols have not been officially identified in the conterminous United States and only occur on the islands of Hawaii, Guam, Micronesia, and Puerto Rico where they account for about 4.3% and 6.7% of the land area, in Hawaii and Puerto Rico, respectively (Table 33.62). Small isolated areas of Oxisols have been observed in the coastal hills of northern California formed from peridotite, but have not been established as a soil series.

33.13.5 Formation and Landscape Relationships of Oxisols

33.13.5.1 Factors of Oxisol Formation

Soil is the cumulative result of the interaction of pedogenetic processes, which are controlled by the factors of soil formation. Although the factorial approach to understanding soil formation has serious conceptual and operational limitations (Smeck et al., 1983; Wilding, 1994), it remains a unifying philosophy in pedology. The formation of Oxisols is, therefore, discussed here in the context of this paradigm.

33.13.5.1.1 Climate

With respect to rainfall, the definition of Oxisols implies two conditions. First, in the case of Oxisols with kandic horizons, there must be some time during the year when evapotranspiration exceeds precipitation as this appears to be a prerequisite for the formation of the kandic horizon (van Wambeke, 1992). Second, precipitation must exceed the capacity of the soil to retain water during some time of the year, so that water percolates through the solum (Miller, 1983). This causes the leaching of soluble weathering products and favors the residual concentration of kaolinite and sesquioxides (van Wambeke et al., 1983). The wide variety of tropical climates that meet these conditions corresponds to the udic, perudic, and ustic soil moisture regimes of *Soil Taxonomy* (Soil Survey Staff, 1998). The few Torrox that now have an aridic soil moisture regime are considered relics of a more humid climate in the past.

Regarding temperature, most Oxisol areas have mean annual air temperatures >15°C with minimal annual fluctuations and, therefore, have isothermic or isohyperthermic soil temperature regimes. A few areas in southern Brazil and South Africa may have a thermic temperature regime. The climate conducive to the formation of Oxisols is thus typically humid tropical.
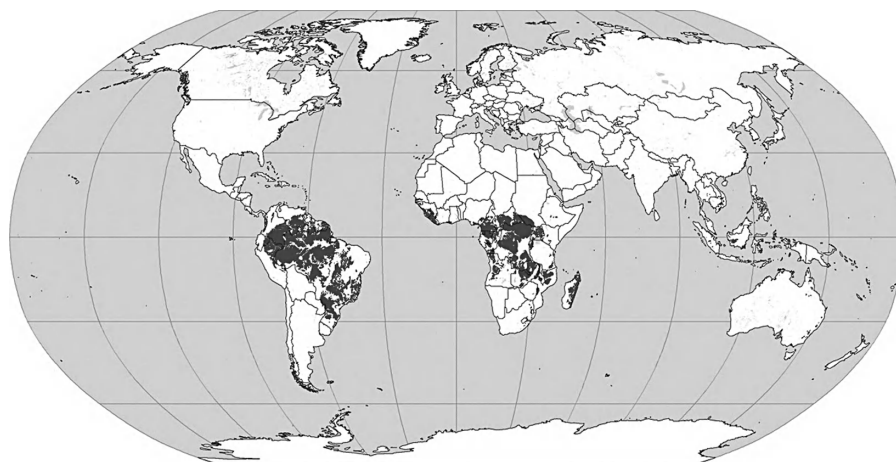
FIGURE 33.82 Global distribution of Oxisols. (Courtesy of USDA-NRCS, Soil Survey Division, World Soil Resources, Washington, DC, 2010.)

TABLE 33.62 Oxisol Series of Hawaii and Puerto Rico

Soil Series		Classification		Hawaii		Puerto Rico	
Hawaii	Puerto Rico	Subgroup	Code	Area (ha)	%	Area (ha)	%
	Delicias	Rhodic Haplustox	CCEN			280	0.03
	Moteado	Humic Haplaquox	DADD			0	0.00
Molokai		Typic Eutrotorrox	DBBC	14,143	0.84		
Molokai variant		Typic Eutrotorrox	DBBC	578	0.03		
	Matanzas	Lithic Eutrustox Kandiustalfic	DCCD			1,786	0.20
Wahiawa		Rhodic Eutrustox	DCCH	8,566	0.51		
Helemano		Rhodic Eutrustox	DCCN	11,250	0.67		
Lihue		Rhodic Eutrustox	DCCN	6,300	0.38		
Niu		Rhodic Eutrustox Inceptic	DCCN	1,368	0.08		
Mahana		Haplustox	DCEJ	4,918	0.29		
Makapili		Humic Haplustox Anionic	DCEM	963	0.06		
Halii		Acroperox	DDBE	1,968	0.12		
Kapaa		Anionic Acrudox	DEBF	9,198	0.55		
Pooku		Anionic Acrudox	DEBF	3,154	0.19		
	Nipe	Anionic Acrudox	DEBF			896	0.10
Kunuweia		Typic Acrudox	DEBO	325	0.02		
	Cotito	Lithic Eutrudox	DECD			280	0.03
Hanamaulu		Humic Kandiudox	DEDK	3,150	0.19		
Puhi		Humic Kandiudox	DEDK	5,205	0.31		
Lawai		Hapludox	DEDN	664	0.04		
	Daguey	Typic Kandiudox	DEDN			6,393	0.72
	Zarzal	Typic Kandiudox	DEDN			0	0.00
	Rosario	Lithic Hapludox	DEED			1,158	0.13
	Los Guineos	Inceptic Hapludox	DEEH			32,553	3.67
	Limones	Humic Hapludox	DEEL			993	0.11
	Catalina	Rhodic Hapludox	DEEM			108	0.01
	Bayamon	Typic Hapludox	DEEO			9,336	1.05
	Coto	Typic Hapludox	DEEO			5,262	0.59
Total Oxisols				71,750	4.28	59,045	6.66
Total land area				1,677,308	100.00	886,216	100.00

As most Oxisols have evolved over geologic time periods, consideration of the paleoclimate is essential for understanding and analyzing their development.

33.13.5.1.2 Parent Material

A distinction must be made between transported or allochthonous and residual or autochthonous parent materials. Autochthonous parent materials have weathered in situ, whereas allochthonous materials have been transported by fluviocolluvial processes. The sediments that form the parent material of Oxisols have frequently been preweathered and may have gone through more than one weathering and transport cycle. Where Oxisol development occurs in situ, the parent material is usually saprolite, which is a highly weathered parent rock that, unless it is collapsed, still preserves most of the original rock fabrics. In terms of area, the allochthonous Oxisols are more extensive than the autochthonous (Tavernier and Eswaran, 1972). The allochthonous Oxisols on the Brazilian and Guayanian Shields (Lepsch and Buol, 1974; Lepsch et al., 1977) and the Central African Plateau (Ruhe, 1956) are formed in sediments on mid to upper Tertiary surfaces and are generally as deep as the sediments. In Hawaii, geomorphological evidence, underlying rock weathering degree, and island age and time frame of cessation of volcanic ash production suggest that these soils have formed from basaltic volcanic ash.

33.13.5.1.3 Biota

There appear to be no clear cause and effect relationships between vegetation and the geography of Oxisols as they occur under both rainforest and savannas (van Wambeke et al., 1983). Termites and ants, however, may play an important role in the formation of Oxisols. Lee and Wood (1971) studied the termite activity in soils and its effect on modifications in the solum. van Wambeke (1992) cites a French study which reports that the amount of soil that termites displace varies between 300 and 1000 kg ha^{-1} year^{-1}, and that their mounds typically comprise 250 m^3. Leaf cutter ants not only transport soil particles upward but also carry large amounts of plant tissue below the surface. A comprehensive account of the role of termites and the mesofauna in tropical pedogenesis has been compiled by van Wambeke (1992).

33.13.5.1.4 Relief

Oxisols occur on many landforms, including uplands, backslopes, pediments, interfluves, and river and marine terraces. The one characteristic that the loci of autochthonous Oxisols have in common is geomorphic stability over usually long periods of time. Commonly, this implies low slope gradients. Small gradients, however, are not necessarily an indication of stability or geomorphic age as a level surface may be a recent floodplain or a Tertiary peneplain.

33.13.5.1.5 Time

The time required for the formation of Oxisols obviously depends on the other pedogenetic factors, notably the weatherability of

the parent rock, the composition and weatherability of transported parent materials, and the climate and its fluctuations over time. There is much scope for variability in these conditions, and consequently, in the age of Oxisols.

For Oxisols developed in transported sediments, their geologic age determines the actual maximum time available for soil formation. In Puerto Rico, for example, Oxisols have developed in preweathered materials on marine terraces of Quaternary age, but the autochthonous Oxisols of the interior commonly occur on Pliocene or Miocene surfaces that have been exposed to subaerial weathering for as long as 15 million years (Beinroth, 1981). Oxisol landscapes of comparable age have been reported in Africa, South America, and Australia (Miller, 1983). Oxisols in Hawaii can be less than 1 million years old.

33.13.5.2 Processes of Oxisol Formation

Various processes combine to produce, either concurrently or sequentially, the unique features of Oxisols. Prominent among these processes are those that (1) lead to the intensive weathering of primary minerals and the removal of soluble weathering products, the formation of 1:1 lattice clays, and the residual accumulation of sesquioxides, and (2) cause an increase of silicate clay in subsurface horizons. The first process, collectively known as laterization, is of paramount importance in the formation of most Oxisols. It involves desilication, ferrallization, ferritization, and allitization and causes the chemical migration of silica out of the solum and the relative concentration of sesquioxides in the soil. Formation of Fe coatings or aggregation at the expense of quartz grains is a common process in the tropical environment (Padmanabhon and Mermut, 1996).

Clay increase with depth that is diagnostic for the kandic horizon may result from (1) the process known as lessivage that causes clay illuviation and results in a clay maximum or bulge in the subsoil; (2) vertical downward translocation of clay without accumulation in an illuvial horizon. French pedologists refer to this process that leads to lighter textured surface horizons as appauvrissement or impoverishment (van Wambeke, 1992); (3) clay depletion in the soil surface may be caused by the selective removal of fine particles from the surface soil by erosion or mesofauna; (4) in Oxisols that are seasonally flooded, ferrolysis can cause the destruction of clay in the topsoil (Brinkman, 1970); (5) as postulated by Simonson (1949), in situ formation of clay in the B horizon may occur; and (6) lithological discontinuities may account for textural changes in the solum. In the past, the process of podzolization has been associated with Oxisols, but the evidence for it is intangible.

The formation of plinthite is often considered an extreme manifestation of laterization. The Soil Survey Staff (2010) characterizes plinthite as an Fe-rich, humus-poor mixture of sesquioxides, clay, quartz, and other diluents that commonly appear as dark red mottles in platy, polygonal, or reticulate patterns and generally forms in a horizon that is saturated with water for some time during the year. In a moist soil, plinthite is soft enough to

be cut with a spade, but it changes irreversibly to ironstone or petroplinthite when exposed to repeated wetting and drying. Plinthite (Eswaran and Raghumohan, 1973) is a diagnostic feature of many soils which are hydromorphic or have gone through a hydromorphic phase during their evolution. Hardening of plinthite takes place slowly when the ground water table is lowered and the soil surface ground cover is removed. If the surface soil horizons are eroded, the underlying plinthite is exposed and hardens rapidly to form petroplinthite (Sys, 1968; Eswaran and Raghumohan, 1973). A related feature is the development of a petroferric contact, which is an abrupt boundary between soil material and an underlying layer of cemented petroplinthite gravel that is hard and impermeable to both roots and water. The definition, genesis, and kinds of plinthite and petroplinthite have been discussed in detail by van Wambeke (1992). The upland soils of Palau and Yap, where eroded and degraded, exhibit surface layers dominated by irregular jagged and vesicular ironstone gravel and gibsite pendants. The uneroded, nondegraded adjacent soils however do not exhibit plinthite (Smith, 1983).

Gleization, which is a process of importance in the formation of some Oxisols, refers to the reduction of Fe and Mn under seasonally anaerobic soil conditions and produces bluish to greenish gray matrix colors with or without yellowish brown, brown, or black mottles and ferric and manganiferous concretions. These redoximorphic features are striking characteristics of the aquic subgroups.

The transformation of raw organic material into soil organic matter (SOM) known as humification occurs in all Oxisols, but is of particular importance in the humic and histic subgroups. A related process is the illuviation of humus that results in the formation of the sombric horizon which is a dark colored subsurface horizon found in soils on old geomorphic surfaces of Central Africa and parts of South America (Eswaran and Tavernier, 1980). Although its origin and genesis are still being debated, it is used as a diagnostic horizon in *Soil Taxonomy* because it is a distinctive feature in an otherwise nondescript soil (Eswaran et al., 1986). Oxisols with sombric horizons are restricted to the cool high plateaus of the tropics at altitudes between 1400 and 3000 m above sea level that have isothermic or colder temperature regimes and a udic soil moisture regime (van Wambeke, 1992).

In summary, a broad range of environmental determinants, and pedogenetic processes and mechanisms may be involved in the formation of Oxisols. Yet, there is no single set of formative factors, processes, and mechanisms that could account for the formation of all Oxisols. The fact that the causative conditions may not have been the same or may have operated at different intensities over time, and may have occurred simultaneously or sequentially, adds complexity to Oxisol genesis.

33.13.5.3 Landscape Relationships

The geomorphic evolution of the landscape is an important factor that is more important in the formation of Oxisols than in many other kinds of soil. As pointed out by Daniels et al. (1971), the occurrence of Oxisols, as that of other soils, is controlled by the interaction of geomorphic and other formative factors and the resulting rates and degrees of expression of pedogenic processes. Beinroth et al. (1974), Lepsch and Buol (1974), and Lepsch et al. (1977) provide illustrative examples of landscape relationships of Oxisols in Hawaii and Brazil that invariably show that Oxisols occupy geomorphic positions that are older and more stable than the surfaces where Ultisols and Inceptisols occur with which they are geographically associated. However, the recent introduction of the kandic horizon, which may be diagnostic for both Ultisols and Oxisols, has blurred the geomorphic boundary between the two orders. In the case of some Hawaiian Ultisols, a truncated oxidic paleosol is overlain by oxidic depositional material of ash origin. Although no significant clay percent differences are present between the two materials, the presence of clay films on the paleosol results in the Ultisol classification but physical and chemical properties and behavior are those of an Oxisol.

33.13.6 Properties

33.13.6.1 Macromorphology

Compared with the often strikingly horizonated soils of other orders, the field morphology of most Oxisols is visually rather uniform. They nevertheless have some distinguishing attributes. Color is a prominent feature of Oxisols. The surface horizon of most lowland Oxisols is a thin, reddish to yellowish colored ochric epipedon. Oxisols at high elevations (>1000 m) frequently have a dark-colored, humus-rich surface horizon (Ruhe, 1956), which may qualify for a mollic or umbric epipedon. Many Oxisols of Central Africa also have a dark-colored layer, the sombric horizon, in the subsoil. As Table 33.63 indicates, subsurface colors range from light gray in the Aquox to various hues of red in the upland Oxisols and are generally a function of the Fe content of the original material or rock (Eswaran and Sys, 1970). Color is also related to the kind of Fe minerals, with goethite producing yellow colors and hematite red colors. Presence of colloidal organic matter darkens the soil. Although organic carbon may be 4% or more, iron oxides dominate the color leading to the typical designation of an ochric epipedon. If there is a fluctuating groundwater table, mottles or plinthite may form in the oscillation zone. If the soil remains saturated with water for long periods, reduction and removal of the Fe results in a whitish horizon.

The texture of Oxisols may vary from sandy loam to clay. A characteristic feature is that the macro structural elements are weak in the subsoil. When the soil is gently pressed between the thumb and forefinger, the material collapses or fails abruptly. This is probably a good field indicator for an oxic horizon. In surface layers with higher clay contents, macrostructure can be moderate to strong, fine to medium subangular blocky. Excessive tillage however can reduce this condition to a seemingly single grain appearance of silt and sand sized microaggregates.

Many Oxisols have stone lines with the stones being quartz or petroplinthite gravel (Ruhe, 1956). The stone line is a mark of a lithologic discontinuity indicating that the material above

TABLE 33.63 Physical Properties of Selected Pedons Representing Oxisol Suborders

Classification	Depth (cm)	Horizon	Bulk Density (g cm⁻³)	Water Retention 1/3 bar (%)	15 bar (%)	WRD	Particle Size (%) Sand	Silt	Clay	Soil Color
Typic Acraquox	0–10	A1	1.3	32.5	26.9	0.1	33.1	10.5	56.4	10YR 6/1
(Brazil)	10–30	Ag	1.4	21.8	17.2	0.1	34.5	11.6	53.9	10YR 7.1
	30–48	Bog1	1.3	28.7	21.1	0.1	25.1	8.8	66.1	10YR 7.1
	48–77	Bog2	1.3	29.4	23.1	0.1	44.5	13.5	42.0	10YR 8.2
	77–90	Bov	1.4	27.3	21.9	0.1	62.3	11.8	25.9	10YR 5/8
Typic Eutrotorrox	0–23	Ap			21.5		17.1	40.2	42.7	2.5YR 2/4
(Hawaii)	23–50	Bo1	1.3	29.6	21.9	0.1	10.4	4.17	47.9	2.5YR 3/4
	50–87	Bo2	1.4	28.2	22.1	0.1	24.1	30.1	45.8	2.5YR 3/4
	87–123	Bo3	1.4	28.9	21.6	0.1	18.8	34.7	46.5	2.5YR 3/4
	123–150	Bo4	1.3	30.9	20.4	0.1	11.2	39.9	48.9	5YR 3/3
Humic Rhodic	0–25	Ap			23.8		18.0	41.2	40.8	2.5YR 3/2
Eutrustox (Brazil)	25–40	AB	1.2	32.4	23.3	0.1	16.8	37.9	45.23	2.5YR 3/2
	40–64	Bo1	1.2	30.8	24.0	0.1	10.9	25.1	64.0	2.5YR 3/2
	64–110	Bo2	1.1	32.2	24.6	0.1	13.3	28.4	58.3	2.5YR 3/2
	110–210	Bo3	1.1	31.4	24.6	0.1	18.5	38.1	43.4	2.5YR 3/2
Typic Kandiperox	0–10	Ap1	0.9	42.0	26.5	0.1	13.2	28.9	57.9	5YR 3/3
(Indonesia)	10–21	Ap2	0.9	39.5	26.6	0.1	11.4	28.7	59.9	5YR 4/3
	21–51	Bo1	1.0	48.0	31.8	0.2	7.0	20.7	72.3	5YR 3/4
	51–81	Bo2	0.9	50.9	32.6	0.2	6.1	18.5	75.4	5YR 3.4
Anionic Acrudox	0–28	A1	1.1	35.4	26.5	0.3	9.2	36.3	53.8	2.5YR 2/4
(Puerto Rico)	28–46	B1	1.2	26.7	22.8	0.2	7.4	34.9	54.5	2.5YR 2/4
	46–71	Bo1	1.1	34.4	24.8	0.4	9.8	30.6	57.7	7.5YR 3/8
	71–97	Bo2	1.3	35.7	25.9	0.4	23.3	21.0	59.6	7.5YR 3/8
	97–120	Bo3	1.4	31.6	26.4	0.2	17.0	23.3	55.7	7.5YR 3/4
	120–155	Bo4	1.3	29.8	24.5	0.1	19.2	27.2	59.7	7.5YR 3/4
Humic Sombriudox	0–15	A	1.0	26.5	11.3	0.2	58.2	8.6	33.2	7.5YR 3/3
(Rwanda)	15–40	B1	1.4	15.4	11.1	0.1	56.3	9.3	34.4	7.5YR 2/4
	40–66	Bo1	1.3	17.9	12.4	0.1	52.6	8.9	38.5	7.5YR 2/4
	66–91	Bh1	1.3	21.6	15.2	0.1	42.8	8.4	48.8	7.5YR 2/4
	91–121	Bh2	1.3	22.7	16.7	0.1	43.2	7.1	49.7	7.5YR 2/4
	121–150	Bo2	1.4	19.9	15.4	0.1	42.7	7.3	50.0	7.5YR 2/4

Sources: Soil Survey Staff. 2011. National Cooperative Soil Characterization Database. Available online at http://ssldata.nrcs.usda.gov. Accessed July 8, 2011.

the line was deposited or formed at a different period than the material below. Stone lines frequently suggest that the soils are formed on transported deposits and point to the allochthonous nature of the material (van Wambeke, 1992). Some Oxisols have multiple stone lines.

33.13.6.2 Mineralogy and Micromorphology

The unique physical and chemical properties of Oxisols result mainly from the mineralogical composition of the colloidal fraction (Uehara and Keng, 1975; Herbillon, 1980). To have a CEC < 16 cmol$_c$ kg⁻¹ clay by NH$_4$OAc at pH 7.0, the clay fraction must be dominated by low activity clays such as kaolinite. Iron oxyhydroxide minerals, such as goethite, hematite, and ferrihydrite, are usually associated with kaolinite, but in some Oxisols, the Fe minerals predominate (Jones et al., 1982), particularly in soils belonging to ferruginous or ferritic families. The Fe minerals

can have a high positive charge (Jones and Uehara, 1973), which accounts for the special physical and chemical properties discussed later.

Gibbsite is present as a secondary mineral in many Oxisols (Eswaran et al., 1977). Weathering of primary minerals releases Si and Al; Si is lost in the soil solution while the Al crystallizes as gibbsite, mostly as nodules. The gibbsite crystals in the nodules are well crystallized as shown in the SEM micrograph taken at a magnification of ×2500 (Figure 33.83). The crystals are euhedral and twinning is common. Typically, gibbsite crystals have the size of fine silt, although dense pendants of almost pure gibbsite with the morphology of the root channels they formed in can be observed in highly degraded Oxisols of Palau (the Babelthuap Series). Oxisols usually have more gibbsite in the silt than in the clay fraction. It is for this reason that gibbsitic families are defined on the basis of the amount of gibbsite in the fine earth (<2 mm) fraction.
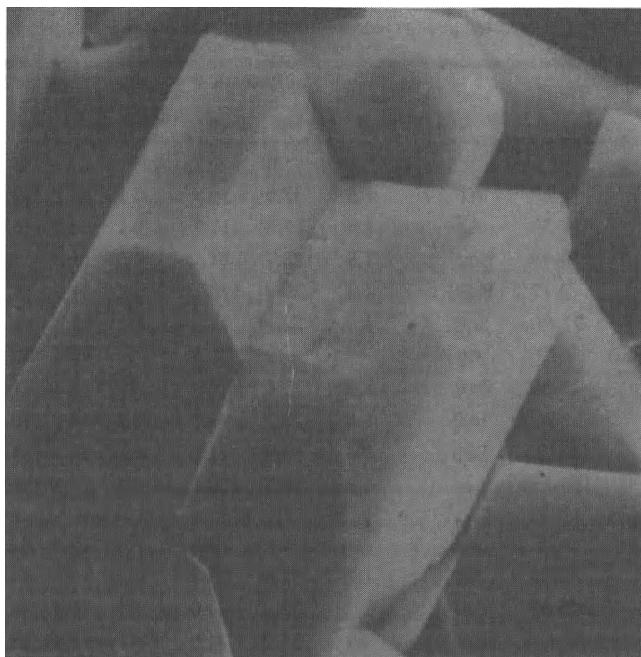
FIGURE 33.83 Scanning electron micrograph of a gibbsite nodule in a Gibbsiudox from Malaysia; the gibbsite crystals are typically euhedral and fine and medium silt size; magnification × 10,000.

The other frequent mineral in Oxisols is goethite (Eswaran et al., 1978). Plinthite, laterite, or petroplinthite frequently exhibit characteristic forms of goethite aggregates in thin sections as illustrated by the SEM micrograph (×10,000) in Figure 33.84. The goethite crystals have a typical lenticular shape and
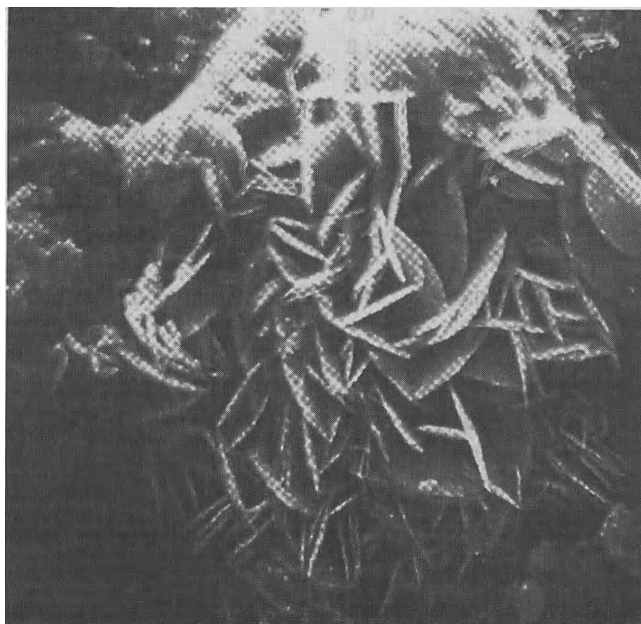

FIGURE 33.84 Scanning electron micrograph of a geothite in a laterite fragment. The crystals have a lenticular habit with split edges; laterite sample from Angadipuram, Kerala, India, which is the type locality of the Buchanan lateriate; magnification × 25,000.

appear welded together, which gives the petroplinthic material its strength. Goethite and hematite have different habits and show different crystal forms.

In most Oxisols, the fabric is homogenous without too many specific entities like those illustrated previously (Buol and Eswaran, 1978). In some oxic horizons, a thin lining of ferriargillans (yellow coatings on the void walls) may be present. Ultrathin sections under TEM suggest that the combination of random orientation of clay particles, organic matter, and aggregates of Fe-bearing minerals accounts for the isotropic nature of the aggregates under a petrographic microscope (Santos et al., 1989). The presence of the clay skins is evidence of the transitional nature of the soil, and that clay illuviation and accumulation was an important process. The transitional nature is indicated by the kandi prefix in the soil name.

Unless eroded, the surface horizons of Oxisols have a relatively high organic matter content with high biological activity including the presence of fungal hyphae and fruiting bodies, which are generally indicators of good soil quality. Typical for the drier tropics is the presence of large termite nests, which can reach 5 m in height. Some species of termites are subsoil dwellers and their galleries may extend several meters into the soil. Bioturbation of the soil is, therefore, an important soil-forming process in tropical soils.

Oxic horizons have a friable consistency. When a large soil clod is gradually crushed in the hands, the material disaggregates and small rounded bodies become evident. These features, which are only observed in Oxisols, have been referred to as pedovites or soil eggs by Belgian pedologists and "earthy lumps" in Hawaiian pedon descriptions. The excellent and stable structure of Oxisols, and their high macroporosity resulting in moderate to moderately rapid infiltration rates, make them resistant to erosion. Where intense tillage has reduced surface structure to microaggregates and has created tillage pans, erosion rates can be dramatically increased.

33.13.6.3 Chemistry and Physics

The chemistry and physics of Oxisols are inextricably linked to the surface charge characteristics of minerals in the clay fraction. Unlike in most other soils, the surface charge of minerals in Oxisols is pH dependant and varies in magnitude and sign. Oxisols, however, are not the only soils dominated by variable charge systems. Andisols, Ultisols, and Histosols can be even better examples of soils with variable surface charge, so that the attribute that makes Oxisols unique is the low permanent negative charge of the silicate clay minerals in the clay fraction. Important physical and chemical properties of selected pedons representing Oxisol suborders are summarized in Tables 33.63 and 33.64.

The sign of the charge is readily established by the sign of the difference in pH (ΔpH) in the following equation:

$$\Delta pH = pH_{KCL} - pH_{water}$$

Oxisols with a positive ΔpH are rare, but not difficult to find if one knows where to look for them. They are rare because

TABLE 33.64 Chemical Properties of Selected Pedons Representing Oxisol Suborders

Classification	Horizon	pH pH_o	pH H_2O	pH KCl	pH ΔpH	OC (%)	Free Fe (%)	Charge (cmol$_c$ kg^{-1}) ECEC	Charge CEC 7EC 8.2	Charge CEC 8	Base Saturation (%) CEC 7	Base Saturation (%) CEC 8.2	Al Sat. (%) ECEC
Typic Acraquox	Al	4.3	4.8	4.5	−0.3	2.4	0.3	1.9	6.8	13.6	3.0	1.0	89.0
(Brazil)	Ag	4.3	4.9	4.6	−0.3	1.6	0.2	1.2	4.8	10.1	4.0	2.0	83.0
	Bogl	5.3	5.5	5.4	−0.1	0.9	0.4	0.1	2.0	6.3	5.0	2.0	0.0
	Bog2	6.8	6.0	6.4	0.4	0.6	1.3	0.2	1.4	5.3	14.0	4.0	0.0
	Bov	6.8	6.0	6.4	0.4	0.4	1.4	0.1	1.0	4.9	2.0	1.0	0.0
Typic Eutrotorrox	Ap	5.9	6.7	6.3	−0.4	1.0	12.2	8.9	9.4	16.1	95.0	55.0	0.0
(Hawaii)	Bo1	6.1	6.5	6.3	−0.2	0.7	11.9	8.1	8.2	14.8	99.0	55.0	0.0
	Bo2	6.3	7.5	6.9	−0.6	0.2	9.9	8.9	9.4	14.9	95.0	60.0	0.0
	Bo3	6.2	6.8	6.5	−0.3	0.2	8.6	9.3	10.2	14.9	91.0	62.0	0.0
	Bo4	6.0	7.0	6.5	−0.5	0.3	8.5	7.9	8.3	13.2	95.0	60.0	0.0
Humic Rhodic	Ap	5.4	6.6	6.0	−0.6	2.8	14.3	15.8	17.6	26.7	90.0	59.0	0.0
Eutrustox (Brazil)Ap	AB	5.9	6.5	6.2	−0.3	2.2	14.7	13.7	15.4	23.6	89.0	58.0	0.0
	Bo1	5.4	6.8	6.1	−0.7	1.2	14.3	8.0	9.0	15.9	89.0	50.0	0.0
	Bo2	5.7	6.9	6.3	−0.6	0.9	14.9	6.1	6.6	13.5	92.0	45.0	0.0
	Bo3	5.9	7.1	6.5	−0.6	0.5	14.8	4.1	4.3	11.1	95.0	35.0	0.0
Typic Kandiperox	Ap1	4.1	4.9	4.5	−0.4	2.0	5.6	5.9	16.9	25.5	30.0	20.0	15.0
(Indonesia)	Ap2	3.8	4.8	4.3	−0.5	1.5	5.7	4.4	15.4	22.9	15.0	10.0	48.0
	Bo1	4.4	5.2	4.8	−0.4	1.9	5.6	5.4	15.1	23.4	34.0	22.0	6.0
	Bo2	4.9	5.3	5.1	−0.2	0.5	5.8	6.0	14.3	21.3	42.0	28.0	0.0
Anionic Acrudox	Al	3.5	5.1	4.3	−0.8	6.0	13.0	7.9	25.4	34.8	11.0	8.0	17.7
(Puerto Rico)	Bl	3.8	5.0	4.4	−0.6	2.0	12.9	1.7	12.1	21.5	1.0	0.0	52.9
	Bo1	4.4	5.0	4.7	−0.3	1.3	16.5	0.0	8.2	15.7	0.0	0.0	0.0
	Bo2	6.2	5.2	5.7	0.5	0.9	19.2	0.0	6.4	12.8	0.0	0.0	0.0
	Bo3	6.7	5.5	6.1	0.6	0.7	23.1	0.2	5.3	12.1	2.0	0.0	0.0
	Bo4	7.1	5.7	6.4	0.7	0.6	25.7	0.0	3.8	12.8	0.0	0.0	0.0
Humic Sombriudox	A	3.3	4.5	3.9	−0.6	1.7	1.2	3.8	10.0	13.7	6.0	4.0	84.0
(Rwanda)	Bl	3.5	4.5	4.0	−0.5	1.3	1.4	3.0	7.7	12.3	4.0	2.0	90.0
	Bo1	3.4	4.6	4.0	−0.6	1.2	1.8	3.8	9.6	14.6	5.0	3.0	87.0
	Bh1	3.4	4.6	4.0	−0.6	1.4	2.2	5.3	12.3	19.9	7.0	5.0	83.0
	Bh2	3.2	4.6	3.9	−0.7	1.6	2.2	5.9	15.5	25.4	5.0	3.0	88.0
	Bo2	3.2	4.6	3.9	−0.7	1.1	1.8	4.7	11.6	17.3	5.0	3.0	87.0

Sources: Soil Survey Staff. 2011. National Cooperative Soil Characterization Database. Available online at http://ssldata.nrcs.usda.gov. Accessed July 8, 2011.
ΔpH, pH in KCl minus pH in H$_2$O; pH$_o$, pH of soil at the zero point of charge where positive and negative charges are equal; OC, organic carbon content.

positive ΔpH values almost always occur in the subsoil. They rarely or almost never occur in surface horizons because negatively charged SOM masks the positive charge in the mineral–organic mixture. Organic matter, like the mineral fraction in the oxic horizon, has variable charge characteristics. The difference lies in their respective points of zero charge (p.z.c.) being <pH 3 for organic matter and >7 for oxic materials. Since soil pH values rarely fall below 3, SOM is net negatively charged in most soils.

The p.z.c. for material in the oxic horizon is highly variable ranging from pH 3–6. As a rule, it increases as organic C decreases and the silica/sesquioxide ratio of the clay fraction decreases. Hematite (Fe$_2$O$_3$), for example, has a p.z.c. of 8.5 (Parks and de Bruyn, 1962). In this sense, Oxisols are products of desilication, and the end products of desilication are the oxides and hydrous oxides of Fe and Al.

Over 82 years ago, Mattson (1928, 1932) showed that p.z.c. increased as the silica/sesquioxide ratio decreased. He also showed that the pH of the gel shifted toward the p.z.c. on leaching with distilled water, which he called isoelectric weathering. This concept is useful in explaining the chemical and physical behavior of Oxisols and related soils dominated by low activity clays. This concept is best illustrated with an example, and one of the best examples is the Nipe soil of Puerto Rico and Cuba (Anionic Acrudox in Table 33.61). (Data is for the original Type location in Puerto Rico. This site has been urbanized. The new type location (data not presented) classifies as a Typic Acrudox. The isoelectric properties of the Nipe soil are mainly determined by the p.z.c., which is measured as pH$_o$ (pH at which positive and negative charges on variable charge surfaces are equal).

When ΔpH = 0, pH values measured in M KCl and water are identical indicating that when the material is net negatively
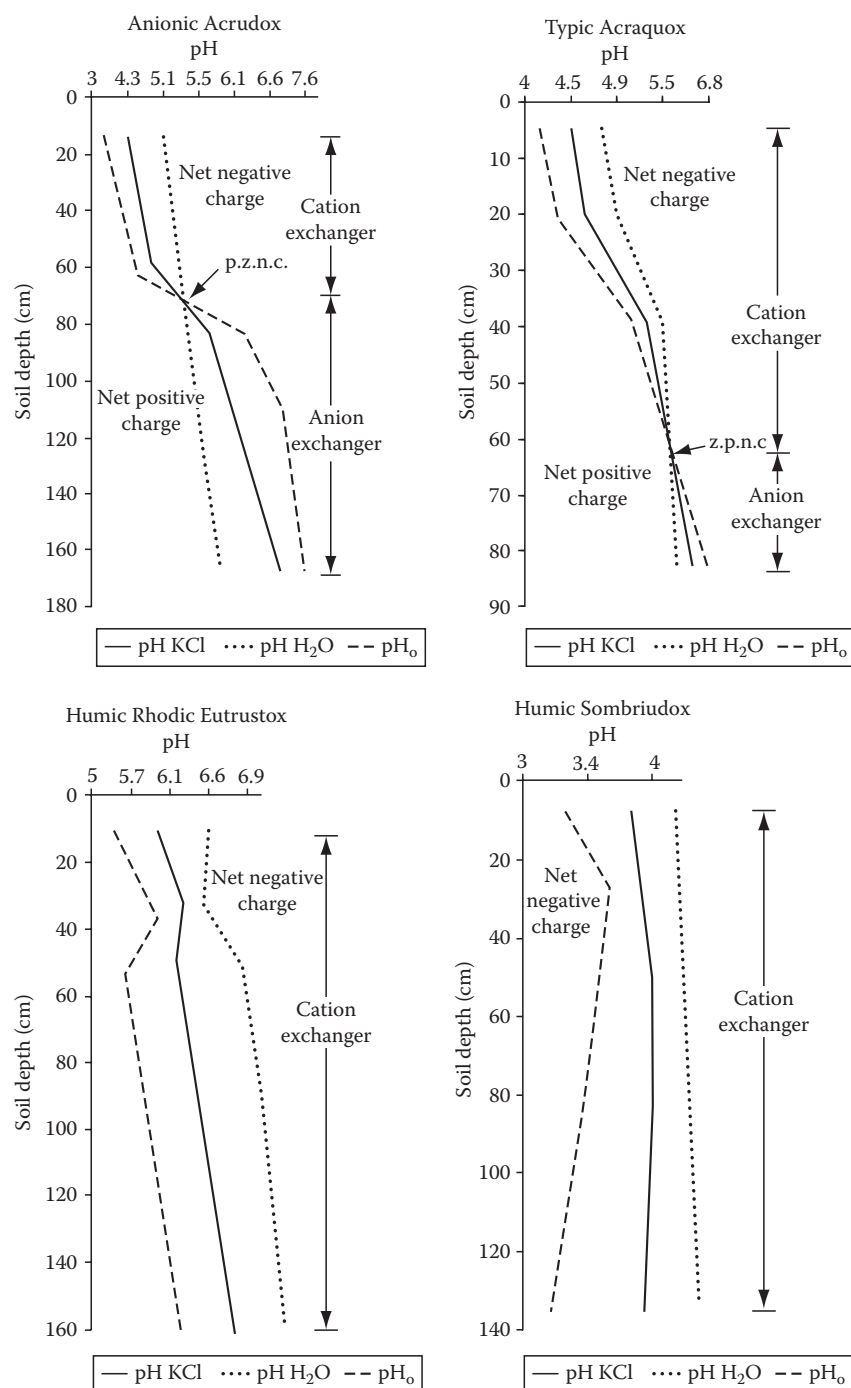
FIGURE 33.85 Surface charge characteristics of Oxisols of different suborders.

charged, pH_o is always lower than pH_{KCl} and vice versa. This does not apply to materials with significant amounts of permanent charge minerals.

The Nipe soil, like all Acric Oxisols, is a cation exchanger in the surface horizon and an anion exchanger in the subsoil. Charge characteristics of the Nipe soil and three other Oxisols are illustrated in Figure 33.85. The anion exchange capacity (AEC) of the subsoil can lead to unexpected consequences. In Hawaii, for example, 3–11 Mg NO_3–N ha^{-1} have been measured in the subsoil

and deep saprolite underlying Oxisols and Ultisols (Deenik, 1997), which explains the low NO_3^- levels in the groundwater underlying these soils even after nearly a century of intensive farming. Although the NO_3^- remains trapped above the water table, pesticides banned decades ago continue to enter the groundwater. Had the soil minerals been of the permanent charge type, the NO_3–N would have reached the groundwater many years ago.

In cation impoverished Oxisols such as the Nipe soil, lime is often added as a Ca fertilizer rather than an amendment.

But because lime raises pH and increases negative charge, Ca^{2+} ions remain in the limed layer and do not move to the impoverished subsoil. To circumvent this problem, gypsum and magnesium sulfate are favored over calcite or dolomite whenever the aim is to raise subsoil Ca and Mg. The factors determining movement of surface-applied amendments in variable charge soils are discussed in detail by Sumner (1995).

The acidic organic matter near the surface and the basic oxides in the subsoil determines, in essence, the chemical properties of the Nipe soil. In most Oxisols, the desilication process has not progressed as far as in the Nipe series. In such instances, the subsoil pH will be lower than that of the Nipe soil because the p.z.c. will be lowered by the higher Si content.

An opposing process that counteracts desilication is humification of the desilicated weathering products. Humic acids, like silicic acid, have a p.z.c. below 3. At pH levels normally encountered in soils, humus and silica are net negatively charged, so that they have strong affinities for positively charged sesquioxides. In one sense, organic matter adsorption on oxide surface has nearly the same effect as resilication of the oxides (Uehara, 1995). The surface of quartz is chemically similar to silicic acid, but its low specific surface renders it virtually inert. Desilicated Oxisols can be rejuvenated by additions of soluble silicates. Large crop responses to additions of calcium silicate above those obtained from similar lime applications have been reported (Plucknett, 1971).

The moisture characteristics curves of well-aggregated Oxisols show two major desorption zones. Water in the interaggregate pores drains rapidly between 0 and -0.01 MPa. Another desorption zone occurs when the intraaggregate pores begin to drain at about -15 MPa (Sharma and Uehara, 1968a, 1968b), which results in a bimodal pore size distribution (Tsuji et al., 1975; Bui et al., 1989). Some have referred to Oxisols as behaving like aggregated sands because of their stable aggregates and high macroporosity. This description tells only half the story. The aggregates that remain nearly water saturated beyond the wilting point impart additional properties to Oxisols. The water-saturated aggregates increase volumetric heat capacity and lower thermal diffusivity. Crops such as pineapple respond to practices that raise subsoil temperature. In Hawaii, plastic sheets used to increase the effectiveness of soil fumigants also aid the crop by raising subsoil temperature (Ekern, 1967). The intraaggregate water retained at high negative pressures also affects tillage operations. This intraaggregate water is freed under the shearing action of tillage implements and causes the soil to adhere to the implement. Some farmers solve this problem by bolting a teflon sheet onto the implement's shearing surface.

In summary, the physics and chemistry of Oxisols are strongly influenced by the extent to which desilication has occurred. Basic and ultrabasic parent materials readily desilicate and produce soils that approach the central concept of Oxisols. Desilication is also aided by warm and humid conditions, so that when Oxisols occur outside the tropics, they are almost always associated with ultrabasic rocks.

33.13.7 Conclusions

Oxisols occupy about 25% of the land area of the tropics where they are the single most extensive soil type. Yet, historically they have been perceived as agriculturally unproductive and problematic for management; under low input agriculture, yields are in fact low, risk is high, and the potential for land degradation is also high (Sanchez and Salinas, 1981). The negative notion of the poor agronomic performance derives substance from the inherent chemical constraints of Oxisols, which include, to varying degrees, a low nutrient retention capacity, anion adsorption, Ca deficiency, and Mn and Al toxicity (van Wambeke, 1974; Sanchez, 1976). It is also true that the inputs required to correct these constraints may be economically prohibitive for many farmers, and in places where better endowed soils are available, the Oxisols are, therefore, at a distinct comparative disadvantage (van Wambeke, 1992). Nevertheless, with science-based management that employs the tools and techniques of modern agriculture, Oxisols can be managed to be both economically and sustainably productive. Consequently, no less an authority than Charles E. Kellogg stated that "some day the most productive agriculture of the world will be mostly in the tropics, especially in the humid parts" (Kellogg, 1967). Presumably, this assessment is based on the realization that the favorable physical attributes of Oxisols outweigh their chemical limitations. While the latter can be amended with purchased inputs, good soil structure cannot be bought. That said, excessive tillage can destroy macro aggregation.

Oxisols are the dominant soils in the humid tropical forest ecosystem, a pristine environment that constitutes an enormous reservoir of sequestered carbon and unique ecological niches of great biological diversity, in addition to being a resource for food, timber, medicine, and other products for people. Yet, the resource-poor farmers invading these areas practice shifting cultivation, and slash and burn agriculture has become the most extensive form of agriculture in the tropics. As a result, over 15 million ha of forests are being burned annually, and some plants and animals are lost permanently when their habitats are destroyed. Moreover, the resilience of these ecosystems is so low that complete regeneration may not be achieved.

Oxisols constitute a major land resource and one of the few remaining frontiers for agricultural development, particularly in Africa and South America, and also support the largest areas of tropical forests. Their conversion to agricultural land will be at the expense of the forest with the concomitant loss of biodiversity and negative impact on global climate. It is imperative, therefore, to develop viable alternatives to traditional agricultural systems. If this challenge can be met successfully, the rewards are not only to provide a means for millions of people to extricate themselves from poverty, but also to ensure the survival of the tropical forests. Landuse policies guided by an understanding of the nature, properties, and ecological functions of Oxisols are critical to sustain the integrity and productivity of these land resources.

References

Alexander, L.T., and J.G. Cady. 1962. Genesis and hardening of laterite in soils. USDA Technical Bulletin No. 1282.

Aubert, G. 1958. Classification des sols. Compte Rendu, Reunion Sous-comitŽ, Brazzaville, Congo.

Beinroth, F.H. 1981. Some highly weathered soils of Puerto Rico. 1: Morphology, formation and classification. Geoderma 27:1–73.

Beinroth, F.H., M.N. Camargo, and H. Eswaran (eds.). 1988. Proc. Eighth Int. Soil Classif. Workshop: Charact. classif. util. Oxisols. EMBRAPA-SNLCS. Rio de Janeiro, Brazil, 12–13 May, 1986.

Beinroth, F.H., H. Neel, and H. Eswaran (eds.). 1983. Proc. Fourth Int. Soil Classif. Workshop, Rwanda. Part 1: Papers. Part 2: Field trip background and soil data. Agriculture Editions 4. ABOS-AGDC. Brussels, Belgium, 2–12 June, 1981.

Beinroth, F.H., and S. Panichapong (eds.). 1979. Proc. Second Int. Soil Classif. Workshop. Part II, Thailand. Department of Land Development, Bangkok, Thailand, 3–9 September, 1978.

Beinroth, F.H., and S. Paramanathan (eds.). 1979. Proc. Second Int. Soil Classif. Workshop. Part I, Malaysia. Department of Land Development, Bangkok, Thailand, 28 August–1 September, 1978.

Beinroth, F.H., G. Uehara, and H. Ikawa. 1979. Geomorphic relationships of Oxisols and Ultisols on Kauai, Hawaii. Soil Sci. Soc. Am. J. -journal. 38:128–131.

Bennema, J., R.C. Lemos, and L. Vetturs. 1959. Latosols in Brazil, p. 273–281. III Inter-African Soils Conf., Dalaba. Vol. I.

Botelho da Costa, J.V. 1954. Sure quelques questions de nomenclature des sols des regions tropicales, p. 1099–1103. Compte Rendu Conf. Int. Sols Africains. Leopoldville, Congo. Vol. 2.

Botelho da Costa, J.V. 1959. Ferralitic, tropical fersiallitic and tropical semi-arid soils: Definitions adopted in the classification of the soils of Angola, p. 317–319. III Inter-African Soils Conf., Dalaba. Vol. I.

Brinkman, R. 1970. Ferrolysis, a hydromorphic soil forming process. Geoderma 3:199–206.

Buchanan, F. 1807. A journey from Madras through the countries of Mysore, Kanara and Malabar, p. 436–461. Vol. 2. East India Co., London, U.K.

Bui, E.N., A.R. Mermut, and M.C.D. Santos. 1989. Microscopic and ultramicroscopic porosity of an Oxisol as determined by image analysis and water retention. Soil Sci. Soc. Am. J. 53:661–665.

Buol, S.W., and H. Eswaran. 1978. Micromorphology of Oxisols, p. 325–328. *In* M. Delgado (ed.) Proc. Vth. Int. Work. Meet. Soil Micromorph. Granada, Spain.

Buol, S.W., and H. Eswaran. 1988. International Committee on Oxisols: Final Report. USDA-SCS Soil Management Support Services, Technical Monograph No. 17. USDA, Washington, DC.

Camargo, M.N., and F.H. Beinroth (eds.). 1978. Proc. First Int. Soil Classif. Workshop. EMBRAPA-SNLCS. Rio de Janeiro, Brazil.

Charter, C.F. 1958. Report on the environmental conditions prevailing in Block A, Southern Province, Taganyika Territory, with special reference to the large-scale mechanized production of ground-nuts. Ghana Department of Soil and Land Use Survey, Occasional Paper 1. Ghana Government Printing Office, 37pp.

Cline, M.G. 1975. Origin of the term latosol. Soil Sci. Soc. Am. Proc. 39:162.

Daniels, R.B., E.E. Gamble, and J.G. Cady. 1971. The relation between geomorphology and soil morphology and genesis. Adv. Agron. 23:51–88.

Deenik, J. 1997. Liming effects on nitrate adsorption in soils with variable charge clays, and implications for groundwater contamination. M.S. Thesis, University of Hawaii. Honolulu, HI.

Ekern, P.C. 1967. Soil moisture and soil temperature changes with the use of black vapor barrier mulch and their influence on pineapple (*Ananas comosus* (L.) Merr.). Soil Sci. Soc. Am. Proc. 31:270–275.

Eswaran, H., H. Ikawa, and J.M. Kimble. 1986. Oxisols of the world, p. 90–123. *In* Proc. Int. Symp. Red Soils. Science Press, Beijing, China.

Eswaran, H., C.H. Lim, V. Sooryanarayanan, and N. Daud. 1978. Scanning electron microscopy of secondary minerals in Fe–Mn glaebules, p. 851–866. *In* M. Delgado (ed.) Proc. Vth. Inter. Work. Meet. Soil Micromorph. Granada, Spain.

Eswaran, H., and N.G. Raghumohan. 1973. The micro-fabric of petroplinthite. Soil Sci. Soc. Am. Proc. 37:79–81.

Eswaran, H., G. Stoops, and C. Sys. 1977. The micromorphology of gibbsite forms in soils. J. Soil Sci. 28:136–143.

Eswaran, H., and C. Sys. 1970. An evaluation of the free iron in tropical basaltic soils. Pedologie 20:62–85.

Eswaran, H., and R. Tavernier. 1980. Classification and genesis of Oxisols, p. 427–442. *In* B.K.G. Theng (ed.) Soils with variable charge. New Zealand Soil Science Society, Lower Hutt, New Zealand.

FAO–UNESCO. 1971–1976. Soil map of the world. FAO, Rome, Italy.

Harrassowitz, H. 1930. Bšden der Tropischen Region. Laterit und allitischer Rotlehm, p. 387–536. E. Blanck (ed.) Handbuch der Bodenlehre. Vol. 3. 387–536.

Herbillon, A. 1980. Mineralogy of Oxisols and oxic materials, p. 109–126. *In* B.K.G. Theng (ed.) Soils with variable charge. New Zealand Society of Soil Science, Lower Hutt, New Zealand.

Jones, R.C., W.H. Hudnall, and W.S. Sakai. 1982. Some highly weathered soils of Puerto Rico, Part 2: Mineralogy. Geoderma 27:75–137.

Jones, R.C., and G. Uehara. 1973. Amorphous coatings on mineral surfaces. Soil Sci. Soc. Am. Proc. 37:792–798.

Kellogg, C.E. 1949. Preliminary suggestions for the classification and nomenclature of great soil groups in tropical and equatorial regions. CAB Soil Sci. Tech. Commun. 46:76–85.

Kellogg, C.E. 1967. Comment, p. 232–233. *In* H.M. Southworth and B.F. Johnston (eds.) Agricultural development and economic growth. Cornell University Press, Ithaca, NY.

Lee, K.E., and T.G. Wood. 1971. Termites and soils. Academic Press, London, U.K.

Lepsch, I.F., and S.W. Buol. 1974. Investigations in an Oxisol–Ultisol toposequence in Sao Paulo State, Brazil. Soil Sci. Soc. Am. Proc. 38:491–496.

Lepsch, I.F., S.W. Buol, and R.B. Daniels. 1977. Soils-landscape relationships in the Occidental Plateau of Sao Paulo State, Brazil. Soil Sci. Soc. Am. J. 41:109–115.

Maignen, R. 1966. Review of research on laterite. UNESCO, Paris, France.

Mattson, S. 1928. The electrokinetics and chemical behavior of the aluminosilicates. Soil Sci. 25:289–311.

Mattson, S. 1932. The laws of soil colloidal behavior: IX. Amphoteric reactions and isoelectric weathering. Soil Sci. 34:209–240.

Miller, B.J. 1983. Ultisols, p. 283–323. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy, II. The soil orders. Elsevier Scientific Publishing Co., Amsterdam, the Netherlands.

Moormann, F.R. 1985. Excerpts from the circular letters of the International Committee on Low Activity Clay Soils (ICOMLAC). USDA-SCS Soil Management Support Services, Technical Monograph No. 8. U.S. Government Printing Office, Washington, DC.

Padmanabhon, E., and A.R. Mermut. 1996. Submicroscopic structure of Fe coatings on quartz grains in tropical environments. Clays Clay Miner. 44:801–910.

Parks, G.A., and P.L. de Bruyn. 1962. The zero point of charge of oxides. J. Phys. Chem. 66:967–973.

Plucknett, D.L. 1971. The use of soluble silicates in Hawaiian agriculture. University of Queensland Paper, Dep Agric. 1(1971), pp. 203–223, Vol. 1.

Robinson, G.W. 1951. Soils: Their origin, constitution, and classification. Thomas Murby & Sons, London, U.K.

Ruhe, R.V. 1956. Landscape evolution in the High Ituri, Belgian Congo. INEAC Ser. Sci. 66.

Sanchez, P. 1976. Properties and management of soils in the tropics. John Wiley & Sons, New York.

Sanchez, P., and J.G. Salinas. 1981. Low-input technology for managing Oxisols and Ultisols in tropical America. Adv. Agron. 34:279–406.

Santos, M.C.D., A.R. Mermut, and M.R. Ribeiro. 1989. Submicroscopy of clay microaggregates in an Oxisol from Pernambuco, Brazil. Soil Sci. Soc. Am. J. 53:1895–1901.

Sharma, M.L., and G. Uehara. 1968a. Influence of soil structure on water relations in low humic latosols: I. Water retention. Soil Sci. Soc. Am. Proc. 32:766–770.

Sharma, M.L., and G. Uehara. 1968b. Influence of soil structure on water relations in low humic latosols: II. Water movement. Soil Sci. Soc. Am. Proc. 32:770–774.

Simonson, R.W. 1949. Genesis and classification of red–yellow podzolic soils. Soil Sci. Soc. Am. Proc. 14:316–319.

Smeck, N.E., E.C.A. Runge, and E.E. Mackintosh. 1983. Dynamics and genetic modelling of soil systems, p. 51–81. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy, I. Concepts and interactions. Elsevier Scientific Publishing Co., Amsterdam, the Netherlands.

Smith, G.D. 1963. Objectives and basic assumptions of the new classification system. Soil Sci. 96:6–16.

Smith, G.D. 1965. Lectures on soil classification. Pedologie Society Special Bulletin No. 4. Ghent, Belgium, Belgishe Bodem Kundige Vereniging, pp. 134.

Smith, C.W. 1983. Soil survey of Islands of Palau, Republic of Palau. USDA-NRCS. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1975. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. USDA agriculture handbook no. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th edn. USDA-NRCS. U.S. Government Printing Office, Washington, DC.

Sumner, M.E. 1995. Amelioration of subsoil acidity with minimum disturbance, p. 147–186. *In* N.S. Jayawardane and B.A. Stewart (eds.) Subsoil management techniques. Lewis Publishers, Boca Raton, FL.

Sys, C. 1968. Suggestions for the classification of tropical soils with lateritic materials in the American classification. Pedologie 18:189–198.

Tavernier, R., and H. Eswaran. 1972. Basic concepts of weathering and soil genesis in the humid tropics, Proc. 2nd ASEAN Soils Conf. Vol. 1: 383–392. Jakarta, Indonesia.

Tavernier, R., and C. Sys. 1965. Classification of the soils of the Republic of Congo. Pedologie Vol: 91–136. Ghent, Belgium. Special Issue No. 5:91–136.

Tsuji, G.Y., R.T. Watanabe, and W.S. Saki. 1975. Influence of soil microstructure on water characteristics of selected Hawaiian soils. Soil Sci. Soc. Am. Proc. 39:28–33.

Uehara, G. 1995. Management of isoelectric soils of the humid tropics, p. 271–278. *In* R. Lal, J. Kimble, E. Levine, and B.A. Stewart (eds.) Soil management and greenhouse effect. CRC Press, Inc., Boca Raton, FL.

Uehara, G., and J. Keng. 1975. Management implications of soil mineralogy in Latin America, p. 351–363. In E. Bornemizsa and A. Alvarado (eds.) Soil management in tropical America. NC State University, Raleigh, NC.

van Wambeke, A. 1974. Management properties of ferralsols. FAO Soils Bulletin No. 23.

van Wambeke, A. 1992. Soils of the tropics: Properties and appraisal. McGraw-Hill, New York.

van Wambeke, A., H. Eswaran, A.J. Herbillon, and J. Comerma. 1983. Oxisols, p. 325–354. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy, II. The soil orders. Elsevier Scientific Publishing Co., Amsterdam, the Netherlands.

Wilding, L.P. 1994. Factors of soil formation: Contributions to pedology, p. 15–30. *In* R.J. Luxmoore (ed.) Factors of soil formation: A fiftieth anniversary perspective. SSSA Special Publication No. 33. SSSA, Madison, WI.

34

Land Evaluation for Landscape Units

J. Bouma
*Wageningen University
and Research Centre*

J.J. Stoorvogel
*Wageningen University
and Research Centre*

M.P.W. Sonneveld
*Wageningen University
and Research Centre*

34.1 Introduction

Sustainable development has become a key goal of various international agencies, and the role of the land in achieving this is increasingly being recognized. More efficient land use is needed to, for example, feed 9 billion people in 2050, to combat climate change, and to enable sustainable biofuel production (e.g., Hartemink and McBratney, 2008). Studies have explored the potential world food supply and associated environmental quality issues, among them land degradation (Penning de Vries et al., 1995). At another scale, new farming systems are being developed to optimize fertilizer and biocide use, applying modern information technology, summarized as precision agriculture (PA; CIBA Foundation, 1997; Bouma et al., 1997a; Stoorvogel et al., 2004c). Regional planners consider alternative land use scenarios combining agricultural production with nature conservation and other functions such as water storage (Bouma et al., 2008). In all cases, land use is the core issue to be studied, while socioeconomic considerations often play a crucial role as well. Consideration of actual and potential land use as a function of land properties fits under the broad umbrella of land evaluation as advocated by FAO (1976, 1983). Land evaluation

can be realized with descriptive, qualitative methods but increasingly includes quantitative simulation models for crop production, land use, and environmental impacts in combination with geographic information systems. All these methods need to be fed with adequate soil data, often derived from soil surveys (e.g., Bouma, 1994). This sometimes occurs mechanistically using existing databases with little attention to natural soil dynamics or landscape relationships, and this may lead to unrepresentative results. Soil survey has much to offer, but this expertise has to be applied and presented more effectively than at present. Soil survey interpretations in soil survey reports present broad possibilities and limitations for a variety of land uses for each mapping unit. However, these represent rather narrowly defined land uses as compared with land evaluation that enables the actual assessment of land performance. Aside from the variation in space, there is also the need to consider variation in time, be it days, growing seasons, or decades. Increasingly, land evaluation is realized in close interaction with the stakeholders, ranging from farmers to planners and politicians. Innovative developments in land evaluation, emphasizing interdisciplinary approaches at different spatial and temporal scales, will be discussed in this chapter.

34.2 Land Evaluation: The FAO Framework

34.2.1 Definitions

Land evaluation is defined as follows:

> the process of assessment of land performance when used for specified purposes, involving the execution and interpretation of surveys and studies of land use, vegetation, landforms, soils, climate and other aspects of land in order to identify and make a comparison of promising kinds of land use in terms applicable to the objectives of the evaluation.

> **(FAO, 1976)**

Clearly, evaluating the performance of land has been the topic of many studies in the past in different disciplines. There is, for instance, a large amount of literature on land use planning from a regulatory and social perspective, but here the focus will be on agroecology as related to sustainable development. Adopting the international FAO approach has the advantage of using a widely known procedure in which definitions are so broad that they offer many opportunities for expansion and modification. Two elements stand out when considering the definition of land evaluation: (1) Performance can only be assessed when a specific land use has been defined. In other words, judgment cannot be made as to performance in general but only for specific types of land use. Land may, for instance, function quite well as a campground but poorly when growing a wheat crop. (2) Attention is paid not only to current land use but also to potential forms of land use, which may be more or less promising depending on the objectives of the evaluation.

Land is defined as follows:

> an area of the Earth's surface, the characteristics of which embrace all reasonably stable, or predictably cyclic attributes of the biosphere, vertically above and below this area including those of the atmosphere, the soil and underlying geology, the hydrology, the plant and animal populations and the results of past and present human activity to the extent that these attributes exert a significant influence on present and future uses of the land by man.

> **(FAO, 1976)**

A land unit is defined as follows:

> an area of land possessing specified land characteristics and/or land qualities which can be demarcated on a map. Land is often represented as a georeferenced land unit on soil maps, which present additional data on climate in a soil survey report.

> **(FAO, 1983)**

The broad term "land use" is specified in terms of land utilization types (LUTs), which define a particular type of land use in varying degrees of detail. In the context of rainfed agriculture, an LUT refers to a crop, crop combination, or cropping system with a specified technical and socioeconomic setting (FAO, 1983). When combining the Land Unit with the LUT, one obtains the so-called land use system (LUS) defined as a specified LUT practiced on a given land unit and associated with inputs, outputs, and possible land improvements (FAO, 1983). Previous work in Costa Rica has defined such LUSs also in terms of the type of technology (T) being used in each particular production system. They refer, therefore, not to a LUS but to a LUST, an example of which is provided in Table 34.1.

A key element in land evaluation is the assessment of land performance. This is done, in principle, by comparing the requirements of a particular type of land use with what the land has to offer. When the two match, the land is suitable for a particular LUST. When they do not match, suitability is less to varying degrees. Land suitability is correspondingly defined as follows: *the fitness of a given type of land for a specified kind of land use* (FAO, 1976). This matching process, which is central in land evaluation, is handled by defining land qualities and land characteristics. Land qualities are *complex attributes of land that act in a manner distinct from the actions of other land qualities in its influence on the suitability of land for a specified kind of use.* The matching process is realized by expressing both land use requirements and what the land has to offer in terms of land qualities and by comparing the two expressions.

Although this is not mentioned by FAO (1976), land qualities usually cannot be measured directly. Examples are the moisture supply capacity, the workability, and the trafficability of land. They vary over the years and are determined by land behavior over extended periods of growing seasons. Modern methods will be discussed later to determine land qualities.

TABLE 34.1   A Listing of Data Documenting an LUT with LUST for Maize, Sown January 15 on a Fertile, Well-Drained Soil, Typical for the Neguev Area of the Atlantic Zone of Costa Rica

Operation	Date	Labor (h)	Equipment	Materials
Land preparation	January 31, 1991	20	Machete	
Herbicide application	January 2, 1992	10	Knapsack sprayer	2 L Gramoxome
Sowing	January 15, 1992	10	Planting stick	20 kg local variety maize seed
Fertilizer application	January 30, 1992	10		50 kg ammonium nitrate
Harvest	May 15, 1992	50		100 bags dry cobs

Source: Adapted from Jansen, D.M. and R.A. Schipper (1995). A static descriptive approach to quantify land use systems. Netherlands Journal of Agricultural Science 43:31–46.

However, in the older land evaluation work, attempts were made to find proxies for land qualities based on land characteristics, which are *attributes of land that can be measured or estimated*. In this context, one may think of soil texture, organic matter, carbonate content, etc.

The basic elements of "classical" land evaluation have now been introduced. One has land that is being used for a particular purpose in a particular way. Not only current land use should be looked at but also other possible forms of land use that are of interest. One wants to assess land performance for these different alternative forms of land use and does so by comparing land requirements for each alternative form of land use with what the land has to offer. This matching process is made possible by defining important land qualities, often defined in terms of land characteristics in different classes. The overall analysis results in statements as to relative suitabilities of a given piece of land for a series of LUSs. Much practical experience with this system is reflected in the work of Sys et al. (1991) who provide excellent case studies. Some examples to further illustrate the procedure and to discuss some underlying concepts will be examined.

34.2.2 An Example

An example given by FAO (1983) addresses land suitability for sorghum (Table 34.2) and may serve to illustrate some basic decision steps to be taken that have general validity. Suitability of the land unit being considered is expressed in terms of highly, moderately, marginally, and not suitable. Three land qualities are distinguished: rooting conditions, oxygen, and nutrient availability. This selection reflects the expert judgment of the land evaluator. Quite possibly, other evaluators would have selected other land qualities. Because land qualities, as mentioned here, cannot be measured directly, land characteristics are used as proxies, called "diagnostic factors," and are as follows: drainage class (soil survey reports), effective soil depth (to be estimated from structure descriptions in soil survey reports), and soil reaction (pH). Clearly, some selections reflect lack of good data. Soil drainage classes are very broadly defined and only remotely related to O_2 diffusion. The effective soil depth is more direct but does not consider any particular demands by sorghum; it could apply to any crop. The pH, finally, is a very general indicator for nutrient availability, and its selection reflects an apparent desperate lack of information on the soil fertility status. Finally, the evaluator has to couple classes of the diagnostic factors somehow to demands by the sorghum crop (crop requirements) to arrive at the suitability classes. In summary, there are three moments when important decisions have to be made based on expert knowledge and data from the literature: (1) selection of land qualities, (2) selection of land characteristics to describe the land qualities, and (3) selection of gradations of these characteristics to form suitability classes based on estimated crop requirements.

34.3 Beyond Classical Land Evaluation

The classical land evaluation scheme has been widely applied and has influenced many land evaluation methodologies (e.g., Sonneveld et al., 2010b). Its application has been facilitated by automated computer-driven decision support systems (e.g., Rossiter, 1990), often with a strong agroecological focus. However, five problems have become clear over the years:

1. Even though the need to define clear land evaluation objectives in close interaction with stakeholders has been stressed from the start, the development of a mechanistic, rather sterile approach has been observed in which land suitability is defined for a large number of LUSTs (Sys et al., 1991). It is not clear who is asking the questions or, worse, whether the answers being provided address the questions that are really being raised. Thus, several projects have failed because of a lack of stakeholder participation (FAO, 2007).

2. Defining land qualities in terms of land characteristics has become a rather rigid qualitative procedure, even, and particularly, in automated computer-driven decision support systems, allowing little input from modern process-driven land research focused on landscape processes and climatic change.

3. The procedure is almost exclusively driven by the properties of the land, and even though the importance of socioeconomic conditions is acknowledged, little is done to take these conditions into account. Land use and its possible changes are usually more a reflection of socioeconomic developments in society than of differences in soil suitabilities for different forms of land use. Implicitly, classic land evaluation assumes that *use* follows *land*, while in reality more often *land* follows *use*. Think of roads and shopping centers on prime agricultural land. This requires a different approach than the soil-centered one.

TABLE 34.2 Land Qualities and Land Characteristics (Diagnostic Factors) for a Sorghum LUT, Expressed in Terms of Crop Requirements

Crop Requirement			Factor Rating			
			Highly Suitable	Moderately Suitable	Marginally Suitable	Not Suitable
Land Quality	Diagnostic Factor	Unit	s1	s2	s3	n
Oxygen availability	Soil drainage class	Class	Well drained/ excessive	Moderately well drained	Imperfectly drained	Poorly/very poorly drained
Rooting conditions	Effective soil depth	cm	>120	50–120	30–50	<30
Nutrient availability	Soil reaction	pH	5.5–7.5	4.8–5.5 and 7.5–8.0	4.5–5.5 and 8.0–8.5	<4.5 and >8.5

Source: FAO. 1983. Guidelines: Land evaluation for rainfed agriculture. FAO Soils Bulletin No. 52. FAO, Rome, Italy.

Moreover, land units, as distinguished in earth science, hardly ever correspond with legal units on which decisions are made. For example, a farmer farms a field with most often different land units, not a single one. A district or county where land use decisions may be made may cut through different land units, etc.

4. The procedure was implicitly defined as being scale independent. Most applications of classical land evaluation have been at the regional level, but many land use questions are raised at farm or field level or at the continental or world level. Not only are the questions then quite different (see (1) above) but procedures to be followed should be different as well.

5. Finally, there was a lack of dealing with uncertainties. Map units, agronomic relationships, and climatic variability are characterized by high levels of uncertainty, requiring expression of land use options and their effects in terms of probabilities. This aspect was not considered in the FAO framework.

What is now needed is a better evaluation of questions being asked at different spatial scales. Next, proper procedures need to be defined to deal with these questions, realizing a wide variety of stakeholders are involved. And, finally, the proper role of the land in affecting land use decisions needs to be defined. Certainly, decisions are not made for land units but for georeferenced surfaces on the Earth that may contain many land units.

34.4 What Is the Question?

Questions regarding land use vary a great deal. Let us analyze some of them:

1. A farmer wants to know how he can obtain a high and secure yield of profitable crops at minimal cost, while operating in line with increasingly detailed environmental laws and regulations. Cutting costs to be achieved by, for example, precision application of fertilizers and biocides or minimum tillage is increasingly important. He certainly is not interested to hear that his land is moderately suitable for wheat growing. He will already know that. He wants quantitative information in terms of what to do and when, where, and how to do it within the context of given environmental regulations.

2. An environmentalist will ask how agrochemicals use can be reduced or even abolished, thus avoiding leaching into ground or surface water. To answer such questions, detailed process-based simulation models may have to be applied at both farm and regional level to estimate the adsorption and leaching of agrochemicals as a function of management. Of course, the ideal is to combine the desires under (1) with those under (2), which is the basic concept of PA.

3. A regional planner may want to formulate alternative land use options within a region all in the context of achieving sustainable development. Here, the existing land evaluation procedure may be useful, but data are not specific enough

to allow quantitative trade-offs among environmental, economic, and social considerations for each of the options.

4. A policy maker may see options for land use in a region or a country, formulated by procedures under (3). His question, however, is how attractive options can indeed be realized? How can stakeholders be influenced to do the right thing: special educational efforts, taxes, bonuses, subsidies? Clearly, land evaluation does not primarily focus on such issues, but they are increasingly important for land use and should, therefore, be considered because in the end the success of any scientific activity will be measured by its effect on society. If land evaluation reports wind up on the shelf, the activity itself will turn out to be nonsustainable. This fourth question can be handled by considering the policy cycle to be discussed later.

Considering the range of questions that may be encountered, an exploratory approach is attractive, defining a number of realistic land use options for the area to be considered. The stakeholders or policy makers make a choice from options presented. The type of question to be answered is as follows: What are the options for land use? How can one optimize land use for certain objectives and what are the trade-offs between these objectives? When dealing with sustainable land use, a balance has to be found for each option between environmental, economic, and social objectives. Different groups of stakeholders or different political parties will have different affinities with different options. The primary task of land evaluation is not to develop and select the "best" option but to present a series of well reasoned and transparent and alternative land use options allowing a rational selection process. Whether or not selected options are realized in practice depends on the stakeholders, but scientists can still be involved in the selection and implementation process because—as stated above—they also have an interest that some of the options they have derived will materialize in practice. This can be achieved by following the policy cycle, including signaling, designing, deciding, implementing, and evaluation (e.g., Bouma et al., 2007). The exploratory approach does not predict but explores what have also been called "windows of opportunity," a challenging concept.

Bouma (1993) has discussed five case studies on land use at different scales, showing that each was associated with quite different questions, while knowledge used in answering the questions varied very much as well. We therefore first analyze the different types of knowledge involved in land evaluation.

34.5 Some Basic Elements in Land Evaluation

34.5.1 Considering Different Types of Knowledge: The Knowledge Chain

A diagram (Figure 34.1) has been helpful to illustrate the use of different types of knowledge when analyzing land use questions at different spatial scales (Hoosbeek and Bryant, 1992; Bouma and Hoosbeek, 1996). They considered two perpendicular axes,
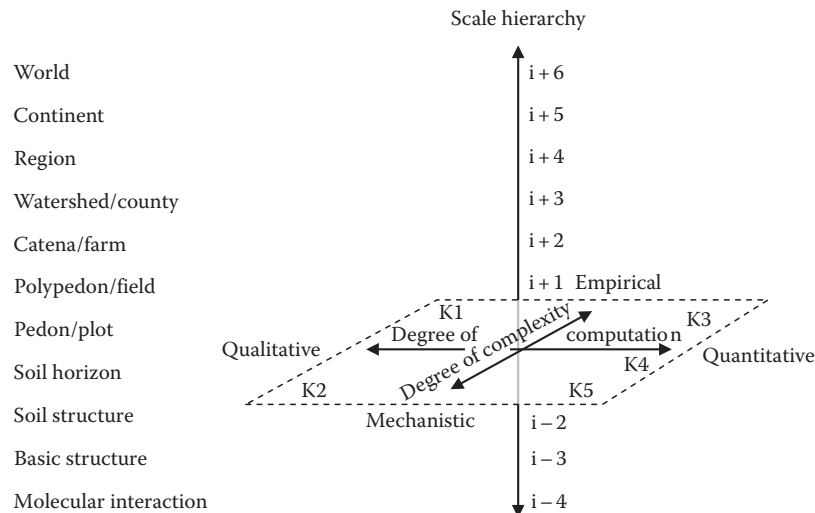
FIGURE 34.1   Scale diagram showing a series of hierarchical scales (i levels) and modeling approaches expressed in terms of four characteristics, which are summarized in terms of knowledge levels K1–K5. (Reprinted from Bouma, J., and M.R. Hoosbeek. 1996. The contribution and importance of soil scientists in interdisciplinary studies dealing with land, p. 1–15. *In* R.J. Wagenet and J. Bouma (eds.) The role of soil science in interdisciplinary research. SSSA Special Publication No. 45. SSSA, Madison, WI. Soil Science Society of America.)

one ranging from qualitative to quantitative and the other from empirical to mechanistic. Different research approaches occur within the plane thus obtained: K1 represents user knowledge; K2 represents expert knowledge; K3 represents knowledge to be obtained through semiquantitative models, in which real soil processes are not known; K4 represents knowledge through quantitative models where processes are characterized in general terms; and K5 represents the same, but processes are described in great detail, which can imply that the entire soil/crop system cannot be characterized anymore, and attention is focused on one aspect only (Bouma et al., 1996a). The vertical axis represents the scale hierarchy, where the pedon level (the individual soil) occupies the central position (i level). Higher levels are indicated as i+, while lower levels are i−. The scale in Figure 34.1 ranges from molecular interaction (i − 4) to the world level (i + 6).

One can now place the classic land evaluation in the scheme at scale hierarchy of watershed or region while the knowledge level is K2. For other questions raised above, one needs different hierarchies and knowledge levels. For example, the farmer and environmentalist would require the i + 1 field scale and a K4 knowledge level to get the necessary quantitative answers for their questions. The regional planner would operate at level i + 4 and would need at least a K3 knowledge level, because the K2 level would be too descriptive not allowing a quantitative trade-off analysis. A planner would be smart, though, to combine K2 with K3, by restricting the more detailed analyses of K3 to areas where a simpler K2 analysis could not provide answers. For example, Van Lanen et al. (1992) made a land evaluation for Europe in which potential for crop growth was established. Using soil maps, they first screened out strongly sloping land and land with shallow bedrock using a K2 approach. Then, in the remaining 40% of the land area, they ran a K3 simulation model to predict crop growth. The K2 approach for these soils (land is

moderately suitable for wheat) would not have been satisfactory. This scheme introduces the possibility of combining approaches. Another example is derived from a study of De Vries et al. (1992). They did a study to determine the possible impact of acid rain on soil acidification. They dealt with nonagricultural areas without fertilization. They divided Europe (scale i + 5) (Figure 34.2) into grids, and for each grid, they determined the dominant soil type, using the soil map of Europe (K2 knowledge). Then, they selected a limited number of soil units (level i) that were considered to be representative for European soils (using K2 knowledge). In these soils, they made some detailed measurements of weathering rates (scale i − 4; knowledge level K5). Next, this knowledge was scaled back up, resulting in what is assumed to be effective K4 knowledge at the European level. It would have been impossible to make the detailed K5 measurements of weathering rates in all European soils. By using expert knowledge at different scales, measurements could be made more efficient. Ideally, measurements should always include a measure for reliability and accuracy. An overall K5 approach for all soils would have the highest reliability but would be too expensive. One must know how much is lost in terms of reliability when one goes from K5 to K4, K3, and K2. Decisions as to what to do can only be based on this type of information. The lines in Figure 34.2 represent a so-called research chain demonstrating how a given problem can be analyzed by combining knowledge at different scales. The study by De Vries et al. (1992) had a disciplinary character, covering one particular aspect of land use. The four questions raised in Section 34.4 have, however, a much broader scope focusing on sustainable land use and require therefore a more comprehensive and interdisciplinary approach than was needed in the study by De Vries et al. (1992), including consideration of the role of different disciplines and of the policy cycle. Before discussing the latter two items, it is important to discuss the role
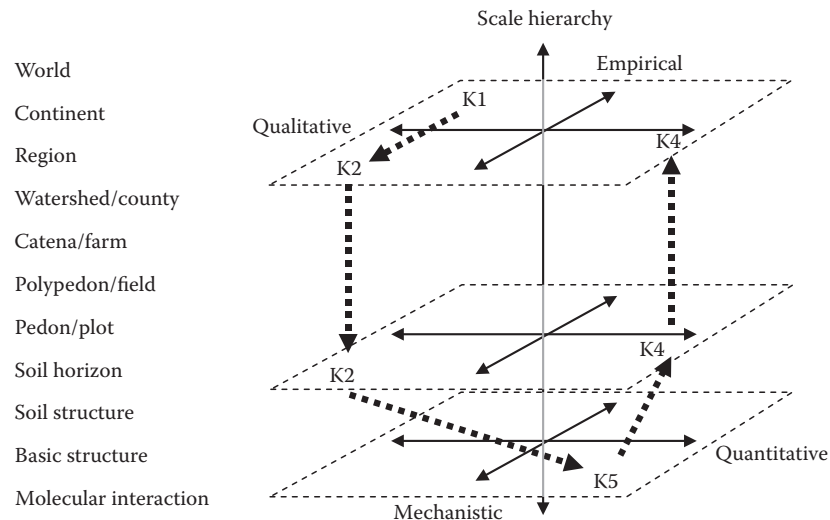
FIGURE 34.2 Scale diagram for the study by De Vries et al. (1992) on acidification of European forests. Three hierarchical scales are distinguished. At each scale, different methods corresponding to different knowledge levels are applied, forming a knowledge chain. (Reprinted from Bouma, J., and M.R. Hoosbeek. 1996. The contribution and importance of soil scientists in interdisciplinary studies dealing with land, p. 1–15. *In* R.J. Wagenet and J. Bouma (eds.) The role of soil science in interdisciplinary research. SSSA Special Publication No. 45. SSSA, Madison, WI. Soil Science Society of America.)

of soil scientists within broad, interdisciplinary land evaluation approaches involving many colleague scientists and often also stakeholders and policy makers. This role is quite different from the one described by FAO (1976, 1983, 1993) where soil scientists were in charge.

34.5.2 Applying Soil Science in the Broader Context of Land Evaluation

Soil science is back on the global agenda as evidenced by policy documents of several international agencies, listing the importance of soils in studying the Millennium Development Goals, climate change, environmental degradation, hunger alleviation, and production of biofuels (Hartemink, 2008; Hartemink and McBratney, 2008). In this context, land evaluation, as discussed in this chapter, can and should play a crucial role. There are, however, some potential difficulties (e.g., Bouma, 2009). Broad issues, as listed, require an interdisciplinary and participatory approach involving many scientists, stakeholders, and policy makers. To define a clear and visible role for soil science in such a complex context is not easy but necessary to allow input of relevant soil science expertise. Two suggestions are made here to strengthen this role. The *first* relates to three ways in which soil can be considered: (1) as an object to be studied with quantitative scientific methods ("it"); (2) as considered by society, expressed, for instance, by rules and regulations or by shear beauty ("we"), and (3) as viewed by the person studying the soil ("I"). Communication by soil scientists to larger groups often suffers when the three approaches are mixed as when personal feelings overly influence scientific study or by whatever society wishes to hear or feels. Being aware of these three views may help to make communications to the outside world more effective (Bouma, 2005). The *second* suggestion relates to the recently introduced

Soil Protection Strategy of the European Union (Commission of the EC, 2006) containing, in turn, two important suggestions: first, it defines seven soil functions that are crucial for land use, and second, it defines the DPSIR approach as a means to characterize soil dynamics. The DPSIR system defines the drivers of land use change (D), the associated Pressures (P) their impact (I) and Response (R) and the subsequent state of a given landscape (S). The seven soil functions are as follows: (1) production of food and biomass; (2) storage, filtering, and transformation of compounds; (3) habitat for living creatures and gene pool; (4) physical and cultural environment; (5) source of raw materials; (6) carbon pool; and (7) archive of geological and archeological heritage. (Note that the term soil, used in the strategy, is equivalent to the term "land," as used in this chapter.) Focusing soil research on these seven functions and consideration of the DPSIR approach can help to clarify and focus the role of soil science in the teams studying the issues mentioned above, all relating to sustainable development (Bouma, 2010). The next question is how to apply this in terms of defining exploratory procedures to obtain options for land use, as discussed in Section 34.4.

34.5.3 Defining Options for Future Land Use

A scheme (Figure 34.3) was introduced by Bouma et al. (2008) allowing the systematic characterization of the state (S) of a given landscape for which a land evaluation is to be made. First it is important to broadly characterize the area in terms of land and water dynamics and ecological conditions. This can be done at K2 level. Hydropedology is a relatively new research activity within the soil science community and is very helpful in this context (Lin et al., 2005, 2006). Also, mapping locations of roads, railways, and canals, as well as settlements, helps to define the context in which the land evaluation will take place. Next, the
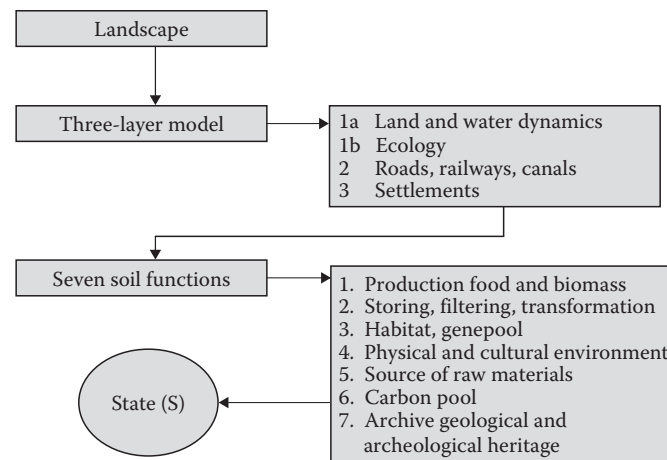
FIGURE 34.3 Flowchart indicating how the state (S) of a given landscape can be characterized by systematically considering the three-layer model and the seven soil functions.
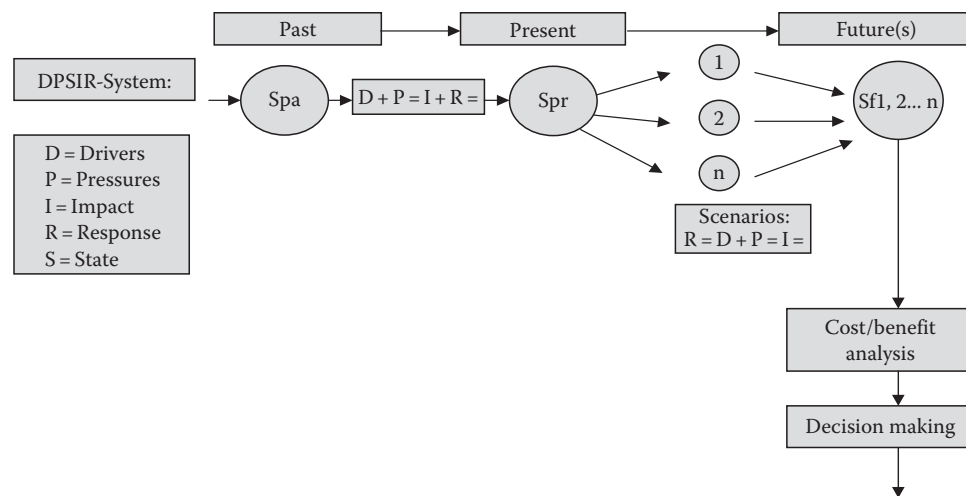

FIGURE 34.4 Illustration of the DPSIR procedure, illustrating how the past (Spa), current (Spr), and future states of landscapes (Sf1, 2, …, n) can be derived from considering D(rivers), P(ressures), R(esponses), I(mpacts), and R(esponses).

seven soil functions are systematically analyzed for the area, where, again, a selection has to be made of the level of knowledge to be applied for each function (see Section 34.5.1). The next step is to put this scheme in a time frame in terms of past, present, and future, where the latter is important for defining land use options as discussed in Section 34.4. Figure 34.4 illustrates the procedure. Drivers for land use change in the past and the associated pressures have led to certain impacts and responses. Effects of selected future options or scenarios are made visible by the responses to drivers and pressures of land use change that are associated with each option. As stated in Section 34.4, options are derived by discussions with stakeholders and policy makers.

34.5.4 Considering the Policy Cycle

Figure 34.4 could suggest that options derived are just presented to stakeholders and policy makers asking them to make a selection of the most attractive one. Even though the procedure of

deriving the various options involved interaction with stakeholders and policy makers, it is important for researchers to more strongly involve both types of users. Again, as discussed above, the success of any land evaluation will in the end be measured in terms of successful acceptance of whatever was created in the procedure. This requires substantial involvement over time. Bouma et al. (2007) have illustrated this by emphasizing the need to consider the policy cycle in planning the research process. This includes various phases:

1. Signaling of a problem, preferably based on a characterization of the current state S.
2. Designing, involving development of options, as discussed above.
3. Deciding: Here, scientists should keep their distance. They design but do not decide. Still, it can be quite effective to give some advice at the right time to keep the decision process moving along.

4. Implementing: Here, researchers have a more distant role, but when they are still involved at this stage, they can answer questions, make minor modifications if certain assumptions turn out to be incorrect, etc. We see many examples where nice options for land use have been derived but where the implementation fails in ways that could have been avoided.

5. Evaluation: Very important for all involved stakeholders including scientists as a learning opportunity.

When dealing with land use issues in land evaluation, scientists often restrict themselves to the design phase (2). To make sure that land evaluations have a real impact in society, we advise the broader involvement described here. Various examples are presented and discussed by Bouma et al. (2007) and Van Latesteijn (1995). Of course, the design phase will remain the core activity and to not become overstretched as a scientist is a real challenge. This calls for a certain division of responsibilities within the scientific community, corresponding with the particular abilities of each member of the team in dealing with various aspects of the policy cycle. In this context, Bouma (2005) proposed establishment of so-called *Communities of Scientific Practice*.

34.6 Operational Procedures

34.6.1 Three Basic Steps

The role and function of soils in a given region, following the approach indicated in Figures 34.3 and 34.4, will become clear when the proper research procedure is selected. This procedure includes three basic steps:

1. Problem definition in interaction with stakeholders and policy makers, resulting in a number of alternative options for land use that are expected to solve problems being identified. The problems are also interpreted in terms of what they mean for the seven soil functions. All functions are not necessarily relevant for each land area, but it is wise to analyze them all. The level of analysis (K1–K5) depends not only on the relevance of each function but also on data availability.

2. Selection of research methods to broadly characterize the land and water dynamics and the soil functions; again, implying a selection in the range K1–K5. This selection not only includes basic data but also simulation models of varying kinds. If not enough data are available, new research can be proposed and initiated. It is attractive to define research needs on the basis of an analysis of shortcomings that would result when using available data only as this is more convincing than research needs that are defined out of context. Attention should be paid to the accuracy and reliability of the various procedures to allow an overall risk analysis in the end, which should, ideally, be an integral part of each land use option being devised (Heuvelink, 1998).

3. Effective presentation of results using modern communication technologies to be focused on prime user groups and policy makers and an evaluation of impacts, including lessons learned.

When performing land evaluation procedures, including an analysis of past drivers of land use change (Figure 34.4), one should realize that there has been a shift in focus during the last decades. Exclusive emphasis on food production and food security after the Second World War has resulted in a technology explosion: problems in food production were there to be solved. Land that was too wet was drained; land that was too poor was fertilized, even at very high rates; land that was too dry was irrigated; and land where crops were suffering from pests and diseases was treated with biocides. Initially, these measures were taken with only food production in mind, and this has led to considerable pollution of land and water locally. Later, concern for the environment played an increasingly important role, and as this process was unfolding, and as a balance had to be struck between agricultural production, on the one hand, and environmental quality of soil and water, on the other, the importance of agroecological features of the land increased dramatically. There used to be the technology-driven spirit: anything can be done anywhere. Now one realizes again that a sand will never be a clay and that the natural dynamics of soils in different agroecological zones are the basis for developing sustainable production systems that are in harmony with natural conditions and the environment.

It is, of course, important to realize that soils do not occur in random patterns in a landscape. Soil scientists know this but the general public and many of our scientific colleagues are hardly aware of this. There is too much talk about "Soils" and "Land" in general terms while, for example, there are 1500 soil types in the Netherlands alone. Much work in soil science has been done on soil classification, grouping soils that are comparable in their basic soil properties. Such groupings are used to describe soil behavior in terms of suitability for a given use (see Section 34.2). But the procedures, illustrated in Figures 34.3 and 34.4, allow extension of this limited approach toward defining characteristic: "windows of opportunity" for any given soil series (which is the lowest hierarchical unit of classification). Different forms of management will lead to different soil conditions within the same soil series, the range of conditions (the window) will be characteristically different for each series. Note that in soil classification, the effects of management have always been ignored for the simple reason that classifications should be more or less permanent and should not change as a result of normal agricultural management. But, in contrast, for land use and land evaluation, the effects of management are very important and need therefore to be associated with the classification.

34.6.2 Considering Effects of Management: The Phenoforms

Droogers and Bouma (1997) studied a prime agricultural soil in the Netherlands (a fine, mixed, mesic Typic Fluvaquent according to Soil Survey Staff, 1998). They suggested the terms genoform for the genetic soil name and phenoform for the management variant, of which three were studied, one resulting from 60 years of organic farming (BIO), one from modern intensive arable farming (CONV), and one from continuous grassland (PERM).
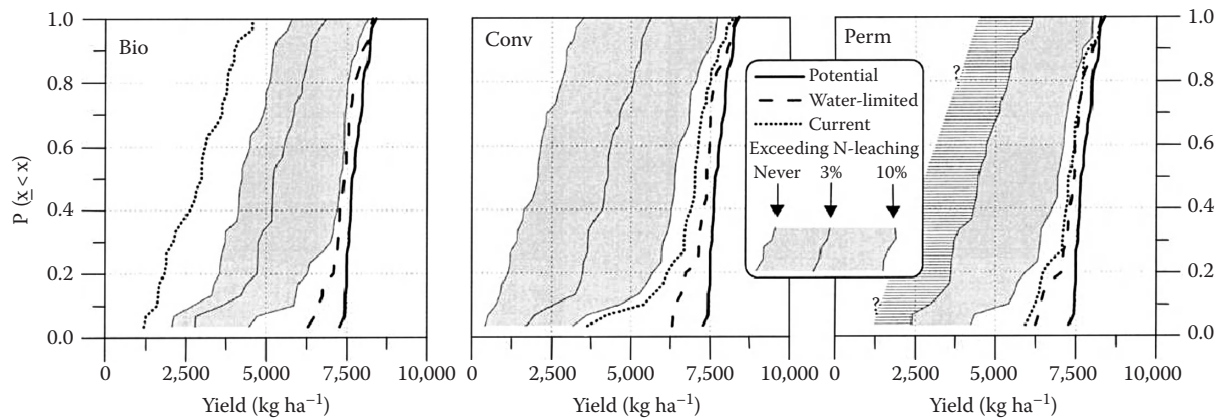
FIGURE 34.5 Cumulative probability function of simulated yields for the defined scenarios (see text). Probabilities were obtained by using 30 years of climatic data and are expressed for three probabilities that the threshold for nitrate leaching are exceeded. (Reprinted from Droogers, P., and J. Bouma. 1997. Soil survey input in exploratory modeling of sustainable soil management practices. Soil Sci. Soc. Am. J. 61:1704–1710. Soil Science Society of America.)

They used the validated WAVE model (Vanclooster et al., 1994) to calculate crop yields for wheat as a function of a wide range of N fertilization scenarios. By using yield calculations for a 30 year period with a wide range of weather conditions, they could express both yields and NO_3^- leaching to the groundwater in probabilistic terms (Figure 34.5). This, in turn, allowed the identification of a soil quality indicator for yield-expressing effects on the environment and risk. This indicator was defined by dividing actual calculated production (×100) by potential production where the latter is characteristic for each agroecological zone (Bouma and Droogers, 1998). Pulleman et al. (2000) applied the genoform/phenoform concept to the same soil series and after studying 50 fields, they derived a surprisingly robust regression equation, relating the organic matter content of surface soil (SOM) to previous land history as follows:

$$\text{SOM (\%)} = 20.7 + 29.7\ C_p + 7.5\ C_c + 7.5\ M \qquad (34.1)$$

where

C_p is the past crop type (31–63 years ago)
C_c is the current crop type (1–3 years ago) with grass = 1 and arable = 0
M is the management type (3–7 years ago) with organic = 1 and conventional = 0 ($r^2 = 0.74$)

Sonneveld et al. (2002) found a similar relationship for a common type of sandy soil in the North of the Netherlands, a coarse loamy, siliceous, mesic plagganthreptic alorthod (Soil Survey Staff, 1998) where soil organic carbon (SOC) was related to land use as follows:

$$\text{SOC (\%)} = 3.40 - 1.54 \times \text{Maize} + 0.19 \times \text{Old} + 0.55 \times \text{GWC} \qquad (34.2)$$

where

Maize is 1 for continuous maize, otherwise 0
Old is 1 for old grassland, otherwise 0
GWC is groundwater class in which 1 for class VB and 0 for the dryer class VI

Distinguishing well-defined phenoforms of existing soil series can be quite effective in extending the use of existing soil survey interpretations. Such effects are of major interest to many users and also define a meaningful new and innovative field procedure once the systematic soil mapping of land has finished, as is the case in many countries. Moreover, these findings illustrate the need to view soils as continuous spatial bodies that respond to land use.

This approach can be used at different scales once a database is filled with a number of phenoforms for any given soil series. Larger landscape units on small-scale maps, such as soil associations, are still defined in terms of their internal composition, for example, 30% soil series x, 50% series y, and 20% series z. Then, the effects of different types of management on yield and environmental side effects can still be expressed in a probabilistic manner, be it nongeoreferenced. Also, the fact that mapping units in the field may have quite some internal variability should be considered when making interpretations (Mausbach and Wilding, 1991; Bouma et al., 1996b; Young et al., 1997). The current development of digital soil mapping techniques, using, for example, gamma-ray spectrometry, does open up new possibilities for mapping soil properties at high spatial resolutions for relatively large areas.

34.7 The Crucial Importance of Effective Interaction with the Stakeholders

Interaction with stakeholders has been emphasized many times in the above sections during problem definition and the research process including final reporting. In the past, much research has been top down and supply oriented, which has led to the failure of several development projects (FAO, 2007). The researcher had an impression as to what the problem was that needed to be investigated, and he or she pressed forward using a favorite model, expert system, or data-gathering technique. In the end, the results of the study were presented to the stakeholders.

For sure, there are many examples of fine and effective research being executed in this way that have led to successful implementation. However, there are also too many examples of research efforts that were less successful, ending up in a desk, covered with dust. The challenge, now, is to involve stakeholders to the extent that research is being executed jointly with frequent interaction, with the desirable effect that the end result of the work is experienced as a joint product. This is easier said than done, and it may take a lot of time as illustrated by Bouma et al. (2008) for a regional study of environmental quality in the Northern Frisian Woodlands (NFW) in the Netherlands (see Section 34.8.5). Necessary interaction with stakeholders and policy makers during the research process is not only time-consuming but also potentially perilous. Any researcher has the obligation to make independent studies, whether or not potential results are liked by the stakeholders or not. If he or she gets too involved, objectivity may indeed suffer. This is a reason for many scientists to avoid interaction altogether. Though understandable, it is unwise certainly in the area of land use research that is so strongly tied to societal concerns (Bouma, 2009).

Modern information technology has an important role to play in stimulating interaction with stakeholders. Visualization of alternative land use patterns associated with different options is a very powerful tool to involve stakeholders. A picture says more than a thousand words. Interactive computer technology allows, for instance, instant generation of alternative land use scenarios with all associated input data by researchers and stakeholders. Also, the mentioned improvement of results when moving from K2 to K3, K4, and K5 approaches can be visualized as well by showing the accuracy of the land use maps obtained (Bouma, 1997b). The stakeholder can decide whether or not the costly improvement of the product is worth the cost. As Bouma (1993) has pointed out, several problems can be solved well with a relatively cheap K2 approach based on soil survey reports. This may be scientifically less challenging, but it is important from a practical point of view.

34.8 Case Studies

Six case studies will be discussed covering three spatial scales: the farm, the regional, and the world level. The three steps in the research procedure, as described in Section 34.6, will be followed for each case. Cases will be reviewed broadly, and reference will therefore be made to more detailed source publications.

34.8.1 Precision Agriculture: Arable Farm in the Netherlands

34.8.1.1 Problem Definition and Associated Soil Functions

PA is receiving increasing attention because fine-tuning of the application of fertilizers and biocides to the temporal needs of plants within spatially variable fields can be attractive from both an economic and ecological point of view (CIBA Foundation, 1997). By not applying more chemicals than are used or necessary, costs and losses to the environment are minimized: a "win-win"

proposition. Prime agricultural lands in alluvial deposits all over the world are stratified. Sandy and more clayey spots alternate at small distances. Also other soils, for example, those formed in glacial deposits, may be quite heterogeneous. In this study, precision fertilization and application of biocides were tested in an operational farm setting to help fine-tune management with the objective of maximizing production while minimizing adverse environmental side effects, particularly nitrate and biocide pollution of groundwater. Research results from the past, as reflected in current fertilization advice by extension services, cannot be used for this fine-tuning as they base their advice on simple and highly generalized input-response functions derived from many different plot experiments under undefined environmental conditions. Also, the fertilizer advice implicitly assumes fields to be homogeneous. Existing exploratory or reconnaissance soil maps cannot be used either because they are not detailed enough to characterize fields. A more fundamental problem is the fact that pedological differences, even when expressed on large-scale soil maps, do not necessarily correspond with functional differences in terms of nutrient and biocide transformations, as defined by soil function 2. Functional characterization of soil is therefore a key element of PA as it is "translated" into so-called management units for the farmer, each of which requiring particular treatment in time. Their boundaries do not necessarily correspond with those of mapping units on soil maps. In all of this, aside from function 2, function 1 is crucial, while function 6 may become important as increasing the organic matter content by management in the long run will improve functions 1 and 2 (see Section 34.5.2).

34.8.1.2 Method Selection and Results

The need to define N dynamics and biocide adsorption in soil in relation to crop development requires the use of a quantitative mechanistic K4 model. More empirical and qualitative approaches are inadequate because environmental regulations are strict, and research results are only convincing when obtained at least K4 level. The WAVE model was therefore applied (Vanclooster et al., 1994). Even though a complete 1:50,000 soil map is available for the Netherlands, we lack the necessary detailed 1:5,000 maps to identify differences within individual fields. Besides, the concept of representative soil profiles is used for mapping units even on such large-scale maps and individual borings taken during mapping are not georeferenced. We therefore decided to do a new detailed soil survey for the 110 ha farm being studied (Van Alphen and Stoorvogel, 2000a). The procedure consisted of (1) detailed soil survey; (2) using geostatistics to estimate optimal boring distances; (3) making the soil borings; (4) performing simulations for each boring in the context of functional characterization, leading to the definition of management units to be used in precision management. Pedotransfer functions were used to obtain physical parameters for the simulation model, by relating them to basic soil characteristics (Van Alphen et al., 2001), and (5) perform simulations for each boring on the adsorption of various biocides being used in practice. Functional characterization (point 4 above) is based on items that are relevant for the particular arable system being studied and is site specific. In this case,

four items were relevant to characterize N dynamics: (1) water stress in a dry year, (2) N stress in a wet year, (3) N leaching from the root zone in a wet year, and (4) residual N content at harvest in a wet year. Simulations of N dynamics were made for each boring, and management units were derived with fuzzy clustering and kriging interpolation techniques (Van Alphen and Stoorvogel, 2000a, 2000b). Calculations for three biocides were also made for each point with the WAVE model, focusing on simulated leaching values and associated risks. Patterns of leaching risks for each biocide were obtained by interpolation. Here, the management units, derived for N fertilization, were not used (Van Alphen and Stoorvogel, 2002). The operational farm system for PA consisted of real-time daily modeling with WAVE, using weather data. The stock of nitrogen in the root zone was tracked by calculated crop uptake and leaching as a function of time. As soon as a threshold was reached, the farmer was advised to fertilize. This moment is different in each different management unit (Figure 34.6), as the water and nutrient dynamics are different, even though the weather is identical. Unit 1 needed three fertilizations, and unit 3 needed four. The PA system reduced the amount of applied fertilizers with 25% as compared with the existing system based on generalized functions and resulted in significantly less leaching (Van Alphen and Stoorvogel, 2000b). The biocide studies showed significantly different risks being associated with using the different biocides, and the farmer was advised to use certain biocides only in certain parts of the farm, while using expensive biocides only in parts where risks were particularly high (Van Alphen and Stoorvogel, 2002).

34.8.1.3 Impact and Lessons Learned

This study on PA clearly showed the significant impact of fine-tuning the application of fertilizers and biocides to the varying needs of the plant in time. The soil science focus of this study is rather unique. Much PA work focuses on using remote or proximal sensing to establish crop conditions and to define management units. There has been much progress in developing these sensing techniques. For example, the N status of crop leaves can be well established by now. However, fertilization is already too late when applied after observing a low N status of plant leaves:

growth retardation has already occurred. The soil-focused procedure, illustrated above, provides a signal before the N supply is too low, leading to N stress in the plant. It also allows estimates of possible leaching of nutrients, which is, of course, not possible with remote sensing. A second very powerful thrust in the PA research arena is the technical focus, developing global positioning systems (GPS) equipment and machinery allowing PA procedures. Unfortunately, the various approaches have not been combined and integrated well enough and that is one of the reasons that PA has not been adopted as widely as one would have expected and as one would wish. One lesson learned is that results from soil-centered studies have to be communicated more effectively to provide a counterweight to one-sided commercially motivated approaches by other disciplines. Finally, the soil studies, briefly reviewed above, were only possible thanks to the rapid development of information technology and computing power. As such, this offers unique and quite attractive new opportunities to "assess the performance of land" as required in land evaluation. Here, no abstract judgment is given as to how suitable a soil might be for a given use, but an operational procedure is presented for a farmer that can help to reach the potential of any given soil within commercial units of production, such as fields.

34.8.2 Precision Agriculture: Banana Finca's in Costa Rica

34.8.2.1 Problem Definition and Associated Soil Functions

Where European agriculture is being dictated by a plethora of legislation controlling agricultural management at the national and European level, institutional infrastructure to enforce legislation is lacking in most tropical countries. Despite the fact that national legislation is minimal, the producers of export crops are increasingly being constrained by the importing countries. A good example is the international banana sector in Costa Rica. Costa Rica provides over 25% of the bananas consumed in Europe. Banana producers are influenced in several ways by the market: (1) The European Union has supported the banana production in African, Caribbean, and Pacific (ACP) through
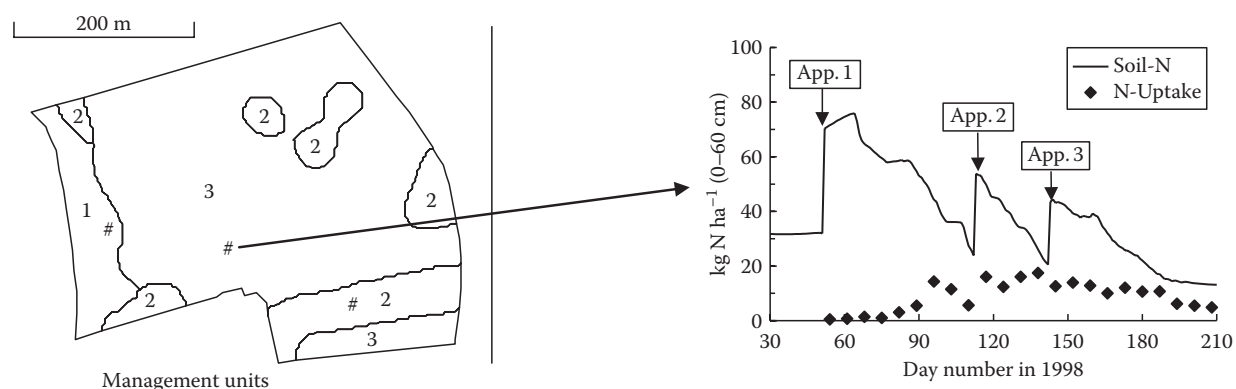

FIGURE 34.6 Management units for a Dutch arable farm with real-time simulations for soil available N contents for management unit 3 (line; kg nitrate-N ha^{-1} in the top 60 cm) and crop N uptake rate (dots; kg N ha^{-1} week^{-1}).

a complex tariff quota system restricting imports from, among others, Costa Rica; (2) the use of biocides is constrained by the regulatory agencies in the United States and Europe. They determine which pesticides are allowed and how they can be applied; (3) a plethora of different labels have been developed in Europe and the United States focusing on chemical use (organic bananas), social production conditions (fair trade), and the environment (ISO 14001, the environmental management standard); (4) in many cases, producers have to follow the guidelines of these labels to be able to export; and (5) large buyers of bananas (e.g., supermarket chains) force the banana plantations to produce under a detailed tracing and tracking system. Given the above, it is clear that the production has moved away from the small farmer. Currently, bananas are produced at large plantations that are either owned by independent producers (50%) or multinational companies (50%). All producers are organized in the Costa Rican Banana Corporation (Corbana). The banana producers have to operate in this complex set of political conditions. That does not only require management adaptations, it also requires the development of alternative management strategies and coordination among producers. Already in 1992 when the environmental lobby pushed hard on the banana sector, Costa Rican producers organized themselves in the Environmental Banana Commission aiming at the improvement of the production system in a more environmental friendly way. Before a plantation can be established, a land suitability assessment has to indicate that the land is suitable for banana production. As a result, Costa Rican banana production takes place under very favorable conditions. One of the plantations, the Rebusca plantation, has been very innovative in the past and maintains close links to the research department of Corbana. The Rebusca plantation is located in the perhumid lowlands in the northeast of Costa Rica (10.5°N, 84.0°W). The plantation measures approximately 124 ha of which 111 ha is used for the cultivation of banana. Like many plantations, the plantation exhibits significant soil variability with soils ranging from Typic Udivitrands to Andic Eutropepts and Typic Dystropepts (Soil Survey Staff, 1998). Like most Costa Rican banana plantations, the production of bananas is intensive with high use of agrochemicals in terms of fertilizers (2400 kg ha^{-1} year^{-1}), fungicides (48 aerial applications per year), and nematocides (3 applications per year). During the last decade, Rebusca has exported through different production channels ranging from a small exporting company owned by a few independent producers to sales through multinational companies. In the first case, production techniques were determined by plantation management, and although they had to fulfill European regulations, there was relatively little control. Exports through the multinationals had many more restrictions and controllers frequently visited the plantation. As a commercial enterprise, the first objectives of management are economical and can be described as profit maximization. Nevertheless, the owner of the plantation is also looking at the environment and tries to minimize the use and indirectly the emissions of agrochemicals to the environment. In contrast to other plantations, the owner is seeking innovations in the production system. This can be explained

in part by a vision toward the future where he expects environmental and social regulations to tighten. After confronting the owner with the concepts of PA, the owner was highly interested. He saw major advantages in terms of chemical use efficiency and the registration of production activities. Just like the previous PA case, the focus is on soil functions 1 and 2 to produce efficiently with minimum losses to the environment.

34.8.2.2 Method Selection and Results

An integrated system for PA was developed for the Rebusca plantation. A detailed description of the system and the process of implementation is presented in the study by Stoorvogel et al. (2004c). In contrast to common practices in PA, we could not make use of standard yield mapping techniques. In addition, the technological focus of PA was not easy adoptable in a continuous cropping system with almost no mechanization and managed by manual labor. Banana plantations in Costa Rica are frequently (2–3 times per week) screened for bunches ready for export. Bunches are harvested and transported by a dense cable system (every 100 m) to the packing plant. For yield monitoring, groups of bunches are coded based on their origin (cable number and the location within the cable). A weighing balance is installed in the main cable just before the packing plant. Codes are registered at the balance, and bunches are weighed. A software package denominated BanMan was developed for data processing and to create and analyze the yield maps. An example is presented in Figure 34.7. In addition to the yield maps, a detailed soil survey has been carried out for the Rebusca plantation Stoorvogel, J.J., Kooistra, L., and Bouma, J. (1999). Spatial and temporal variation in nematocide leaching, management implications for a Costa Rican banana plantation. In: Corwin, D.L., Loague, K., and Ellsworth, T.R. (eds.). Assessment of non-point source pollution in the Vadose Zone. Geophysical Monograph 108. p. 281–289. Key element in an effective system for PA is the translation of site-specific soil and production data into management recommendations. There is an increasing call for the use of crop growth simulation models. However, we have to realize that for many tropical crops, these simulation models do not exist. In this case, we analyze the yield maps in relation to the soil conditions and determine the location of so-called problem areas. These areas are characterized by a relative low production compared to the average production for areas with similar soil conditions. This is an important signaling function after which more detailed studies have to pinpoint to the exact causes of the low production and management can intervene. In addition, fertilizer experiments were carried out on the three main soil types in the plantation where during a period of a year 1 ha blocks received 75%, 100%, and 150% of conventional fertilization. In these blocks, crop performance as well as nutrient leaching was monitored. Depending on the soil type, changes in fertilization resulted in changes in production and/or nutrient leaching. Results for the three main soil types are presented in Table 34.3. In soils 1 and 2, almost all of additional nitrogen is taken up by the plant if we move from 75% to 100% fertilization. However, further increasing fertilization resulted in increased leaching.
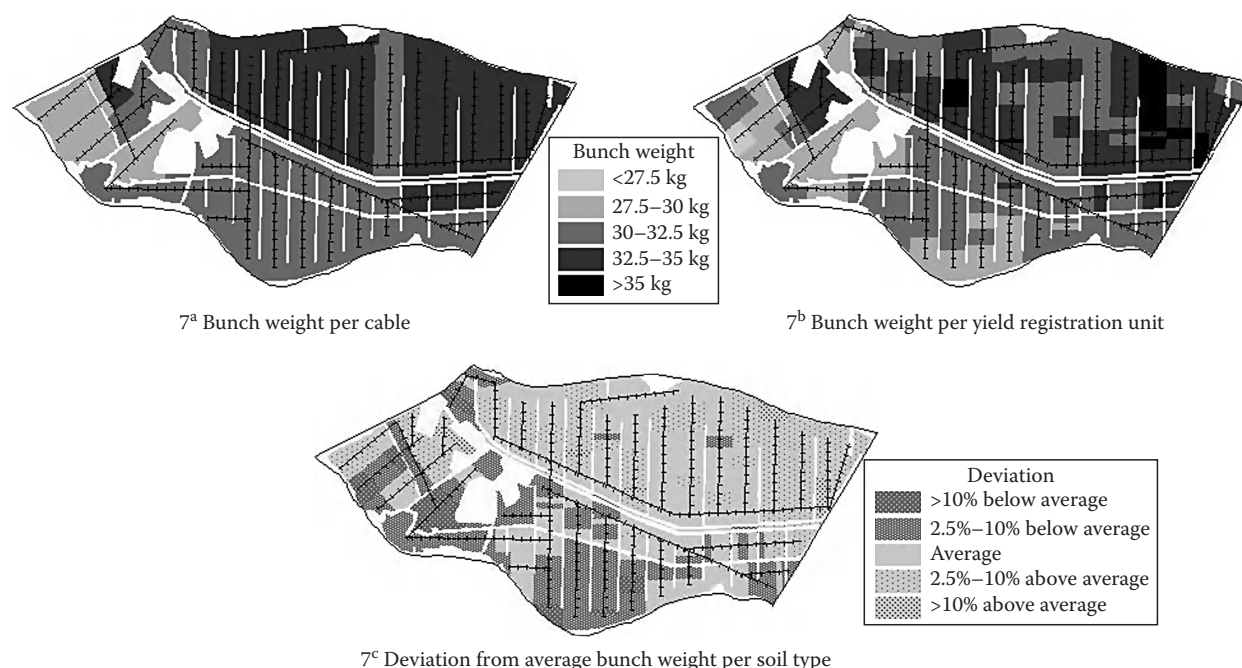
7ᵃ Bunch weight per cable

7ᵇ Bunch weight per yield registration unit

Bunch weight
- <27.5 kg
- 27.5–30 kg
- 30–32.5 kg
- 32.5–35 kg
- >35 kg

Deviation
- >10% below average
- 2.5%–10% below average
- Average
- 2.5%–10% above average
- >10% above average

7ᶜ Deviation from average bunch weight per soil type

FIGURE 34.7   Yield maps and problem areas for a Costa Rican banana plantation.

TABLE 34.3   Nitrogen Concentrations in Soil Water

Soil	Treatment (%)		
	75	100	150
1	3.61	3.81	4.96
2	4.12	4.08	6.36
3	1.91	2.34	3.29

In soil 3, however, leaching already increased moving from 75% to 100% (with a stable production) indicating that fertilization could be reduced. The low production was caused by the heavy soil textures, poor internal drainage, and low pH. These results formed the basis for site-specific fertilizer recommendations. The recommendations resulted in a decrease in fertilizer use of 12%. An important new development in banana management has been the renovation of areas that experience a decline in production and for which no clear cause can be identified. Often, the deterioration of the plants has been given as the main cause. Due to the site-specific yield monitoring, we can accurately identify the areas that may improve through renovation.

34.8.2.3 Impact and Lessons Learned

The experiences with PA in the banana crop have been very favorable. Farm management has been able to improve the productivity of the plantation, and it has provided them with an improved insight into the performance of the plantation. In addition, they are equipped with an excellent system for the registration of farm management and productivity to answer questions in terms of tracking and tracing. Since the concepts of PA have been introduced in the Costa Rican banana sector, many companies and plantations have used particular elements of PA. A good example

is the site-specific application of nematocides in those areas where nematode concentrations exceed the threshold of 16,000 per 100 g of roots. However, the spatial variation in nematode populations makes sampling and detection an awkward practice. The success of PA at the Rebusca plantation can be attributed to the integrated approach in which all the elements of the cropping system are included and the intensive involvement of the owner of the plantation in the development and implementation phase. Currently, the banana sector is under heavy pressure. Corbana recognizes the value of PA and starts to link their extension work to PA. The strategy of Corbana is that the sector needs to take a proactive approach. Tighter regulations with respect to the environment, the social conditions of workers and tracking and tracing are to be expected and the plantations have to be well prepared. This study presents a high-tech, interdisciplinary approach, developed in close and continuous interaction with the banana growers, to the "assessment of the productivity of the land." In fact, this PA system, operating in a developing country, functions much better than PA systems in so-called developed countries.

34.8.3 Organic Farming in the Netherlands

34.8.3.1 Problem Definition and Associated Soil Functions

The need to create sustainable agricultural production systems that are economically feasible, ecologically acceptable, and socially attractive has led to much discussion in society as to the possible effects of introducing alternative farm management systems. Many of these discussions have an ideological character. For example, to some, organic farming serving regional markets presents a unique road to sustainable development.

To others, industrialized agriculture with a global perspective offers better perspectives. Obviously, this type of discussion is beyond the scope of this text. It is, however, of interest to see if and, if so, how soil conditions may change as different management practices are followed in a given type of soil. This way soil scientists can have independent input into broad discussions on sustainable development. This line of thinking is different from the classical one followed in soil survey interpretations where each soil series is defined in terms of a number of use limitations for a variety of land uses, implicitly assuming that soil properties within a given soil series remain the same, independent of use. To assess the effects of management on soil properties of a given soil series, representing prime agricultural land in the Netherlands, we compared soil conditions on an organic and a conventional arable farm. In contrast, also permanent grassland was added as a variant (Droogers and Bouma, 1996, 1997; Droogers et al., 1996). Several soil functions (Section 34.5.2) are relevant here: (1) biomass production, (2) transformations in the soil, (3) habitat and gene pool, and (6) acting as a carbon pool, function (4) does to a certain extent relate to organic farming as this tends to better preserve the local traditional character of the land as compared with industrial agriculture.

34.8.3.2 Method Selection and Results

Soil conditions can be well expressed in terms of land qualities (FAO, 1976; see Section 34.2.1), which are the moisture supply capacity, nutrient supply, trafficability, and workability, representing a K2 approach. This, however, is descriptive and not diagnostic. We, therefore, decided to use a deterministic, quantitative K4 model (WAVE; Vanclooster et al., 1994) that allows quantitative characterization of soil moisture regimes and associated N transformations. Common soil properties such as bulk density, porosity, moisture retention, and hydraulic conductivity reflect the effects of short-term management. Tillage or soil traffic under wet conditions or poor management may in a given year lead to compaction, puddling, and structure degradation, which may not occur in the same soil where soil traffic is avoided or where the farmer does a better job. We, therefore, focused here on soil properties that are not significantly influenced by short-term management such as the organic matter content, which is affected by long-term management in terms of decades. Moisture retention and hydraulic conductivity data needed to be measured because estimates with pedotransfer functions could not be made as they did not cover the range of organic matter contents observed. To adequately represent the effects of different contents of organic matter on N mineralization, we measured rate constants to be used in the N module of WAVE. Thus, a relatively high data demand materialized. Simulation runs were made for 30 year periods using real weather data, allowing expressions of production and leaching of NO_3^- in probabilistic terms. Predictions of moisture contents in surface soil allowed estimates to be made for trafficability and workability (Droogers et al., 1996). Runs for 1995 were compared with measured data, showing that the model performed satisfactorily. Validation was thus assured, but no thorough sensitivity or error propagation analysis was performed.

Results were presented in graphical form. The increased organic matter content originating from organic farming resulted in a 20% increase of potential productivity (Droogers and Bouma, 1996, 1997). However, trafficability and workability decreased because the higher moisture contents during the year resulted in shorter periods with adequate trafficability and workability (Droogers et al., 1996). Indeed, strong compaction was observed in fields of the organic farm where conventional plowing to 30 cm depth was used. We also addressed the question as to how production and NO_3^- leaching could be balanced, using a probabilistic graphical expression (Figure 34.5). This graph was also used to define soil quality in terms of production and leaching of NO_3^-, illustrating that the user must choose the level of risk he or she is willing to take. This risk relates to yields (probabilities that a certain yield is exceeded) and to NO_3^- leaching (probabilities among the years that the threshold value for NO_3^- leaching is exceeded).

34.8.3.3 Impact and Lessons Learned

Organic farming resulted in higher organic matter contents that, in turn, allowed higher potential productions because of a higher water supply capacity. However, risks of causing structure degradation were higher as well, and it was shown clearly that traditional plowing to 30 cm depth did indeed result in such degradation. We advised to explore minimum tillage as an alternative tillage practice, and this was followed up by the farmers. This study also led to the initiation of a broader study relating past land use successfully to organic matter content of surface soil, as reported in Section 34.6.2 by the study of Pulleman et al. (2000). The K4 land evaluation approach followed in this chapter is a far cry from the traditional K2 procedure. But it still focuses on the central issue of land evaluation, that is, "the assessment of land performance when used for specified purposes," as it addresses real questions of the land user. Computer simulations for a series of years allow expressions in terms of risks, and this is important: scientists do not only provide a verdict as to the suitability of a given form of land use but define risks associated with particular actions. It is up to the land user to decide what to do based on the risks he or she is willing to take. One lesson learned was the fact that this study defined a welcome "niche" for soil science in the very broad and often emotional societal discussion about sustainable land use and the role of organic farming. Looking back, we should have pursued this approach systematically for other major soil series offering a model for post soil mapping research. In fact, this type of work was done later only for a major soil series in sandy soil (see Section 34.6.2).

34.8.4 Alternative Land Use in Ecuador

34.8.4.1 Problem Definition and Associated Soil Functions

Although the commercial farmers of the Carchi region in Northern Ecuador are probably some of the better endowed in the rural communities of the Andes, it became increasingly apparent that the use of pesticides in intensive potato production

was not only a blessing but a curse as well. In the early 1990s, researchers, farmers, and NGOs paid increasing attention to the negative effects of the intensive use of pesticides. The farmers in the Carchi region were very much dependent on these pesticides to control pernicious pests and blight. Intensive on-farm research revealed some of the major health and environmental impacts associated with these pesticides. The intensive use of highly toxic pesticides resulted in significant neurobehavioral effects on farmers, and pesticides were detected in groundwater as well as streams that serve irrigation and domestic use downstream. A broad signaling phase with large groups of stakeholders, however, also drew the attention to soil erosion. Many fields revealed the light-colored subsoil on the upper parts of the fields, and the common perception was that water erosion was the main cause. However, research showed that water erosion was not responsible for the erosion of the topsoil due to low rainfall intensities in combination with a high infiltration capacity of the volcanic ash soils. The steep fields cultivated with potatoes require intensive tillage. All tillage and harvest operations transport topsoil material down the slope. On fields tilled with tractors, the situation is even more serious. As slopes are too steep for contour plowing, tractors plow downslope. Rather than water erosion, tillage erosion was in this case the main cause for the observed erosion processes. This example clearly shows not only the importance of stakeholder input but also the impact of scientists in the signaling phase, avoiding misperceptions that could easily have led to irrelevant routine research on water erosion. After signaling and quantifying the key sustainability factors, that is, human health and environmental impacts of pesticide use and tillage erosion, the question that remained was how to intervene. Clearly, the reduction of pesticide use and pesticide handling issues were the two elements to focus on to reduce the environmental and health impacts of pesticides. To reduce the impact of tillage erosion, very few technical solutions were available. The potato crop requires intensive tillage, and any tillage practice on these steep slopes automatically results in soil erosion. The only practice that significantly reduces tillage erosion is based on the little-known pre-Columbian limited tillage/cover potato system: Wachu rozado that is still being practiced in Northern Ecuador. In this system, seed potatoes are placed on top of the pasture, and the grass mat is folded over the potatoes.

34.8.4.2 Method Selection and Results

A 2 year dynamic farm survey provided insight into the management decisions of the farmers in the Carchi region. This resulted in an economic simulation model (TOA) allowing for the ex ante evaluation of alternative policy and management scenarios (Crissman et al., 1998; Stoorvogel et al., 2004a, 2004b). The TOA methodology addresses two key elements: first, it provides an organizational structure around which to design successful interdisciplinary research that assesses the sustainability of production systems; second, it provides a successful means to communicate research findings to policy makers and the public. Farmers are the Andes' most numerous and most important soil resource managers. Agricultural technology ranges from traditional, extensive, low-input, low-output systems to modern, intensive, high-input, high-output systems. The traditional systems have to be maintained within their ecological constraints and, as a result, are generally perceived as environmentally friendly and sustainable. However, due in part to shrinking farm size, traditional systems have proven to be economically and socially nonsustainable. With a closed agricultural frontier in most parts of the Andes, the fundamental option for Andean farmers is to increase the physical and financial output from the existing farm. This inexorable pressure provides a strong incentive to shift to the higher output modern systems. The basic quest of agricultural and environmental research for sustainable farming systems is to match the environmental friendliness of traditional farming systems while reaching the higher outputs and, thus, the economic and social sustainability found in modern farming systems. Figure 34.8 shows an example of the type of answer one can expect from the TOA system. The trade-off curves between net returns and carbofuran (one of the most commonly applied highly toxic insecticides) leaching are
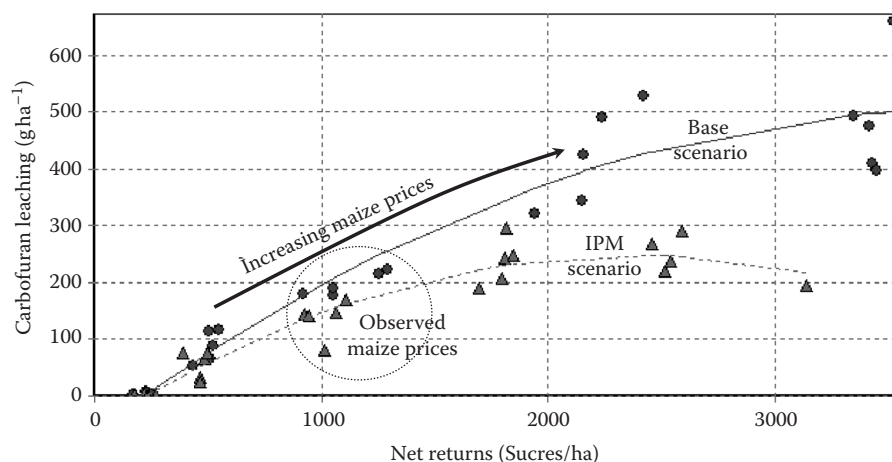

FIGURE 34.8 Trade-offs between net returns and carbofuran leaching under two different management scenarios for the potato-pasture system in Northern Ecuador.

constructed by varying the potato prices. In the base scenario, the current management system is evaluated. The trade-off curve shows that with increasing potato prices, the net returns of the systems increase coinciding with an increase in carbofuran leaching. The latter is not only due to an increase in the potato area but also due to a more intensive management of the potatoes. Two alternative scenarios have been evaluated: the effect of tillage erosion and the effect of the introduction and adoption of Integrated Pest Management (IPM). Due to tillage erosion, we see that carbofuran leaching increases. This can be explained by the removal of the organic matter-rich topsoil that fixes carbofuran. IPM shows an opposite effect where the trade-off curve is moving down. Less carbofuran is being applied, and as a result, leaching is going down. However, the reduction of IPM comes at a cost. Alternative management practices are required to control the pest requiring farm labor. As a result, net returns are slightly decreasing.

34.8.4.3 Impact and Lessons Learned

The TOA provided information on possible interventions both at the political level and at the farm level. In various follow-up projects, the policy cycle has been closed. While pesticides have not been eliminated from the Carchi communities, they are now generally used more cautiously. There is also momentum at the policy level for reducing pesticide dependence. In 1999, all stakeholders were brought together to discuss pesticides and health. This meeting resulted in the Carchi declaration demanding the elimination of the highly toxic products, the inclusion of IPM in university level agriculture training, and a wider dissemination of information on the effects of pesticides. In addition, farmer field schools have been set up in which farmers and the research community developed not only IPM technologies but also pesticide handling measures to reduce the impact of pesticides. The effects of various management changes, as part of IPM, on these farmer field schools are striking. The number of pesticide applications was reduced from 12 in conventional plots to 7 in plots with IPM. Even more important, the overall amount of pesticides applied dropped dramatically. The amount of fungicides decreased by 50%, while insecticide quantities dropped between 40% and 75%. The Carchi story is illustrative for a combined effort in which farm surveys, advanced simulation modeling, geographical information system (GIS) techniques, IPM research, stakeholder meetings, and farmer field schools led to a strong reduction of pesticide use in the Carchi study area. It illustrates the strength of the research chain rather than a single method and/or project. The project was successful in designing innovative production systems for potatoes that were environmentally friendly while protecting the health of farmers. Political decisions were made about environmental and health regulations, and they were implemented, so far only at the regional level. Educational programs for farmers were initiated during this research projects and are continued up to this day. This study not only "assessed the performance of land" but also added and integrated crucial economic and social components, which were part of the original definitions of land evaluation but were hardly incorporated in conventional land evaluation studies focusing on land suitabilities and limitations.

34.8.5 Environmental Quality in the Northern Frisian Woodlands, the Netherlands

34.8.5.1 Problem Definition and Associated Soil Functions

Environmental losses from dairy farming systems in the Netherlands are among the highest in Europe. A strong focus on increasing dry matter production though increases in fertilizer N inputs resulted in average N losses of more than 300 kg N ha^{-1} in the beginning of the 1990s. The excessive loading of groundwater and surface water in the past has stimulated the development of environmental policies at European and national level. Especially, the European Nitrates Directive (EC, 1991) and its translation through the Dutch MINAS legislation have had a profound impact on management practices at farm and field level.

The Nitrates Directive has adopted a threshold for the nitrate concentration in the upper groundwater that corresponds with the E.U. drinking water directive: 50 mg L^{-1}. In 1998, the Netherlands introduced the MINAS policy, a budgeting tool for N and P and farm level with a maximum N surplus of 180 kg N ha^{-1} to be reached in 2003 for sandy soils. Dairy farmers in the Friesian Woodlands in the North of the Netherlands were also confronted with high N surplus levels in the 1990s. Provincial groundwater monitoring networks reported nitrate concentrations for this region well above the 50 mg L^{-1} threshold (Sonneveld and Bouma, 2003). A regional nutrient project was initiated by Wageningen University in close collaboration with local farmers to develop and disseminate knowledge on possibilities for reducing N losses and hence increasing N efficiencies at farm level. A need was expressed in 2004 to provide regional policy makers and farmers with information on the quality status of groundwater in the Friesian Woodlands.

A number of land functions (Section 34.5.2) are especially relevant for the question raised: (1) biomass production; (2) transformations in the soil; and (6) acting as a carbon pool, which is specifically linked to land use on the dairy farms. The functions, physical and cultural environment (4) and geological and archeological heritage (7) became relevant in 2004. In that year, the area was designated as a National Landscape, because of the small-scale elongated fields in the area bordered by hedges and a unique high concentration of pingo-remnants from the Weichsel glacial period.

34.8.5.2 Method Selection and Results

The NFWs cover 60,000 ha and occur in the northern part of the Netherlands. Land use consists predominantly of grassland (80%) with some silage maize (5%). The area mostly consists of hydromorphic podzols developed in Pleistocene coversands. The regional nutrient project (Verhoeven et al., 2003) that aimed at lowering N surpluses adopted a strong participatory approach with study groups among farmers and

contributions from scientists from different disciplines (animal science, soil science, and grassland science). Farm management practices and N surpluses were monitored and used for feedback within the study groups. Different methods were adopted to assess nitrate concentrations in the upper groundwater and N and P concentrations in surface waters in the NFW region:

1. The available STONE model was used to calculate N and P losses at regional level. This quantitative model is based on process description and can be described as a K5 model. The STONE model (Wolf et al., 2003) is used to calculate nutrient loads from the soil system to the surface water and groundwater. STONE calculates nutrient fluxes for different forms of land use like grassland, arable land, and nature and nitrate concentrations in the upper groundwater. The region was divided into 398, 250×250 m grid cells for which specific hydrological, soil physical, and soil chemical characteristics could be derived by downscaling existing soil survey maps. All cells are hydrologically separated but are connected to various drainage levels, like ditches, canals, and subsoil. This implies that STONE behaves spatially like a semi-3D model. All cells have a specific soil type and corresponding soil hydraulic characteristics for the unsaturated zone. The P status and mineralization capacity of the soil are also input for the model. In the STONE model, the water balance model SWAP (Belmans et al., 1983; Van Dam et al., 1997) is used to simulate water fluxes, and the ANIMO model (Groenendijk and Kroes, 1999) is used to simulate nutrient dynamics and nutrient transport. In the initial STONE simulations, the amount of fertilization on the 250×250 m grid was calculated based on "downscaled data 2000" of the Dutch national database on manure application and the assumption that mineral fertilizer was used according to the Dutch national fertilization recommendation. The available data represent the situation in the year 2000. In the second STONE simulation, regional data (2004) on manure application and fertilizer use were used ("regional data"), and grid cells of 25×25 m were used to be able to apply the regional data of manure application on the scale of the fields. Data available for farms were allocated to the small grids. Meteorological data on a daily basis are used as input.

2. A regression method was used based on a modeling study for sandy regions in the Netherlands (Roelsma et al., 2003) where nitrate concentrations had been related to autumn mineral N contents in soil (K3 level) was also applied for a selection of 29 farms on sandy soils (Sonneveld et al., 2010a). In general, agrienvironmental indicators that can be used for monitoring by stakeholders are useful in providing learning feedbacks. The following model was applied:

$$NO_3^-[\text{year} + 1] = C + 0.764N_m[\text{year}] - 37.9P \quad (34.3)$$

where

NO$_3^-$ is the nitrate concentration in the upper groundwater in spring

year is a particular year

C is a constant depending on hydrology and soil type

N_m is soil mineral nitrogen (kg ha^{-1}) in autumn for 0–90 cm

P is a dummy for the occurrence of a peat layer (1 = yes and 0 = no)

For application of this regression model, a monitoring campaign for soil nitrate contents was performed in the autumn of 2006. This method was introduced to see whether soil mineral N contents could be used as an easy indicator by farmers for assessing nitrate concentrations in the upper groundwater.

3. Finally, to validate the above mentioned K3 model, an intensive monitoring campaign was performed in 2007 on more than 300 locations to measure nitrate concentrations in the upper groundwater. In itself, this can be described as a K4 approach.

N losses at farm level were considerably reduced from 1997 to 2003. In this period, the average N surplus dropped from 327 to 168 kg N ha^{-1} (Verhoeven et al., 2003). The reductions were largely achieved through reduced inputs of fertilizer N inputs and improved efficiencies of manure N. For the years 2000 and 2004, STONE calculated relatively low average nitrate concentrations in the upper groundwater in the area with a mean value for the region of 22 mg NO$_3$ L^{-1} for the "downscaled national 2000 data" and 15 mg NO$_3$ L^{-1} for the "local 2004 data" (Figure 34.9). The estimates are far below the E.U. limit of 50 mg NO$_3$ L^{-1}. However, there is a large variation in nitrate concentrations for the year 2000 ranging from higher values in the drier sandy soils to lower values in the peat soils, which is related to differences in denitrification rates. The results for the year 2004 indicate overall low values.

The regression model predicted average nitrate concentrations in the upper groundwater for 2007 to be around 82 mg L^{-1}, which was much higher than the predictions for 2004 by the STONE model. Predicted 90 percentiles of nitrate concentrations in the upper groundwater using the K3 regression method were >150 mg L^{-1} for farmers on sandy soils. Based on the validation data (the K4 method), however, the estimated 90 percentiles were <50 mg L^{-1}. Several observations with high concentrations were especially found on field with arable cultivation (silage maize). Thus, Sonneveld et al. (2010a) concluded that the regression model based on national data was invalid for this particular region and lead to significantly different conclusions. The field measurements appeared to correspond with the STONE model calculations for the year 2004.

34.8.5.3 Impact and Lessons Learned

The results of the participatory approach used in this nutrient management project gave a strong impetus to involve farmers in searching for possibilities for reducing N and P losses to the environment. The strongly reduced N losses in the Northern sandy
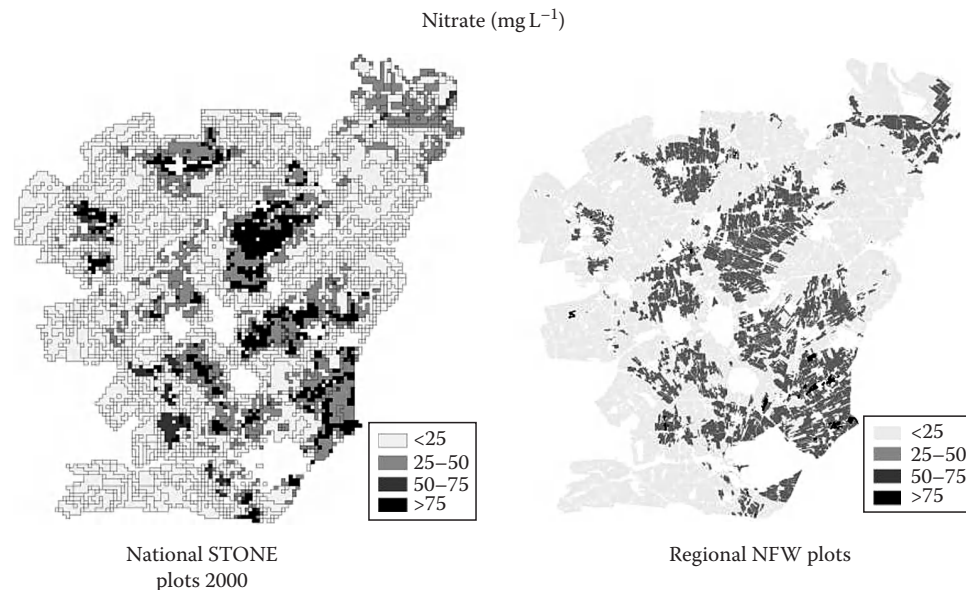
Nitrate (mg L^{-1})

National STONE
plots 2000

Regional NFW plots

FIGURE 34.9 Mean nitrate concentrations in upper groundwater calculated for the years 2000 (a) and 2004 (b).

regions have also been documented in national inventories. The annual environmental report reported that currently, high nitrate concentrations are mainly a problem of southern sandy regions (>80 mg L^{-1}) and not of the northern regions. The STONE model simulations yielded especially relevant results when regional data on land use were introduced in the 2004 model simulations. This strongly suggests that regional calibrations remain a vital component of introducing and applying K5 models at regional scales. The K3 regression models clearly overpredicted nitrate concentrations substantially. It was also found that actual measurements on nitrate concentrations were found to be most convincing for farmers and regional policy makers. Several scale levels were involved in reaching the environmental objectives. Policy formulation at E.U. level resulted in national legislation (MINAS), which was translated to individual farms through a regional nutrient management project. Although national legislation has changed in past years, the insights that were gained have not been lost, and surpluses have continued to decrease over the years. The questions raised clearly went beyond classical land evaluation that is strongly focused on biomass production and follows a top-down strategy. Here, questions did not so much center around options for main types of land use as grassland is considered to be the most important component of dairy farming systems in the region. However, specific management strategies as to how to manage and fertilize pastures (LUTs) were being developed. Options for changing land management practices especially evolved from close collaborations between scientists and farmers and resulted in decreased environmental losses. Model simulations at the regional level as well as field measurements at local level confirmed that the reduced N surpluses resulted in nitrate concentrations <50 mg L^{-1}. Discussions have risen whether the small proportion of silage maize at some dairy farms cannot be replaced with grassland as the higher nitrate concentrations were largely found under fields with maize cultivation.

34.8.6 The World Food Crisis: Assessment of Food Security

34.8.6.1 Problem Definition and Associated Soil Functions

The world population is expected to double by the year 2030. Enough food is produced now, even though its distribution is poor, leaving 800 million people with inadequate food supplies. Many studies have been made on the potential of the Earth to feed its people. We will refer here to the still highly relevant work by Penning de Vries et al. (1995) because they used a modern land evaluation procedure and made some interesting decisions on procedures that correspond to the global level of detail. This study tried to answer questions about global food security by considering three population growth scenarios and two types of agriculture, one with high external inputs (HEI) and the other with low external inputs (LEI). Also, three diets were considered to reflect the demand side: a vegetarian, moderate, and affluent diet. The reader is referred to the above publication for more details on the interesting results of the study. Research at this scale can only be exploratory in a broad sense, and many simplifications are involved. Emphasis was placed on exploring agricultural production potentials. Whether or not these are reached is realistically considered to be beyond the scope of the work as it will depend on socioeconomic and political decisions. What is offered are characteristic windows of opportunity and, as such, the study has led to interesting insights into possible future developments. This study at world level focuses exclusively on biomass production (soil function 1) under a series of conditions relating to management, food consumption, and population growth.

34.8.6.2 Method Selection and Results

A K2 approach might appear to be realistic at first sight. However, no expert can oversee the immensely diverse production

conditions prevailing in different areas of the world. The authors chose, therefore, a K3 approach using a very simple model to calculate crop yields based on available radiation, water, and nutrients. Thus, a universal approach was introduced allowing comparisons among all calculated values. They assigned all possible crops to a cereal (wheat in moderate climates and rice in the tropics) or grass equivalent to obtain a so-called grain equivalent (GE), which makes comparisons relatively easy. The NASA database was used for 1° × 1° grid cells, which occupy a 110 km × 110 km area at the equator. Confining attention here to soils for each grid cell, the authors selected values for slope, soil phase, and soil texture. Selections were based on the 1:5 million FAO World Soil Map and other available data. Clearly, choices were highly arbitrary. Cells often contain different land units, each representing major soil associations with many soil types. Soil texture of the dominant association was applied to the entire cell. Dominant soils in each cell were considered to be well drained, homogeneous without layers or cracks, to be 60 cm deep and without runoff. Degradation of soils was ignored, and so were current land use patterns. For the entire world, 15,500 land units had to be considered along with 700 climatic zones. The model was run for all cells, divided into 15 major regions, which were also distinguished for U.N. population studies, and for different scenarios representing population increase and management types. Estimates for grid cells could have been compared in some areas with estimates obtained using more reliable basic data derived from large-scale soil maps. This, however, has not been done. Therefore, the accuracy of the resulting data cannot be determined, which illustrates the broad exploratory character of the exercise. Still, the alternative is to have nothing at all, and we prefer such a rough estimate, which can be improved upon, over a lack of any estimate.

GIS was used to present maps for the different scenarios, which was effective in communicating the major results of the study. For example, a problem was perceived for Southeast Asia, where food shortages are likely in future even at low rates of population increase and with high-input agriculture.

34.8.6.3 Impact and Lessons Learned

This exploratory study has functioned well to define some major worldwide trends in food security, some of them rather surprising. Food production at the global level can increase significantly, but a strong spatial variability exists. When HEI farming is practiced, all regions can produce the required food, except some areas in Asia. Europe, the Americas, and Central Africa are well-off. Depending on the level of consumption selected, Europe, for example, can grow its food on 30%–60% of its suitable soils and in the America's, this is 20%. When choosing the LEI approach, South Asia will have a food shortage even at the minimum food demand. The impact of studies such as these is difficult to measure. Reports of U.N. agencies have, for example, recently taken a more specific approach to defining food security by focusing on particular regions, and this is supported by the type of studies discussed here. Also, showing the implications of HEI versus LEI management systems in different regions serves

to focus ideological discussions dealing with the most desirable type of farming by presenting plausible consequences of following either HEI or LEI approaches. The study adds an important element to the concept of land evaluation, by including the demand side of the issue of food security. "Assessment of land performance" is not only determined here by the productive capacity as such, which is divided into the LUTs HEI and LEI, but also as a function of demand, which depends on the type of food consumption and on population growth. Thus, land evaluation is placed in a relevant societal context, while the input of soil data is rudimental, to say the least. The tendency of soil scientists is to implicitly assume that the more detailed soil data are provided, the better the land evaluation. The study shows that this is not necessarily true at world level. Unfortunately, soil scientists have not followed up this study by showing what might be gained by improving the soil input data for the model. Finally, one lesson learned is the necessity to repeatedly present studies to a large audience as this work has not received the attention it deserves. The work is not finished when the research report is finished (see also remarks about implementation in the policy cycle in Section 34.5.4).

34.9 New Thrusts for Soil Survey and Land Evaluation

34.9.1 Using New Technologies in the Proper Context

New technological developments are likely to change soil studies in the years to come (see NRC, 2009). Low-cost GPS allow accurate positioning of observers and equipment anywhere in the world. New satellites will strongly increase the opportunity to observe the surface of the Earth in more detail than ever before using an increased number of diagnostic features. Remote sensing from satellites or airplanes will allow better evaluation of crop conditions allowing improved yield predictions. In addition, on-the-go yield monitoring is becoming well established and will revolutionize the assessment of the production capacity of the land by providing a continuous record for many years. In turn, measured differences in yield will give rise to specific research on the underlying causes, which may be many.

Technological developments not only apply to measurements above the soil surface but also to measurements below the soil surface. Many new sensors are being developed to be used for the continuous registration of soil water and solute contents. Already, time domain reflectometry is widely used to measure soil water contents, replacing neutron probes. Transducer tensiometers allow accurate, instant registration of pressure heads in the soil within the range of interest for plant growth. Sensors for N contents in soil are being developed in the context of PA. Information technology allows not only rapid registration but also transmission to central computers. Indeed, some study sites resemble patients in intensive care.

Opportunities to interpret soil data have dramatically increased as well. A wide range of simulation models is available

to predict soil water contents and solute fluxes, even in heterogeneous soils that swell and shrink and may show hydrophobic behavior. This must be calibrated and validated to fit the needs of specific areas where land use questions are raised. Also, expert systems have been perfected allowing a more efficient application of the vast body of knowledge residing with users of the land. GISs, finally, allow integration of different types of data, coupling of databases with models, and construction of digital terrain models with unique opportunities to visualize landscape processes. Interactive use of GISs will enhance effective contacts with end users and policy makers. How will this new technology be used to improve our land evaluation practices? The above case studies already indicated some approaches, but this is only a start. A major challenge for the future is to use this modern technology in a meaningful way keeping in mind the overall objectives of the particular land evaluation being considered. There is a clear risk of a technology push where application of fancy new technologies becomes a purpose in itself. Continued contact with stakeholders and policy makers, as advocated above, is usually quite effective in keeping track of objectives.

34.9.2 Future Land Evaluation Based on Soil Studies

In many, so-called developed countries, standard soil surveys of agricultural lands at scale 1:20,000–1:50,000 have been completed. Revisions of old maps are being made, but funds are often not available to do this. Rather, surveys for specific purposes are increasingly commissioned, or maps are generated using available data that are available in well-accessible databases in many countries. Such data, rather than classical soil maps are then used for a wide variety of applications, using pedotransfer functions and modeling techniques and other means of interpretation. The use of modern GIS software allows rapid generation of flashy reports by nonsoil scientists who are hardly aware of the underlying processes and restrictions to be imposed on data and interpretations being used. Soil scientists should be quite aware of this and make sure that they inject and generate new data as needed to keep their evaluations relevant and up to date. If we only mine existing data, soil science will die (Bouma, 2009). Recent developments in digital soil mapping illustrate innovative state-of-the-art approaches for soil characterization and land evaluation, proving that soil science is still very much alive in contributing to land evaluation in a manner that is highly relevant for society at large now and in the future (Lagacherie and Mc Bratney, 2007; Hartemink et al., 2008).

References

Belmans, C., J.G. Wesseling, and R.A. Feddes. 1983. Simulation model of the water balance of a cropped soil: SWATRE. J. Hydrol. 63:271–286.

Bouma, J. 1993. Soil behavior under field conditions: Differences in perception and their effects on research. Geoderma 60:1–15.

Bouma, J. 1994. Sustainable land use as a future focus of pedology. Soil Sci. Soc. Am. J. 58:645–646.

Bouma, J. 1997a. Precision agriculture: Introduction to the spatial and temporal variability of environmental quality, p. 5–13. *In* J.V. Lake, G.R. Bock, and J.A. Goode (eds.) Precision agriculture: Spatial and temporal variability of environmental quality. John Wiley & Sons, Chichester, U.K.

Bouma, J. 1997b. Role of quantitative approaches in soil science when interacting with stakeholders. Geoderma 78:1–12.

Bouma, J. 2005. Soil scientists in a changing world. Adv. Agron. 88:67–96.

Bouma, J. 2009. Soils are back on the global agenda: Now what? Geoderma 150:224–225.

Bouma, J. 2010. Implications of the knowledge paradox for soil science. Adv. Agron. 106:143–171.

Bouma, J., H.W.G. Booltink, and P.A. Finke. 1996a. Use of soil survey data for modeling solute transport in the vadose zone. J. Environ. Qual. 25:519–526.

Bouma, J., H.W.G. Booltink, P.A. Finke, and A. Stein. 1996b. Reliability of soil data and risk assessment of data applications, p. 63–81. *In* W.D. Nettleton, A.G. Hornsby, R.B. Brown, and T.L. Coleman (eds.) Data reliability and risk assessment in soil interpretations. SSSA Special Publication No. 47. SSSA, Madison, WI.

Bouma, J., J.A. de Vos, M.P.W. Sonneveld, G.B.M. Heuvelink, and J.J. Stoorvogel. 2008. The role of scientists in multiscale land use analysis: Lessons learned from Dutch communities of practice. Adv. Agron. 97:175–238.

Bouma, J., and P. Droogers. 1998. A soil quality indicator for production, considering environment and risk. Geoderma 85:103–110.

Bouma, J., and M.R. Hoosbeek. 1996. The contribution and importance of soil scientists in interdisciplinary studies dealing with land, p. 1–15. *In* R.J. Wagenet and J. Bouma (eds.) The role of soil science in interdisciplinary research. SSSA Special Publication No. 45. SSSA, Madison, WI.

Bouma, J., J.J. Stoorvogel, R. Quiroz, S. Staal, M. Herrero, W. Immerzeel, R.P. Roetter et al. 2007. Ecoregional research for development. Adv. Agron. 93:258–314.

CIBA Foundation. 1997. Precision agriculture: Spatial and temporal variability of environmental quality. CIBA Foundation Symposium 210. Wiley, Chichester, New York.

Commission of the EC. Commission of the European Communities. 2006. Proposal for a directive of the European Parliament and the Council establishing a framework for the protection of soil. COM-232. European Commission, Brussels, Belgium.

Crissman, C.C., J.M. Antle, and S.M. Capalbo. 1998. Economic, environmental, and health tradeoffs in agriculture: Pesticides and the sustainability of Andean potato production. Kluwer Academic Publishers, Boston, MA.

De Vries, W., M. Posch, G.J. Reinds, and J. Kamari. 1992. Critical loads and their exceedance on forest soils in Europe. Winand Staring Center Report No. 58. Wageningen, the Netherlands.

Droogers, P., and J. Bouma. 1996. Effects of ecological and conventional farming on soil structure as expressed by water-limited potato yield in a loamy soil in the Netherlands. Soil Sci. Soc. Am. J. 60:1552–1558.

Droogers, P., and J. Bouma. 1997. Soil survey input in exploratory modeling of sustainable soil management practices. Soil Sci. Soc. Am. J. 61:1704–1710.

Droogers, P., A. Fermont, and J. Bouma. 1996. Effects of ecological soil management on workability and trafficability of a loamy soil in the Netherlands. Geoderma 73:131–145.

EC (European Commission). 1991. Directive 91/676/EEC concerning the protection of water against the pollution caused by nitrates from agricultural sources. European Commission, Brussels, Belgium.

FAO. 1976. A framework for land evaluation. FAO Soils Bulletin No. 32. FAO, Rome, Italy.

FAO. 1983. Guidelines: Land evaluation for rainfed agriculture. FAO Soils Bulletin No. 52. FAO, Rome, Italy.

FAO. 1993. FESLM: An international framework for evaluating sustainable land management. FAO World Research Report No. 73. FAO, Rome, Italy.

FAO. 2007. Land evaluation: Towards a revised framework Land & Water Discussion Paper 6. FAO, Rome, Italy.

Groenendijk, P., and J.G. Kroes. 1999. Modeling the nitrogen and phosphorus leaching to groundwater and surface water with ANIMO 3.5 Report No. 144. DLO Winand Staring Centre, Wageningen, the Netherlands.

Hartemink, A.E. 2008. Soils are back on the global agenda. Guest editorial. Soil Use Manage. 24:327–330.

Hartemink, A.E., and A.B. Mc Bratney. 2008. A soil renaissance. Geoderma 148:123–129.

Hartemink, A.E., A.B. Mc Bratney, and M.L. Mendoca-Santos. 2008. Digital soil mapping with limited data. Springer Science + Business Media, Dordrecht, the Netherlands.

Heuvelink, G.B.M. 1998. Uncertainty analysis in environmental modelling under a change of spatial scale. Nutr. Cycl. Agroecosyst. 50:255–264.

Hoosbeek, M.R., and R. Bryant. 1992. Towards the quantitative modeling of pedogenesis: A review. Geoderma 55:183–210.

Jansen, D.M. and R.A. Schipper (1995). A static descriptive approach to quantify land use systems. Netherlands Journal of Agricultural Science 43:31–46.

Lagacherie, P., and A.B. McBratney. 2007. Spatial soil information systems and spatial soil inference systems: Perspectives, p. 3–24. *In* P. Lagacherie, A.B. McBratney, and M. Voltz (eds.) Digital soil mapping: An introductory perspective. Developments in soil science. Vol. 31, Chapter 1. Elsevier, Amsterdam, the Netherlands.

Lin, H., J. Bouma, Y. Pachepsky, A. Western, J. Thompson, R. van Genuchten, H.J. Vogel, and A. Lilly. 2006. Hydropedology: Synergistic integration of pedology and hydrology. Water Resour. Res. 42:W05301. doi:10.1029/2005WR004085

Lin, H., J. Bouma, L.P. Wilding, J.L. Richardson, M. Kutilek, and D.R. Nielsen. 2005. Advances in hydropedology. Adv. Agron. 85:1–89.

Mausbach, M.J., and L.P. Wilding. 1991. Spatial variability of soils and landforms. SSSA Special Publication No. 28. SSSA, Madison, WI.

NRC (National Research Council). 2009. Frontiers in soil science research. The National Academic Press, Washington, DC.

Penning de Vries, F.W.T., H. van Keulen, and J.C. Luyten. 1995. The role of soil science in estimating global food security in 2040, p. 17–37. *In* R.J. Wagenet and J. Bouma (eds.) The role of soil science in interdisciplinary research. SSSA Special Publication No. 45. SSSA, Madison, WI.

Pulleman, M.M., J. Bouma, E.A. van Essen, and E.W. Meijles. 2000. Soil organic matter content as a function of different land use history. Soil Sci. Soc. Am. J. 64:689–694.

Roelsma, J., C.W. Rougoor, and P.E. Dik. 2003. Regionaal nitraatmonitoringsconcept RENIM; Ontwikkeling en toetsing van een eenvoudige methodiek voor het monitoren van de uitspoeling van nitraat naar het grondwater in zand- en loss gebieden 911. Alterra, Wageningen, the Netherlands.

Rossiter, D.G. 1990. ALES: A framework for land evaluation using a microcomputer. Soil Use Manage. 6:7–20.

Soil Survey Staff. 1998. Keys to soil taxonomy. U.S. Government Printing Office, Washington, DC.

Sonneveld, M.P.W., and J. Bouma. 2003. Methodological considerations for nitrogen policies in the Netherlands including a new role for research. Environ. Sci. Policy 6:501–511.

Sonneveld, M.P.W., J. Bouma, and A. Veldkamp. 2002. Refining soil survey information for a Dutch soil series using land use history. Soil Use Manage. 18:157–163.

Sonneveld, M.P.W., D.J. Brus, and J. Roelsma. 2010a. Nitrate concentrations in the upper groundwater on dairy farms with different N strategies: Validation of regression models for sandy soils. Environ. Pollut. 158:92–97.

Sonneveld, M.P.W., M.J.D. Hack-ten Broeke, C.A. van Diepen, and H.L. Boogaard. 2010b. Thirty years of systematic land evaluation in the Netherlands. Geoderma 156:84–92.

Stoorvogel, J.J., J.M. Antle, and C.C. Crissman. 2004a. Trade-off analysis in the Northern Andes to study the dynamics in agricultural land use. J. Environ. Manage. 72:23–33.

Stoorvogel, J.J., J.M. Antle, C.C. Crissman, and W. Bowen. 2004b. The tradeoff analysis model: Integrated biophysical and economic modeling of agricultural production systems. Agric. Syst. 80:43–66.

Stoorvogel, J.J., J. Bouma, and R.A. Orlich. 2004c. Participatory research for systems analysis: Prototyping for a Costa Rican banana plantation. Agron. J. 96:323–336.

Stoorvogel, J.J., L. Kooistra, and J. Bouma. 1999. Spatial and temporal variation in nematocide leaching, management implications for a Costa Rican banana plantation, p. 281–289. *In* D.L. Corwin, K. Loague, and T.R. Ellsworth (eds.) Assessment of non-point source pollution in the Vadose zone. Geophysical Monograph. Vol. 108. AGU, Washington, DC.

Sys, C., E. van Ranst, and J. Debaveye. 1991. Land evaluation. Part I: Principles in land evaluation and crop production calculations. Part II. Methods in land evaluation. International Training Center for Post Graduate Soil Scientists. University of Gent, Gent, Belgium.

Van Alphen, B.J., H.W.G. Booltink, and J. Bouma. 2001. Combining pedotransfer functions with physical measurements to improve the estimation of soil hydraulic properties. Geoderma 103:133–147.

Van Alphen, B.J., and J.J. Stoorvogel. 2000a. A functional approach to soil characterization in support of precision agriculture. Soil Sci. Soc. Am. J. 64:1706–1713.

Van Alphen, B.J., and J.J. Stoorvogel. 2000b. A methodology for precision nitrogen fertilization in high-input farming systems. Precis. Agric. 2:319–332.

Van Alphen, B.J., and J.J. Stoorvogel. 2002. Effects of soil variability and weather conditions on pesticide leaching: A farm evaluation. J. Environ. Qual. 31:797–805.

Van Dam, J.C., J. Huygen, J.G. Wesseling, R.A. Feddes, P. Kabat, R.E.V. Van Walsum, P. Groenendijk, and C.A. Van Diepen. 1997. Theory of SWAP version 2.0. Simulation of water flow, solute transport and plant growth in the soil–water–atmosphere–plant environment. Technical Document 45. DLO-Winand Staring Centre, Wageningen, the Netherlands.

Van Lanen, H.A.J., M.J.D. Hack ten Broeke, J. Bouma, and W.J.M. de Groot. 1992. A mixed qualitative/quantitative physical land evaluation methodology. Geoderma 55:37–54.

Van Latesteijn, H.C. 1995. Scenarios for land use in Europe: Agro-ecological options within socio-economic boundaries, p. 43–65. *In* J. Bouma, A. Kuyvenhoven, B.A.M. Bouman, J.C. Luyten, and H.G. Zandstra (eds.) Eco-regional approaches for sustainable land use and food production. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Vanclooster, M., P. Viane, J. Diels, and K. Christiaens. 1994. WAVE. A mathematical model for simulating water and agro-chemicals in the soil and vadose environment: Reference and user manual (release 2.0). Institute for Land and Water Management, University of Leuven, Leuven, Belgium.

Verhoeven, F.P.M., J.W. Reijs, and J.D. Van der Ploeg. 2003. Re-balancing soil–plant–animal interactions: Towards reduction of nitrogen losses. Neth. J. Agric. Sci. 51:147–164.

Wolf, J., A.H.W. Beusen, P. Groenendijk, T. Kroon, R.P. Roetter, and H. Van Zeijts. 2003. The integrated modeling system STONE for calculating nutrient emissions from agriculture in the Netherlands. Environ. Model. Softw. 18:597–617.

Young, F.J., R.D. Hammer, and F. Williams. 1997. Estimation of map unit composition from transect data. Soil Sci. Soc. Am. J. 61:854–861.

Hydropedology

Phillip Owens
Purdue University

Henry Lin
Pennsylvania State University

Zamir Libohova
*United States Department
of Agriculture*

35.1 Introduction

Hydropedology is an emerging interdisciplinary science that grew out of the need to address complex natural processes in the earth's critical zone (Wilding and Lin, 2006). Increasingly, scientific research is being conducted by multidisciplinary teams as demanded by the complex, dynamic, and spatial–temporal variability of natural systems that require inputs from various disciplines (Sposito and Reginato, 1992). Hydropedology seeks to bridge disciplines to address (1) knowledge gaps between pedology, soil physics, hydrology, and other related bio- and geosciences; (2) scale differences in microscopic, mesoscopic, and macroscopic studies of soil and water interactions; and (3) data translations from soil survey databases into soil hydraulic properties (Lin, 2003).

Hydropedology is a union of hydrology, pedology, and soil physics disciplines with a focus on soil–water interactions. Hydrology is defined as *the science that treats the waters of the Earth, their occurrence, circulation and distribution, their chemical and physical properties, and their reaction with their environment, including their relation to living things* (NRC, 1991). Pedology is defined as *the branch of soil science that integrates and quantifies the formation, distribution, morphology, and classification of soils as natural or anthropogenically modified entities* (Wilding, 2000; Buol et al., 2001; Lin et al., 2006). Soil physics is defined as *the study of the physical properties of the soil and the relation of the soil physical properties to the study of the state and transport of matter and energy* (Scott, 2000). Although these

definitions are specific and cover the breadth of each discipline, water dynamics, response, and the interactions with the environment are common underlying themes. As a need to unite these disciplines, hydropedology has been defined as *an intertwined branch of soil science and hydrology that encompasses multiscale basic and applied research of interactive pedologic and hydrologic processes and their properties in the variably-unsaturated zone* (Lin, 2003). More specifically, hydropedology focuses on the synergistic integration of pedology and hydrology to enhance the holistic study of soil–water interactions and landscape–soil–hydrology relationships across space and time. Its aim is to understand pedologic controls on hydrologic processes and properties and hydrologic impacts on soil formation, variability, and functions (Figure 35.1; Lin et al., 2008b). Even though hydropedology has its foundation in pedology, soil physics, and hydrology, it is also linked to other bio- and geosciences such as geomorphology, geology, geography, hydrogeology, hydroclimatology, ecohydrology, biology, and other branches of natural sciences (Figure 35.2). In a broader sense, hydropedology seeks to identify feedback mechanisms that allow for a holistic approach to the study and prediction of ecosystem functions.

This chapter provides an overview of the fundamentals and applications of hydropedology, including a review of some guiding principles and several recent advances in soil architecture and preferential flow, soil hydromorphology, scaling, digital soil mapping (DSM), pedotransfer functions (PTFs), and coupling biogeochemistry with hydropedology.
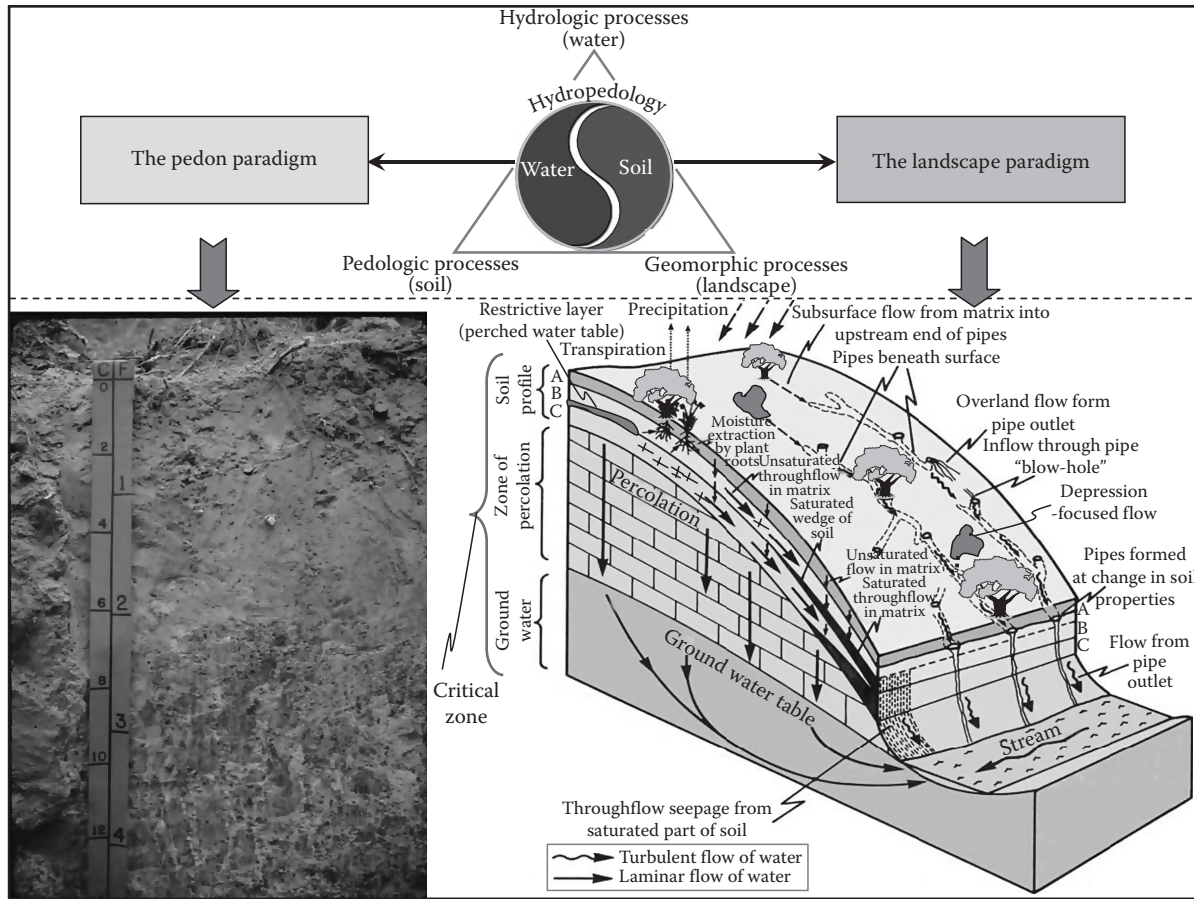
FIGURE 35.1 Hydropedology connects the pedon and landscape paradigms through linking phenomena occurring at the microscopic (e.g., pores and aggregates) to mesoscopic (e.g., pedons and catenae), macroscopic (e.g., watersheds and regional), and megascopic (e.g., continental and global) scales. (From Lin, H.S. 2010. Earth's critical zone and hydropedology: Concepts, characteristics, and advances. Hydrol. Earth Syst. Sci. 14:25–45; and Atkinson, T.C. 1978. Techniques for measuring subsurface flow on hillslopes. *In* M.J. Kirkby (ed.) Hillslope hydrology. John Wiley & Sons, Chichester, U.K.)
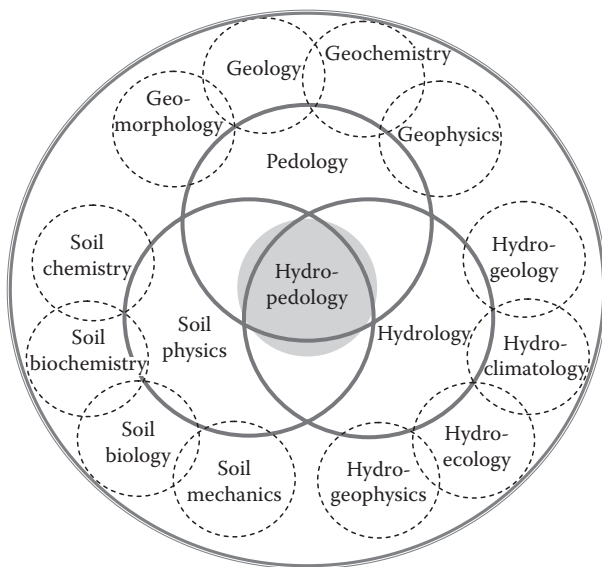

FIGURE 35.2 A conceptual diagram illustrating the relationship of hydropedology to other related disciplines. (From Lin, H.S. 2003. Hydropedology: Bridging disciplines, scales, and data. Vadose Zone J. 2:1–11.)

35.2 Fundamental Questions and Basic Characteristics of Hydropedology

Two fundamental questions of hydropedology are as follows (Lin et al., 2008b, 2008c):

1. How do soil architecture and the distribution of soils over the landscape exert a first-order control on hydrologic processes (and associated biogeochemical and ecological dynamics) across spatio-temporal scales?
2. How does water at the landscape scale (and the associated transport of energy, sediment, chemicals, and biomaterials by flowing water) influence soil genesis, evolution, variability, and functions?

Water at the landscape scale encompasses the source, storage, availability, flux, pathway, residence time, and distribution of water in the near-surface terrestrial environment. While source, storage, availability, and flux of water in the soil have been studied extensively in the past, attention to flow pathways (especially flow networks), residence time (age of water), and spatiotemporal pattern of flow dynamics (and its underlying

organizing principle) has been much limited (Lin et al., 2006; McDonnell et al., 2007). Two basic characteristics of hydropedology are linked to the above two questions:

First, hydropedology emphasizes in situ soils in the landscape, where distinct pedogenic features (e.g., aggregation, horizonation, and redox features) and soil–landscape relationships (e.g., catena, soil distribution patterns, and soil map units) are essential in understanding interactive pedologic and hydrologic processes. Developing quantitative relationships between complex natural soil architecture and soil hydrologic functions across scales is an important research area of hydropedology. Three related key aspects are as follows:

- Hydropedology calls for a new era of soils research that is based on soil architecture (broadly defined as the entirety of how the soil is structured) that is beyond soil texture alone, so that the prediction of flow (and reaction) pathways, patterns, and residence times can be made realistically. This requires innovative techniques for improved quantification of soil architecture at different scales, especially in situ, noninvasively, and linking such soil architectural parameters to field-measured soil hydraulic properties.
- Hydropedology considers the soil in the real world as a "living" entity in the landscape rather than "dead" material that is manipulative mechanically. As Kubiena (1938) pointed out, a crushed or pulverized sample of the soil is related to the soil formed by nature like a pile of debris is to a demolished building. The full benefit of understanding this living system lies in studying them close to their natural setting in the real-world setting.
- Hydropedology attempts to link the form and function of soil systems across scales (Lin et al., 2006), rather than mapping soils without considering soil functions or modeling soils without incorporating soil architecture and soil–landscape patterns. Jenny (1941) clearly noted at the end of his famous book *Factors of Soil Formation*,

The goal of soil geographer is the assemblage of soil knowledge in the form of a map. In contrast, the goal of the "functionalist" is the assemblage of soil knowledge in the form of a curve or an equation… Clearly, it is the union of the geographic and the functional method that provides the most effective means of pedological research.

- Such a union of soil maps and soil functions is what hydropedology hopes to promote in quantitative ways.

Second, hydropedology deals with the variably unsaturated zone in the terrestrial near-surface environment, including the shallow root zone, deeper vadose zone, temporally saturated soil zone, capillary fringe associated with groundwater table, wetlands, and subaqueous soils (soils formed in sediment found in shallow permanently flooded environments such as in an estuary; Demas and Rabenhorst, 2001). Three related key aspects are as follows:

- Hydrology has the potential to be an integrating factor for quantifying soil formation and evolution and for understanding soil changes upon global climate and land use changes (Lin et al., 2005). Hence, a focus on water can provide a potentially powerful means of quantifying and predicting dynamic soil functions.
- New ways of characterizing and mapping soils could or should be linked to hydrology, such as wetland boundary delineations, riparian zones, and hydropedologic functional units that are considered as soil–landscape units with similar pedologic and hydrologic functions (Lin et al., 2008b).
- The interpretation and quantification of soils as historical records of environmental changes in the past could be significantly improved if hydrologic data are considered simultaneously. This has been demonstrated in some paleosols and paleohydrology studies (Ashley and Driese, 2000).

35.3 Fundamental Scientific Issues of Hydropedology

At this stage of its development, the fundamental issues of hydropedology may be considered under the following four headings (Lin et al., 2005; Lin, 2010):

1. *Soil structure and horizonation in relation to in situ water flow and chemical transport*: Hydropedology emphasizes quantitative soil architecture of field soils across scales and their links to preferential flow in different spatial and temporal dimensions (see Section 35.4.1).
2. *Soil morphology and pedogenesis in relation to soil hydrology and soil change*: Hydropedology utilizes quantitative soil hydromorphology as a signature of soil hydrology and also uses soils as valuable records of environmental change over time (see Section 35.5).
3. *Soil catena and distribution pattern in relation to water movement over the landscape*: Hydropedology focuses on quantitative relationships between field soils and their surrounding landscape and the impacts of such relationships on hydrologic (and related biogeochemical/ecological) processes (see Section 35.6).
4. *Soil functions and maps in relation to carriers of soil quality and soil–landscape heterogeneity*: Hydropedology promotes quantitative delineations of functional soil units in the landscape as well as precision soil–landscape mapping for diverse applications (see Section 35.9).

35.4 Applications of Hydropedology

35.4.1 Soil Architecture and Preferential Flow

The natural soil "architecture" is of essence in understanding soil physical, chemical, and biological processes as well as landscape and ecosystem dynamics. It has been suggested that most that can be learned from sieved and repacked soil samples

has already been done, so whatever we do with soil hydrology should be done on intact, undisturbed soils, and preferably in situ. A new era of soils research will have to rely on "structure-focused," passing the stage of "texture-focused," to achieve better ways of quantifying flow pathways, residence times, and spatiotemporal patterns of landscape water.

Soil architecture is used here broadly to mean the entirety of how the soil is structured, which encompasses at least three parts: (1) solid components, including soil matrix (represented by soil texture and soil microfabric) and soil aggregation (represented by the type, quantity, and size distribution of peds and aggregate stability); (2) pore space, including the size distribution, connectivity, tortuosity, density, and morphology of various pores; and (3) interfaces between solid components and the pore space, such as coatings on peds or pores, and the macropore–matrix, soil–root, microbe–aggregate, and horizons interfaces.

Since soil structure generally refers to a specific soil horizon, the broader term of soil architecture used here also encompasses the overall organization of a soil profile (e.g., horizonation), a soil's relationship with the landscape (e.g., catena), and the overall hierarchical levels of soil structural complexity (Lohse and Dietrich, 2005; Lin et al., 2005).

Soil horizonation or layering is ubiquitous in nature, so it must be adequately addressed when measuring, modeling, and interpreting hydrologic processes in watersheds. Various kinds and thicknesses of soil horizons and how they organize in soil profiles reflect longtime pedogenesis and the past and current landscape processes (Mausbach and Wilding, 1991). The fact that natural soils are layered has at least two significant implications for hydrology: (1) interface between soil layers of contrasting textures and/or structures would slow downward water movement, which often leads to some kind of preferential flow (e.g., fingering flow, macropore flow, or funnel flow); and (2) soil layering or discontinuity in soil hydraulic properties between layers would promote lateral flow or perched water table, especially in sloping landscapes with a water-restricting layer underneath.

A catena (also called toposequence) is a chain of related soil profiles along a hillslope, with about the same age, similar parent material, and similar climatic condition, but differs primarily in relief that leads to differences in drainage and soil thickness. Catenary soil development often occurs in response to the way water runs down the hillslope and recognizes the interrelationship between soil and geomorphic processes (Hall and Olson, 1991; Moore et al., 1993; Thompson et al., 1997; Lin et al., 2005). Catenae are often called hydrosequences of related soils, especially in depositional landscapes (Schoeneberger and Wysocki, 2005). Another important aspect of soil architecture along the hillslope is related to preferential flow network, which is further discussed later in this section.

While the importance of soil architecture across scales has long been recognized in soil science and hydrology, its quantification and incorporation into models have been notoriously lagged behind. This problem is due to many factors, including (1) inconsistent and fragmented concepts of soil structure;

(2) overemphasis on ground-sieved soil materials and soil texture in the past, thus ignoring or downplaying the importance of the soil's "natural architecture." That natural soils are structured to various degrees at different scales is the rule, whereas the existence of a macroscopic homogeneity is the exception (Vogel and Roth, 2003); (3) lack of a comprehensive theory of soil structure/architecture formation, evolution, quantification, and modeling that can bridge orders of magnitude in scale and integrate physical, chemical, biological, and anthropogenic impacts. Although a hierarchical organization of soil aggregates has been well recognized (Tisdall and Oades, 1982; Vogel and Roth, 2003) and fractal characterization of soil structure has been proposed (Bartoli et al., 1998; Perrier et al., 1999), there is still a lack of means of representing field soil structure at different scales in a manner that can be coupled into models of flow, transport, and rate processes (Lin, 2003; Lin et al., 2005); and (4) lack of appropriate techniques and devices to quantify soil architecture directly, especially in situ noninvasively. Traditionally, soil structure has been evaluated by pedologists in the field using morphological descriptions or thin section observations, while soil physicists have employed wet and dry sieving, elutriation, and sedimentation to conduct aggregate analysis. In the absence of direct quantification, soil structure has been frequently evaluated by methods that correlate it to the properties or processes of interest (such as water retention, saturated hydraulic conductivity, infiltration rate, and gas diffusion rate). In recent years, noninvasive methods that permit soils to be investigated without undue disturbance of their natural architecture have become increasingly attractive, allowing 3D visualization of internal soil structure and its interactions with water. These methods include x-ray computing tomography, soft x-ray, nuclear magnetic resonance, gamma-ray tomography, ground-penetrating radar, and other methods (e.g., Anderson and Hopmans, 1994; Perret et al., 1999; Luo et al., 2008). Image analysis has brought new opportunities for analyzing soil structure, especially that of pores, their sizes, shapes, connectivity, and tortuosity (Vogel et al., 2002; Vervoort and Cattle, 2003). However, although numerous attempts have been made to find either statistical relations or deterministic links between soil structural data and hydraulic properties, a significant gap remains between in situ soil structure/architecture and field-measured soil hydraulic properties at different scales.

Quantifying soil architecture in the field across scales and at desirable spatial and temporal resolutions has been technologically limited (Tillotson and Nielsen, 1984). While landforms (e.g., digital elevation models or DEMs) and vegetation (e.g., land use/land cover) can now be mapped with high resolution (e.g., using LiDAR and IKONOS, respectively), there is a "bottleneck" phenomenon for in situ high resolution (e.g., submeter to cm) and spatially temporally continuous and noninvasive mapping or imaging of subsurface architecture including flow networks. This "technological bottleneck" has constrained our predictive capacity of many soil and hydrologic functions.

There is also a *conceptual bottleneck* that needs to be resolved for developing a new generation of hydrologic models. That is, should a continuous field or discrete objects be used to model surface and subsurface flow? Traditionally, hydrologic processes are generally conceptualized within the field domain (e.g., the Navier–Stokes equation and the Darcy's law) (Goodchild, 2007). Classical hydrology has applied findings from fluid mechanics, together with the necessary constitutive relations to develop sets of governing equations (much the same as atmospheric and ocean sciences have done). However, heterogeneities in land surface, hierarchical structures of soils, channel geometries, and preferential flow networks all make the land surface and subsurface different from the continuous field assumption (Kung, 1990; Noguchi et al., 1999) (CUAHSI, 2007; Lin, 2010). It is becoming more and more recognized that solid earth is not a continuous fluid; rather, it poses hierarchical heterogeneities with discrete flow networks embedded in both the surface and the subsurface. As McDonnell et al. (2007), Kirchner (2006), Beven (2002), and many others have noted, current models in watershed hydrology are based on well-known small-scale physics or theories such as the Darcy's law and the Richards equation built into coupled mass balance equations. It has been observed that the dominant process governing unsaturated flow in soils may change from matrix flow to preferential flow under different conditions when moving from the pore scale to the pedon scale (Blöschl and Sivapalan, 1995; Hendrickx and Flury, 2001). When moving from the pedon scale to the landscape/watershed scale, our knowledge for extrapolating the Darcy–Buckingham's law and the Richards equation to large heterogeneous area is further constrained (Weiler and Naef, 2003).

Because of heterogeneous soil architecture, variability in energy and mass inputs to soils (Addiscott, 1995; Warrick, 1998) diversity in biological activities, and nonlinear dynamics of hydrologic processes, preferential flow can occur in practically all natural soils and landscapes (Lin, 2010). As Clothier et al. (2008) summarized, preferential flow can occur spatially at the pore scale of spatial order 10^{-3} m, at the core scale (10^{-1} m), in pedons (10^0 m), down hillslopes (10^1–10^3 m) (Lehmann et al., 2007), through catchments (10^4–10^5 m), and across large regions of ≥10^6 m. Time-wise, preferential flow can operate during fluid flows at the temporal order of 10^0–10^1 s, during hydrological events 10^0–10^2 h, throughout seasonal changes 10^0 year, and across interannual variations of 10^1 years.

Based on three theoretical considerations and numerous published evidence, Lin (2010) has attempted to justify the likely universality of preferential flow in natural soils, inferring that the potential for preferential flow occurrence is everywhere in nature, although the actual occurrence of preferential flow depends on local conditions (Lin and Zhou, 2008). Lin (2010) also showed that networks are abundant in soils, such as root branching networks, mycorrhizal mycelial networks, animal borrowing networks, crack and fissure networks, artificial subsurface drainage networks, and pore networks between soil particles and aggregates. These networks provide preferential flow conduits, which in return reinforces or modifies the existing networks.

35.5 Soil Hydromorphology and Quantification

35.5.1 Water Tables and Inferred Soil Hydrology

Much of the research conducted in hydromorphology has focused on identifying the depth and duration of water tables to understand soil hydrology and the effects on pedogenesis and land use. Water tables are described as the top of the zone of saturation where water fills all or most of the pores in the soil to create saturated or satiated conditions (Daniels and Buol, 1992). The knowledge of soil hydrology has many practical applications for engineering, plant growth, reclamation, and remediation. The presence of free water in soils influences many related properties such as oxygen concentration (decreased aerobic respiration), reduction and transformation of redox couples (i.e., denitrification, metal reduction, and carbon dioxide production), change in solubility of compounds (bioremediation), change in soil pH, and decreased soil strength. Soils with high water tables generally lead to some challenges for land use such as decreased landscape stability (landslides), poor trafficability, limitations for onsite wastewater disposal, increased compaction, and poor fertility status (denitrification) in agriculture fields (Veneman et al., 1998). However, in more recent years, the benefits of wetlands and hydric soils have been recognized for their contribution to renovating contaminated waters and providing a source for groundwater recharge (Rabenhorst et al., 1998). Horizons are considered saturated when the soil–water pressure is zero or positive (Soil Survey Staff, 1999). The depth of the water table is related to multiple factors: (1) landscape position; (2) precipitation; (3) evapotranspiration; and (4) permeability of the surface and subsurface (Vepraskas, 1995). Understanding and interpreting the presence of water tables from landscape and soils information is important for understanding pedogenesis and land use. Soil morphology and redoximorphic features (formerly known as mottles) are used to imply the presence of water tables and inferred oxic–anoxic conditions.

### 35.5.2 Process of Redoximorphic Feature Formation

The presence of a water table results in a distinct morphology within a pedon and is the result of reduction–oxidation processes. The redox process results in the redistribution of primarily iron in natural soils, causing the soil to appear mottled. Although manganese is sometimes present and is an indication of alternating redox conditions, manganese is not as ubiquitous as iron. Additionally, the relationship between water tables and manganese is not as well correlated to water table durations and periodicity (Khan and Fenton, 1996; Fielder and Sommer, 2004). The iron segregations in areas within a soil are recorded as iron depletions and concentrations (Vepraskas, 1995; Vepraskas and Faulkner, 2001). The processes of reduction is driven by microorganisms; therefore, the soil conditions required to cause reduction

are heterotrophic facultative aerobic and anaerobic microorganisms, carbon source, saturated conditions (to decrease diffusion of O_2), and a temperature above biologic zero.

Microorganisms use carbon sources to obtain energy for life. For a simplistic example, Equation 35.1 represents the use of glucose and Equation 35.2 represents oxygen as final electron acceptor (Ponnamperuma, 1972).

$$C_6H_{12}O_6 \text{ ------->} 2CH_3\text{--}CO\text{--}COO^- + 6H^+ + 4e^-$$
$$\longmapsto \quad ATP \tag{35.1}$$
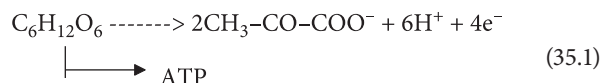
In aerobic soils, O_2 acts as the final electron acceptor as in Equation 35.2.

$$1.5O_2 + 6H^+ + 4e^- \rightarrow 2H_2O \tag{35.2}$$
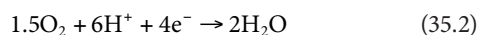
As soils become saturated, the microorganisms continue to use oxygen until it is depleted. Once the oxygen is depleted, there are several other molecules that act as a final electron acceptor for the microorganisms. They are NO_3^-, Mn oxides, Fe oxides, sulfates, and CO_2. Given the temperature is above biologic zero, saturated conditions and a carbon source, the microorganisms through the production of electrons decrease the Eh, which creates conditions where reduced phase of the aforementioned compounds is thermodynamically stable.

The difference in solubility of iron and manganese provides the mechanisms for redistribution and formation of redoximorphic features. Iron is stable and insoluble in aerobic environments as Fe^{3+} valence state. Once iron is reduced from Fe^{3+} to Fe^{2+}, it is soluble and mobile and can be redistributed in the soil matrix. The solubility product constant for $Fe(OH)_3$ is 6.3 × 10^{-38}, whereas for $Fe(OH)_2$, it is 7.9 × 10^{-15} (Kotz et al., 1994). The reduced ferrous iron will remain as Fe^{2+} until it encounters enough higher redox potential O_2 to oxidize, commonly when the water table begins to drop allowing O_2 to enter the system and increase the Eh. The formation of iron oxide minerals often occurs in pore channels, on structure surfaces, and within the soil matrix (Vepraskas, 1995; Figure 35.3). Soils with a dominant gray matrix color are depleted of iron and display the color of minerals with minimal or no iron oxide coatings.
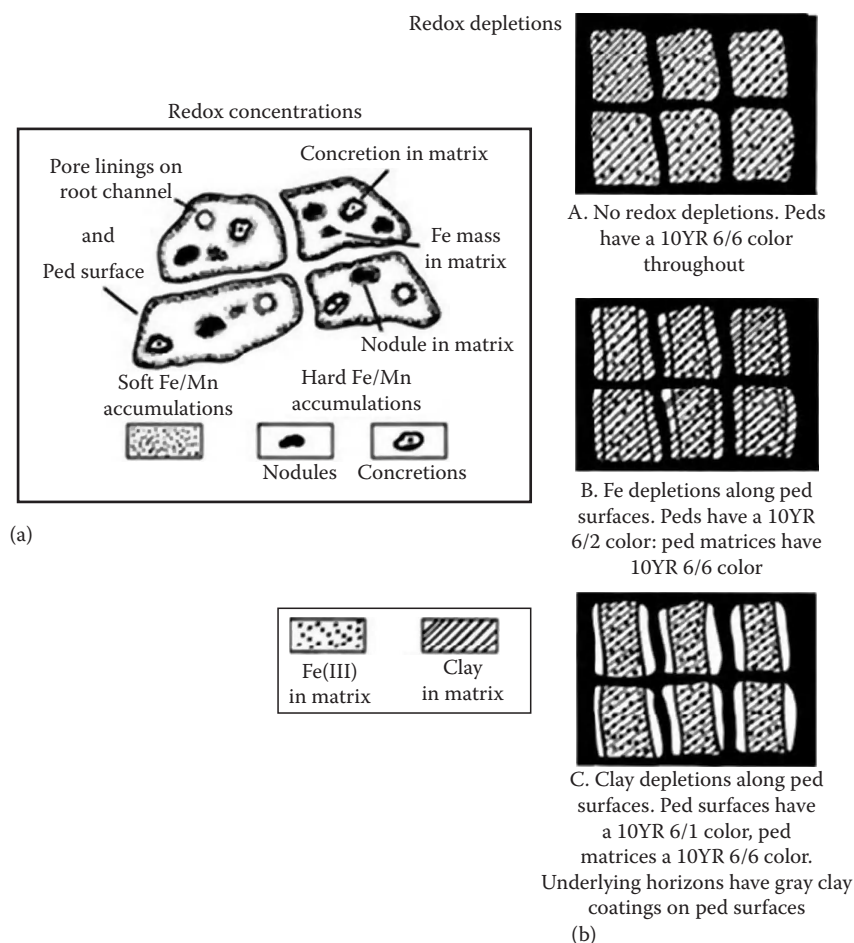
FIGURE 35.3 Schematic illustration showing different kinds of redox concentrations and their relationship to soil macropores and matrices (a). Schematic illustrations of redox depletions showing changes in color and texture as ped surfaces develop Fe depletions and clay depletions (b). (From Vepraskas, M.J. 1995. Redoximorphic features for identifying aquic conditions. Technical Bulletin No. 301. North Carolina State University, Raleigh, NC.)

35.5.3 Relationship of Redoximorphic Features and Soil Hydrology

Data on soil morphology and soil–water relationships are important because collecting water table information with wells and peizometers is time-consuming and expensive. There have been many studies relating soil morphology to the presence or absence of free water in soils (Daniels et al., 1971; Simonson and Boersma, 1972; Veneman et al., 1976; Franzmeier et al., 1983; Vepraskas and Wilding, 1983a, 1983b; Zobeck and Richie, 1984; Griffin et al., 1992; Jenkinson and Franzmeier, 2002). Researchers have focused on correlations between water table levels and morphological soil characteristics. One of the diagnostic redox features commonly used to define limitations due to the duration of water tables is the depth to chroma ≤2 depletions. Depth to gray mottles has been characterized extensively in many soils (Latshaw and Thompson, 1968; Simonson and Boersma, 1972). Yakovleva (1980) found a good general relation between depth to water table and degree of gleying in soils, and Daniels et al. (1971) found that the gray mottles were associated with different degrees of saturation in different landscape positions. More recent investigations by Evans and Franzmeier (1988) and Jenkinson et al. (2002) in Indiana and Megonigal et al. (1993) in South Carolina correlated length of soil saturation with color in alfisol soil toposequences derived from glacial till and a floodplain-terrace soil toposequence developed from tertiary bedrock, respectively. Thompson and Bell (1996) in Minnesota, observed a relationship between the surface accumulation of organic matter in mollisols and duration of saturation. Several researchers (Veneman et al., 1976; Vepraskas and Bouma, 1976; Richardson and Lietzke, 1983) have found that, in general, very brief periods of saturation are associated with high chroma peds with manganese and iron concentrations. Intermittent periods of saturation are associated with bright Fe concentrations, ped ferrans, and some Mn cutans. Saturation persisting for several months in a year is associated with <2 chroma matrix interiors, ped surface depletions, pore and root channel depletions, and virtually no Mn cutans or nodules. Since soil color patterns are strongly influenced by the chemistry and mineralogy of iron and manganese compounds, their solubility status determines the color of the soil matrix and argillans and determines the type and distribution of iron concentrations in a horizon.

Quantification of the time of saturation related to redox features has been reported in a few studies. In Indiana, Franzmeier et al. (1983) found the following soil morphology–soil–water relationships: (1) gleyed horizons with dominant colors of chroma ≤2 are saturated most of the year, (2) horizons with a dominantly brown matrix with gray depletions are saturated a few months of the year, and (3) horizons with a chroma of 5–6 that lack depletions ≤3 are seldom saturated during the growing season unless they were near a gleyed horizon. In study within the coastal plains of Georgia, Jacobs et al. (2002) found that soil horizons with chroma <2 were saturated >50% of the time. Horizons with low chroma iron depletions can be saturated as little as 18% of the time. In three representative floodplains in Maryland

and Delaware, Vaughan et al. (2009) found that with lower temperatures, longer durations of saturation were needed for iron reduction with respect to ferrihydrite (FH). For temperatures greater than 90°C, less than 2 days were required for iron reduction with respect to FH; and as many as 20 days for iron reduction with soil temperatures between 1°C and 3.9°C (Vaughan et al., 2009).

However, these observations and measurements have been mostly limited to a point pedon scale and other research has shown that there are various degrees of expressions of redoximorphic features described that relate to movement and residency of water as controlled by landscape (Richardson et al., 1992; Jenkinson, 1998).

35.5.4 Soil Hydrology within Landscapes

Landscape hydrology has been the focus of research since the 1960s culminating with an intensive group of studies titled the Wet Soil Monitoring Project (WSMP). The WSMP was funded and coordinated by USDA–NRCS between 1990 and 2001. The studies were national in scope and included most climatic temperature and moisture regimes. This group of studies found that the geomorphic surface and the stratigraphy of soil horizons control water table depths in the soil, the amount of time the soil stays saturated, and the direction (vertical and or horizontal) water flows into or across the landscape. In turn, landscape hydrology governs the range of soil moisture regimes, oxidation–reduction processes, and soil color patterns (Arndt and Richardson, 1988; Griffin et al., 1992; Richardson et al., 1992; Hopkins, 1996; Thompson and Bell, 1996; Jenkinson, 1998, 2002; Feigum, 2000; Owens, 2001).

The presence and movement of groundwater through the landscape influences many of the physical and chemical processes that are involved in soil development. The water table, duration of saturation, and the flow of water across and through the landscape are controlled by topography and stratigraphy (Jenkinson, 1998). Richardson and Daniels (1992) concluded that changes in topography exert a strong influence on groundwater flowing through a landscape, and that landscapes can be subdivided into a system of groundwater recharge, flowthrough, and discharge that, under stable long-term conditions, can transfer soluble materials through the groundwater system in distances measured in kilometers. Richardson and Daniels (1992) observed these effects on landscapes in North Dakota and in the southeastern region of the United States. Thompson and Bell (1996) related the hydrological profile of soils along a toposequence in Minnesota to the geomorphology of the landscape. Daniels et al. (1971), in characterizing North Carolina landscapes, found a statistical correlation between water table depth and closeness to the dissected edge. These data indicated that shoulder positions within landforms are well drained due to lateral water movement and decreased residency time.

Additionally, the knowledge accumulated by the WSMP studies created a platform for the interpretation of hydrosequences, especially of soil drainage conditions, that are critically important when making land use decisions. The WSMP studies, gave

researchers insights into what types of instrumentation worked well and what measurement processes could benefit from new technologies. Traditionally, water tables and anaerobic conditions induced by poor drainage conditions have been evaluated by wells, peizometers, tensiometers, open boreholes, alpha, alpha-dipyridyl, and platinum electrodes for Eh measurement. However, the Eh values obtained from platinum electrodes were generally considered circumstantial and occasionally suspect (Owens et al., 2008; Rabenhorst et al., 2009). Recent research by Rabenhorst et al. (2009) indicated that more accurate Eh values can be determined by using voltmeters with high input resistance (20 GΩ) and data loggers with standard configuration that maintains an open circuit during the instantaneous measurement. The need for a new method to address the issue of saturation and anaerobiosis together led to the development of a new device called indicator of reduction in soils (IRIS). IRIS was developed at Purdue University (Jenkinson, 2002) at the end of the WSMP period and further refined by Rabenhorst et al. (2008). IRIS mimics natural soil processes, visually indicates soil reduction, that can be quantified and be robust (Castenson and Rabenhorst, 2006). IRIS is a PVC tube coated with colored Fe minerals, mostly FH, that dissolves under reducing conditions (Jenkinson and Franzmeier, 2006). In practice, IRIS tubes are inserted into a soil, removed after a few weeks or longer, and evaluated by visually inspecting the coating in the field. If the coating was not dissolved, no reduction occurred, but if it was dissolved, reducing conditions must have prevailed (Figure 35.4). Another field technique using zero valent iron rods was developed to infer O_2 concentration (Owens et al., 2008). Iron rods were polished and placed in toposequences and compared with O_2 measurements, piezometer data, and Eh measurements. The coatings on iron metal rods correlated well with specific oxygen ranges. Rods in soils with O_2 concentrations below

about 3 mg L^{-1} did not develop bright (7.5YR 4/4–5/8) oxide/oxyhydroxide coatings but instead formed black (10YR 2/1–2/2) coatings. Rods in soils with O_2 concentrations between about 2% and 5% developed variegated bright (7.5YR 4/4–5/8) oxide/oxyhydroxide coatings indicating microsite differences in O_2 concentrations. Rods in soils with O_2 concentrations above about 5% with adequate moisture were almost completely coated with bright (7.5YR 4/4–58) iron oxide/oxyhydroxides. This method provides a simple and inexpensive means to qualitatively estimate the ranges of O_2 status in soils.

Many soil hydrology studies were conducted at the pedon scale and along catenae. In these studies, soil–water tables were highly correlated to landscape position or topography of the landscape. The term hydrosequence is used to a series of soils with different degrees of wetness due to topography (Schoeneberger and Wysocki, 2005). Soil moisture regime and presence of water tables can be attributed to several factors; however, precipitation and topography are the two most dominant factors at field scale. Other important factors contributing to water table dynamics are the underlying stratigraphy and particle size of the parent material. Water may flow through hydraulically conductive upper horizons and "perch" on horizons with low hydraulic conductivity. In sandy parent materials with high hydraulic conductivity, the presence of a water table may not be evident.

Studies have indicated that within similar climate conditions, topography and parent material stratigraphy are the two most important controlling factors of water tables. In central Iowa, Khan and Fenton (1994) found that within an internally drained glacial till landscape, soils in depressions had longer duration of water tables with higher organic carbon when compared to higher topographic soils. Soils in the depressions also had different chemical properties. In the wettest part on the landscape, Bkg horizons formed where calcareous water was discharged. Similarly Richardson et al. (1992) found that flownet analysis based on stratigraphy and water table position was useful for describing observed differences in soils related to recharge and discharge positions in the landscape. Considering the soils in the context of the landscape, Richardson et al. (1992) identified four kinds of water movement that dominate soil development: (1) recharge or dominantly downward flow of water; (2) flowthrough or lateral groundwater movement; (3) discharge or movement from the water table either to or near the soil surface; and (4) stagnation or slow water movement creating water table mounds. This analysis indicated that leached soils and absence of carbonates were found in the recharge soils. Soils within discharge zones and flowthrough positions had accumulations of carbonates and gypsum reflecting the addition of carbonate-rich water at lower landscape positions from higher landscape positions (Figure 35.5). Rolling topography with steep hillslopes commonly have complex shallow hydrologic system where water moves laterally above a limiting layer (Daniels and Buol, 1992). Evans and Franzmeier (1988) found water flowing laterally on hillslopes remained oxygenated and did not lead to reducing conditions. Therefore, the interpretations of redoximorphic features must be placed in the context of landscapes.
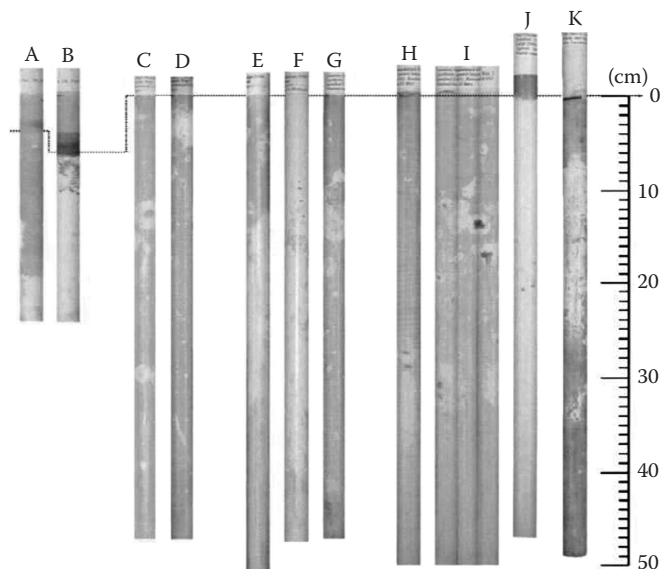

FIGURE 35.4 Photos of IRIS illustrating the degree of iron reduction on the rods. The degree of reduction and stripping of iron relates to the intensity and duration of saturated conditions in the soil.
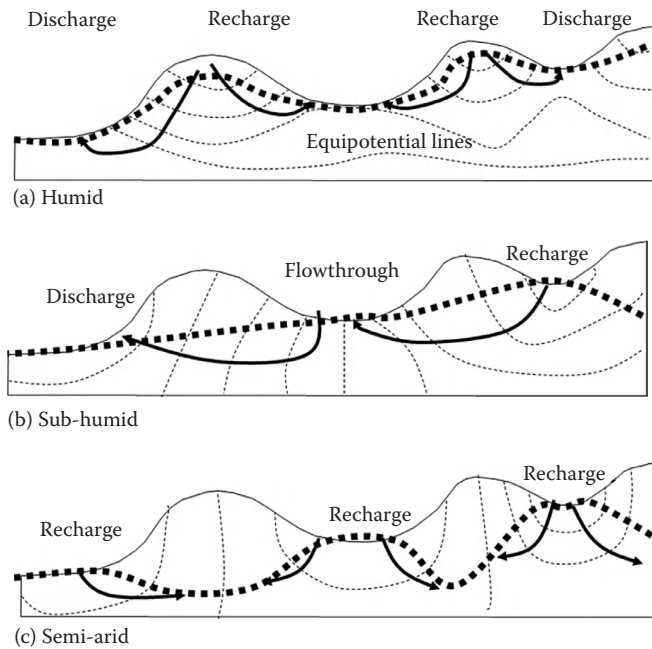
FIGURE 35.5 An examination of soils from (a) Iowa, (b) North Dakota, and (c) Saskatchewan found that the soils in the depressions varied such that the soils from Iowa indicated discharge, North Dakota had recharge in the higher areas and flowthrough and discharge in the lower areas. The Saskatchewan had a dominance of recharge wetlands with argillic horizons. (From Richardson, J.L., L.P. Wilding, and R.B. Daniels. 1992. Recharge and discharge of groundwater in aquic conditions illustrated with flownet analysis. Geoderma 53:65–78.)

Utilizing hydropedology concepts in soil morphologic studies is promising because soil properties that influence water distribution can be better understood and water–soil morphology relationships can be placed in the landscape context. The availability of digital information coupled with the ability to display spatial information will provide the potential to make significant advances for quantifying the spatial–temporal relationship between soil morphology and soil hydrology within larger landscapes. This will be discussed further in Section 35.9.

35.6 Coupling Hydropedology and Biogeochemistry

A biogeochemical cycle is a pathway by which a chemical element moves through biotic (biosphere) and abiotic (lithosphere, pedosphere, atmosphere, and hydrosphere) compartments of the earth system. Biogeochemical cycles are inseparable from the hydrologic cycle (NRC, 1991), and fluxes of elements and rates of biogeochemical cycling are often linked to hydrologic conditions (Richardson et al., 2001). Zones of enhanced fluxes and reaction rates where terrestrial and aquatic ecosystems meet have been recognized, or suspected, for decades (Lohse et al., 2009). *Biogeochemical hot spots* are patches in the landscape that show disproportionately high reaction rates relative to the surrounding matrix, whereas *hot moments* are defined as short periods of time that exhibit disproportionately high reaction rates relative

to longer intervening time periods (McClain et al., 2003). Hot spot and hot moment activity is often enhanced at terrestrial–aquatic interfaces. For example, using examples from the carbon and nitrogen cycles, McClain et al. (2003) showed that hot spots occur where hydrological flow paths converge with substrates or other flow paths containing complementary or missing reactants. Hot moments occur when episodic hydrological flow paths reactivate and/or mobilize accumulated reactants.

Since hydrology often triggers hot spots and hot moments of biogeochemical reactions and ecological functions (Bundt et al., 2001; McClain et al., 2003), improved modeling and prediction of soil architecture and preferential flow will have considerable implications for determining chemical fluxes and calculating elemental budgets in soils and ecosystems. Interpretation of point measurements without knowing preferential flow paths is now often questioned (Gottlein and Manderscheid, 1998; Netto et al., 1999), because the uncertainty of whether soil solution is extracted from stagnant or high velocity flow paths makes it practically impossible to reliably determine mass flux rates. Additional complications arise in structured soils for reactive components due to locally variable chemical conditions. In addition, macropore linings and aggregate coatings restrict lateral mass transfer and reduce sorption and retardation; hence, physical and biochemical nonequilibriums are enhanced. All these suggest the need to identify and model preferential flow pathways and networks in real-world soils if we are to improve the modeling and prediction of interactive soil physical, chemical, and biological processes.

At the pedon scale, soil aggregation, macropore networks, horizonation, and the soil–bedrock interface have important impacts on preferential flow and biogeochemical reactions in soil profiles (Deurer et al., 2003; Field et al., 1984; Flury et al., 1994). At ped surfaces, for example, C and N contents and microbial activities are higher than in those in ped interiors (Tisdall, 1995; Alef and Kleiner, 1989; Kavdir and Smucker, 2005). Soil layering can trigger on or off preferential flow between soil horizons with different textures and structures, which tend to promote lateral flow in sloping landscapes. The soil–bedrock interface has also been recognized as important subsurface preferential flow pathway (Li et al., 1997; Freer et al., 1997; McGlynn et al., 2002). Flühler et al. (1996) have suggested three regimes of flow and transport within a soil profile during a preferential flow process: (1) lateral distribution flow in the *attractor zone* where preferential flow is initiated; (2) downward preferential flow in the *transmission zone* where water moves along preferential flow pathways and bypasses a considerable portion of the soil matrix (Quisenberry et al., 1993; Vervoort et al., 1999); and (3) lateral and downward dispersive flow in the *dispersion zone* where preferential flow pathways are interrupted and water flow becomes more or less uniform again.

Based on a review of studies documenting soil microbial biomass distributions with depth (Balkwill and Ghiorse, 1985; van Gestel et al., 1992; Dodds et al., 1996; Murphy et al., 1997; Richter and Markewitz, 2001; Blume et al., 2002; Taylor et al., 2002; Fierer et al., 2003), the system of Flühler et al. (1996) can be applied. Various studies have demonstrated a consistent pattern of changing microbial biomass size and activity distribution,

characterized by three broad zones with distinct biogeochemical potentials that correspond to the three flow zones of Flühler et al. (1996): (1) *attractor zone*, equivalent to the A horizon and representing high-biomass surface soils under the strong influence of plant roots, moisture, and temperature; (2) *transmission zone*, equivalent to B and C horizons and representing lower more spatially heterogeneous biomass than in the *attractor zone* (Konopka and Turco, 1991; Rodríguez-Cruz et al., 2006); and (3) *dispersion zone*, representing higher-biomass, water-capture zones at the soil–bedrock interface or right above a water-restricted soil layer because of moister and more nutrient-rich conditions resulting from impeded or lateral flow (Dupuy and Dreyfus, 1992; Buss et al., 2006). These three zones are distinguished from "deep-subsurface" aquifers and geologic materials, many of which have extremely low biomass (Balkwill et al., 1985; Kieft et al., 1995). Contributions of greenhouse gas fluxes from the *attractor zone* have been studied intensively (Konopka and Turco, 1991; Krasovskaia et al., 2003; Rodríguez-Cruz et al., 2006), whereas contributions from the *transmission and dispersion zones* are essentially uncharacterized and will likely depend on their connectivity to the atmosphere or on preferential gas flux associated with preferential water distribution.

35.7 Scale Dependency of Soils and Modeling

As previously discussed, a lot of work has been conducted on measuring and understanding of preferential flow at point pedon scale and its role on soil development and hydrological process at landscape scale; however, a much broader context especially in terms of varying scales is needed. The spatial and temporal scales have been long recognized by soil science (Jenny, 1941; Soil Survey Division Staff, 1993; Wilding, 2000; Soil Survey Staff, 2003; Sposito, 1998; Tugel et al., 2005). Different schemes of soil classifications have captured soil variability at different scales from field to hillslope, landscape, watershed all the way to regional continental and world scales (Soil Survey Division Staff, 1993; Boul et al., 2003; Soil Survey Staff, 2004). Various criteria and principles have been used to classify soils and their properties across various scales represented by different modes such as state factor model (Jenny, 1941), process model (Simonson, 1959), energy model (Runge, 1973), statistical landscape models (Shovic and Montagne, 1985; Bell, 1990; Zhu et al., 1996, 1997, 2001, 2006; Havens, 1998; Zhu, 2000), and geostatistical models (Odeh et al., 1992, 1994; McBratney et al., 2003). The evolution of soil models along with that of tools and purposes of soil mapping have highlighted some of the issues related to the current soil maps and soil survey in general. The issues identified encompass a wide range related to the type and level of soil information provided and current delivery format of soil data (McBratney et al., 2003; USDA–NRCS–NGDC, 2006; Hempel et al., 2008; MacMillan, 2008).

The majority of the soil information comes from field descriptions of pedons and laboratory measurements of soil samples collected from these pedons and it is of three types: (1) quantitative, (2) qualitative, and (3) semiqualitative/quantitative. The field descriptions include qualitative statements about soil horizons, soil structure, consistency, color, and texture; quantitative statements about soil horizon thickness, redoximorphic features, degree of mottling, and nutrients; and semiqualitative statements about roots and pores among others (Schoeneberger et al., 2002). The need to use this information in hydrologic modeling has led to the development of PTFs as a way to translate qualitative and semiqualitative soil information to quantitative one (Bouma, 1989; Pachepsky et al., 1999; McBratney et al., 2002; Pachepsky and Rawls, 2003, 2004; Wagenet et al., 1991). The value of PTFs has been widely recognized because of their multiple applications in estimating and predicting soil hydraulic properties (McBratney et al., 2000; Lin et al., 2005; Pachepsky et al., 2006) and also for their use in modeling in general (McBratney et al., 2002). McBratney et al. (2002) points out that PTFs are typically limited to specific geomorphic regions or soil types and lack quantifiable uncertainty. Such efforts have led to the emerging of hydropedology as a promise to bridging pedology and hydrology (Lin et al., 2005; Bouma, 2006; Pachepsky et al., 2006, 2008). These advancements have highlighted one of the major challenges related to the use of PTF and hydropedology, that is, the transferability of the qualitative, quantitative, and semiqualitative/quantitative field measured and estimated soil information across multiple scales (Bouma, 2006). In his discussion about different conceptual frameworks and methods that have been developed to address this issue, Bouma (2006) challenges the soil scientists for the fact that soil expertise is not adequately represented in modeling. Bouma (2006) recognizes the fact that pedologists in general are not comfortable with representation of soils in terms of homogeneity and isotropy needed by hydrologists for modeling purposes. Evidence shows that certain measured hydraulic soil properties display both stochastic and deterministic behavior (Harlan and Franzmeier, 1974; King and Franzmeier, 1981; Evans and Franzmeier, 1986; Franzmeier, 1991; Haws et al., 2004; Zeleke and Cheng Si, 2005; Botros et al., 2009). Franzmeier (1991) measured K_{sat} on different soils across Indiana and found similar values by grouping soils based on the parent materials, soil horizon, and soil texture categories indicating the presence of stochastic behavior of K_{sat} within each category and deterministic behavior across categories. Haws et al. (2004) used the representative measurement area (RMA) concept to describe the behavior of measured K_{sat} and found that an RMA greater than $400 \, cm^2$ was needed to filter out smaller-scale heterogeneities and a distance between 120 and 200 m to capture the effective or deterministic behavior of K_{sat} at landscape scale. The conclusion of the study was that in modeling water flow into and through soils, the measured K_{sat} should integrate across the range of variability at the scale of interest. Iqbal et al. (2005) using geostatistical analysis (kriging and structured variograms) of soil physical properties like soil bulk density (ρb), saturated hydraulic conductivity (K_s), and soil–water content (θ) showed that soil texture displayed spatial structure or
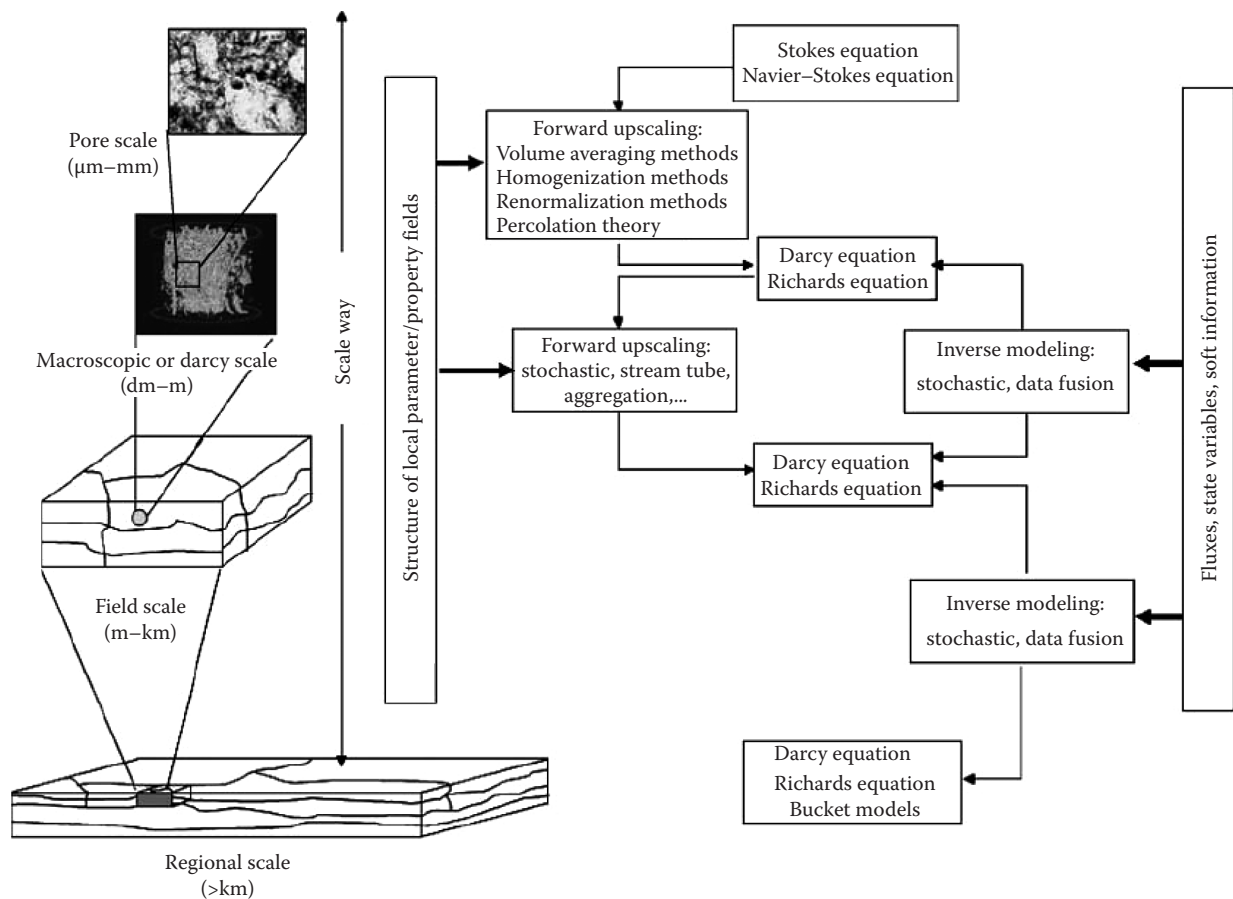
FIGURE 35.6 Forward and inverse upscaling methods to upscale soil–water flow processes from the local to the field scale. (From Vereecken et al., 2007.)

autocorrelation within a 400 m distance while hydraulic properties and bulk density within a 100 m distance beyond which the distribution of measured values displayed random behavior. Similar results have been found by other researchers (Ersahin, 2003; Kozak and Ahuja, 2005; Vereecken et al., 2007) for soil infiltration rates and other soil properties (Iqbal et al., 2005). Vereecken et al. (2007) discusses the "scaleway" approach first introduced by Vogel and Roth (2003) in upscaling soil hydraulic properties across multiple scales (Figure 35.6) and provides a summary of the upscaling methods.

Methods like fractals have also been proposed for upscaling soil hydraulic properties (Shepard, 1993; Peyton et al., 1994; Baveye and Boast, 1999; Rodriguez-Iturbe and Rinaldo, 1997; Perret et al., 2003) along with other approaches. The method is based on the concept of self-similarity of geometrical shapes displayed by disordered systems that appear similar over a range of length scales. Such structures are self-similar and are referred to as geometrical fractals (Perret et al., 2003). The characteristic of a geometrical fractal is its fractal dimension, D, which can be 1, 2, and 3 for 1D, 2D, and 3D geometric shapes. The dimensions can be determined by measuring the length, area, or volume depending on the geometric shape. The method has been used to quantify pore volumes and pore tortuous paths and relate them to soil hydraulic conductivity (Shepard, 1993). The fractal

methodology approach can also be useful for understanding and upscaling soil properties using pattern recognition.

The majority of the discussions about soil spatial variability has been focused on soil properties especially hydraulic properties (Grayson and Bloschl, 2000; Ersahin, 2003; Kozak and Ahuja, 2005; Corwin et al., 2006; Vereecken et al., 2007). However, the majority of soil properties are represented spatially by soil series and soil map units. In this context, two levels of information could be distinguished to characterize the representation of spatial distribution of soil property data, that is, within and between soil map units. Research conducted on deriving certain soil hydraulic properties from soil series and soil map units (Ersahin, 2003; Kozak and Ahuja, 2005) has highlighted some issues related to the spatial representation of these properties. The majority of soil property data has been measured or estimated at point or pedon scale and transferred to polypedon, soil series, and soil map unit scales (Soil Survey Division Staff, 1993). The spatial distribution of these transferred properties comes in the form of representative values (Soil Survey Staff, 2004) or mean values with ranges with no indication about their spatial distribution within soil map units (Libohova et al., 2010). Furthermore, often ranges between soil series and soil map units overlap rendering them not statistically significant. In addition, soil map units are not always pure with only one representative

soil as they could be complexes and/or associations with smaller inclusions of other soil types (Soil Survey Division Staff, 1993). Taxonomy and management related interpretations have both played a role in the current design of soil map units. Least but not last, political boundaries and time differences between soil survey conductance has introduced further variability that may not necessarily reflect the reality. How much of this variability is real and how much is overrepresented and could be simplified is a question that relates directly to the representation of spatial variability of soil properties. Efforts to address these issues have led to the development of most recent concepts like "functional homogeneity" versus "structural heterogeneity" of soil series and soil map units for parsimonious watershed modeling and predictions (Basu et al., 2009). The presence of overlapping ranges of soil properties between soil map units constitutes at the same time recognition of their continuum spatial variability from the soil scientists and surveyors. It is a reflection of the purpose of soil surveys and the limitations of the tools used at the time of soil survey campaigns. Recent developments in spatial tools and technologies offer an opportunity to significantly improve the current delivery format of spatial soil information (McBratney et al., 2003). U.S. Soil Survey Program has adapted the major land resource area (MLRA) concept (Indorante et al., 1996) and started the so-called edge-matching in an effort to create seamless soil maps for the conterminous United States. European Union has launched similar efforts despite the surmountable obstacles due to differences in soil classification systems, time, and purpose of the soil survey conductance and resistance to share data deemed often a matter of national security for some countries (European Communities, 2007). The most recent global soil map initiative aimed at creating the first world digital soil map (McMillan, 2009) has taken these efforts to another level and created an opportunity for the soil surveys around the world to come together like never before.

A plethora of methods have been developed to aid the efforts for creating spatially continuous soil maps and especially soil property maps from the existing soil information. McBratney et al. (2003) provide an extensive review of methods developed for DSM by establishing quantitative relationships between soil properties or classes and their "environment" often referred to as DEM-derived terrain attributes. Some of the methods discussed include generalized linear models, classification and regression trees (CART), neural networks, fuzzy systems, and geostatistics (McBratney et al., 2003). These and other methods have been necessary to address the existence of different levels and types of soil information and in some cases the lack of soil information. The major drives for these developments have been the great progress in geographic information systems (GIS), GPS, remote sensing, and data sources.

DEMs coupled with increasing demands for soil data and information from environmental sciences and modeling (McBratney et al., 2003). Soil science is now faced with the challenge of looking beyond its own domain and sharing paths more than in the past with other disciplines such as ecology and hydrology, just to name few (Lin et al., 2005; Wilding and Lin,

2006). This may require soil scientists to become familiar with the concepts and terminology used by other disciplines in order to establish mutually beneficial communication. A closer look at hydrological and/or ecological sciences reveals similarities with soil science for engaging in modeling different aspects of natural process being soil, water, vegetation, and wildlife (Corwin and Wagenet, 1996; Laurenson, 1974; Ibanez et al., 1995; Sayfried and Wilcox, 1995; Phillips, 2001; Austin, 2002; Burke and Lauenroth, 2002; Epstein et al., 2002; Kie et al., 2002; Shirazi et al., 2003; Hopmans and Schoups, 2005; Lu et al., 2005; Sivapalan, 2005; Logsdon et al., 2007; McDonnell et al., 2007; Wagener et al., 2007; Simmons et al., 2008). A revisit of different concepts used by these disciplines in the light of spatial variability representation is very important as "a better understanding of spatial soil variability, its development over time (pedogenesis) and its functional relationships to recent processes in soil landscapes is one of the biggest challenges in soil science" (Sommer, 2006).

In soil science, the pedon definition and its spatial limits are used to separate soils within a small-scale pattern of local variability, usually $1-10 \, m^2$, while polypedon is presented as a unit of classification or a relatively homogenous soil body whose limits are extremely hard to find (Soil Survey Division Staff, 1993). Because of this difficulty, the polypedon concept has been ignored by most soil scientists due to its circulatory nature, and, thus, its properties and extent have been transferred to the pedon. The concept has been further expanded spatially by the use of soil series and soil map units that are viewed often as a collection of polypedons (Soil Survey Division Staff, 1993). The practical applications of these concepts have been difficult in representing the soil as true continuum. This has resulted in soils and soil map units with overlapping ranges for most of the soil properties (Lin et al., 2005), thus introducing more spatial variability than could be present. Using the analogy from Logsdon et al. (2007), the spatial variability within soil map units would be considered as random fluctuations or noise. However, the spatial variability between soil map units may or may not be random depending on the soil property and the predicted function (Libohova et al., 2010). The distinction between these two spatial variability categories is scale dependent (Logsdon et al., 2007). Hydrology and hydrologic modeling in particular utilizes similar concepts but uses terminology such as "deterministic length scale" (Sayfried and Wilcox, 1995), representative elementary area (REA) (Wood et al., 1988; Wood, 1999) closely related to the representative elementary volume (REV) (Peck, 1983), and RMA (Haws et al., 2004) used in soil science in order to emphasize scale dependence of variability. Other terms like "stochastic," "deterministic," "homogenous," and "heterogeneous" are used in the same manner in describing hydrologic processes; and the distinction among them is scale dependent (Haws et al., 2004). Sayfried and Wilcox (1995) acknowledge that spatial variability is "rarely entirely deterministic or stochastic" as noted also by other researchers (Laurenson, 1974; Vogel et al., 2000). Both trends are found in every hydrologic parameter, and the dominance of one or the other has been found to be scale dependent. In this context, the authors describe the definition of "deterministic length

scale," meaning that depending on the scale, spatial variability of any parameter can be characterized as deterministic, thus the variability can be represented by a single value or function rather than a range of values with some probability distribution as in the stochastic case. Ecological sciences use terms like "numerical richness" (Kempton, 1979) and "species density" (Hurlbert, 1971). Ibanez et al. (1995) relates these terms to the ratio of the number of different soil taxa over the total number of pedons described and the number of soil taxa per sampling area, respectively. Burke and Lauenroth (2002) discuss the challenges related to representation of the regional scale variability facing the ecological sciences especially the limitations of local-scale and short-term studies in capturing regional scales. The common feature of discussions about the spatial distribution of soils or soil properties, hydrological process, or ecological species diversity is the fact that they are all bound within a certain spatial domain and the nature of their variability is scale dependent. The focus so far has been on spatial variability, but the same can be stated for temporal variability (Tugel et al., 2005; Sommer, 2006; Logsdon et al., 2008) especially when considering the human impact as Basu et al. (2009) showed on the role of tile drains on controlling hydrologic response in the Midwest. The difference in scale representation of variability can be addressed, and soil structure and hydrologic functioning at different scales can be related within the same spatial units as pointed by Pachepsky et al. (2008).

The continuous soil maps and soil property maps most likely will be in a raster format; thus, the selection of the appropriate pixel resolution is also important. The selection of pixel size is also scale related and depends on the soils and soil properties to be inferred. Soil properties manifest varying degrees of spatial dependence (McBratney and Webster, 1983) as do the terrain attributes (Gessler et al., 1995), and they need to be taken into consideration; otherwise the derived terrain attributes may have no meaning and the soil pattern or process could be lost as Moore et al. (1991, 1994) have found. McBratney et al. (2000) provides a review of resolutions and scales used for soil survey and suggests pixel sizes of >2 km, 20 m^{-2} km, and <20 m for national to global, catchment to landscape, and local extents, respectively. In a more detailed overview, McBratney et al. (2003) provide pixel resolution based on the order and scale of the soil survey and identify the 5–20 m pixel resolution as the suggested one for U.S. soil survey order 1 and 2 corresponding to scales 1:5,000–1:20,000. Winzeler et al. (2008) found that a pixel size resolution between 5 and 10 m in relatively flat glaciated landscapes in Northern Indiana can capture the local relief and the spatial distribution of selected soil properties without losing valuable information on patterns and variability at the landscape scale, avoiding at the same time unnecessary microrelief details often described as "noise."

The challenge for soil science is not only to accurately and completely represent the spatial variability of soil properties but also to relate this variability to scales. For hydrologic purposes, this means that soil morphological differences need to be viewed also from the hydrologic processes perspective, thus leading to the generation of soil functional hydrologic property maps with a focus on commonality rather than morphological differences.

The approach parallels that of the catchment hydrology major challenge described by Sivapalan (2005) and McDonnell et al. (2007) as a shift from prescribing patterns of catchment heterogeneity to focusing on the search for (1) geomorphic or landform processes that generated these patterns; (2) processes that underlined this heterogeneity and complexity; and (3) the ecological, pedological, and geomorphological functions of these processes. Just as the focus on pattern, process, and function will revolutionize hydrology (Sivapalan, 2005), the same might be stated for soil science in general and soil modeling in particular. This means that mapping soils and soil map units, soil morphological and taxonomic differences as well as soil heterogeneity, need to be linked with scales, hydrological, and soil-forming processes in order to look at the soil map units from the functional in addition to compositional perspective. The variability is everywhere, and before any attempts to characterize it quantitatively, we should ask questions about "purpose" and "scale" or "boundary conditions" as there may not be a universal solution or a silver bullet to address all the variability due to the fact that the nature of variability (stochastic, deterministic, chaotic, or homogenous) is scale dependent as are the hydrological process in hydrology and other disciplines from ecology all the way to social sciences. Recent developments in DSM (McBratney et al., 2003), coupled with PTFs and the emergence of hydropedology (Pachepsky et al., 2006), provide a platform for addressing the scale dependency of soil variability related to modeling.

35.8 Pedotransfer Functions

"Data rich, information poor" (information here connotes interpretation, synthesis, and utilization of data) has been a common syndrome in many disciplines (Lin, 2003). This problem is largely due to data fragmentation, incompleteness, incompatibility, inaccessibility, or lack of interpretation and synthesis in spite of past extensive and costly data collections. In soil science and hydrology, it is recognized that gaps exist between what is available (e.g., the National Cooperative Soil Survey [NCSS] Program databases) and what is needed (e.g., soil hydraulic parameters and PTFs needed for simulation models). Improved procedures to extract useful information from the available databases and to improve and interpret soil survey data for flow and transport characteristics in different soils are needed.

With the increasing popularity of coupling GIS with vadose zone models and soil survey databases for diverse natural resource applications, the demand for soil hydraulic information has increased significantly in recent years. However, the lack of sufficient field data on soil hydraulic properties often limits the application of contaminant transport and hydrological modeling. Existing methods for direct field measurement of soil hydraulic properties remain complex, time-consuming, and costly (Mualem, 1986; Bouma, 1989), despite decades of work by soil physicists, hydrologists, and others representing different disciplines. Another limitation of direct field measurement is significant spatial and temporal variability, hence demanding a large number of measurements that are often prohibitive in terms of

time and money (van Genuchten et al., 1999b). This has prompted efforts to indirectly estimate soil hydraulic properties using more readily available data often found in soil surveys (such as particle-size distribution, bulk density, organic matter content, and others). Such indirect methods, now often referred to as PTFs as suggested by Bouma and van Lanen (1987, 1990), have been attempted for estimating water retention curve parameters, saturated hydraulic conductivity, the unsaturated hydraulic conductivity function, and other soil hydraulic parameters (Vereecken et al., 1990; Batjes, 1996; Lin et al., 1999b; Tietje and Hennings, 1996; van Genuchten et al., 1999a; Wösten et al., 2001; Shaw et al., 2000). Compared to other methods of estimating soil hydraulic parameters (e.g., pore-size distribution models and inverse methods), PTFs are inexpensive, easy to derive and use, and, in many practical cases, they provide good estimators for missing hydraulic parameters (Verhagen and Bouma, 1998; van Genuchten et al., 1999b; Wösten et al., 2001). Besides conventional regression or functional analyses, new techniques such as neural networks (Schaap et al., 2001), group methods of data handling (Pachepsky and Rawls, 1999), and CART (McKenzie and Jacquier, 1997) are increasingly being explored for developing PTFs using a growing number of large soil databases such as the NCSS databases, UNSODA (Leij et al., 1996), HYPRES (Wosten et al., 1999), WISE (Batjes, 1996), SoilVision (SoilVision Systems Ltd., 2002), and many others.

While various degrees of success have been achieved with different PTFs (Pachepsky et al., 1999; Wösten et al., 2001), limitations of existing PTFs remain. For example, the vast majority of existing PTFs are completely empirical, and limited efforts have been put into systematic probing of underlying mechanisms for

the existence of such functions. There is a tendency among soil scientists and others to estimate soil hydraulic functions by regression analysis in a given region, but efforts to apply PTFs derived from one area to soil hydrologic studies in other soil regions are futile (Kutílek and Nielsen, 1994). Only a few quasi-physical PTFs exist, such as those by Arya and Paris (1981), Haverkamp and Parlange (1986), and Arya et al. (1999), in which particle-size distribution is first translated into an equivalent pore-size distribution model and then further related to water retention curve. However, the flow system in the bundle of capillary tubes differs considerably from the water flow network in real-world soils (Miller and Miller, 1956; Beven and Germann, 1982). Thus, the practical application to field soils of these quasi-physical PTFs, based essentially on the Hagen–Poiseuille's law, is very limited. In addition, existing PTFs have not fully incorporated soil structure and land use information and have lacked scale and temporal considerations. As such, the accuracy, reliability, and utility of existing PTFs are constrained. In the mean time, the NCSS databases developed over the past century have been underused in addressing environmental issues. Interpretations and applications of the NCSS databases are challenges facing soil scientists in general and pedologists in particular. There are pressures on both pedologists and soil physicists/hydrologists to disseminate soil survey information and utilize it in a variety of applications (Simunek et al., 2003).

The combined efforts of pedologists, soil physicists, and hydrologists could open up new opportunities for the next generation of PTFs. For instance, five general categories of PTFs may be identified for potential improvement in estimating dynamic soil properties (Figure 35.7). PTF type I relates use-dependent
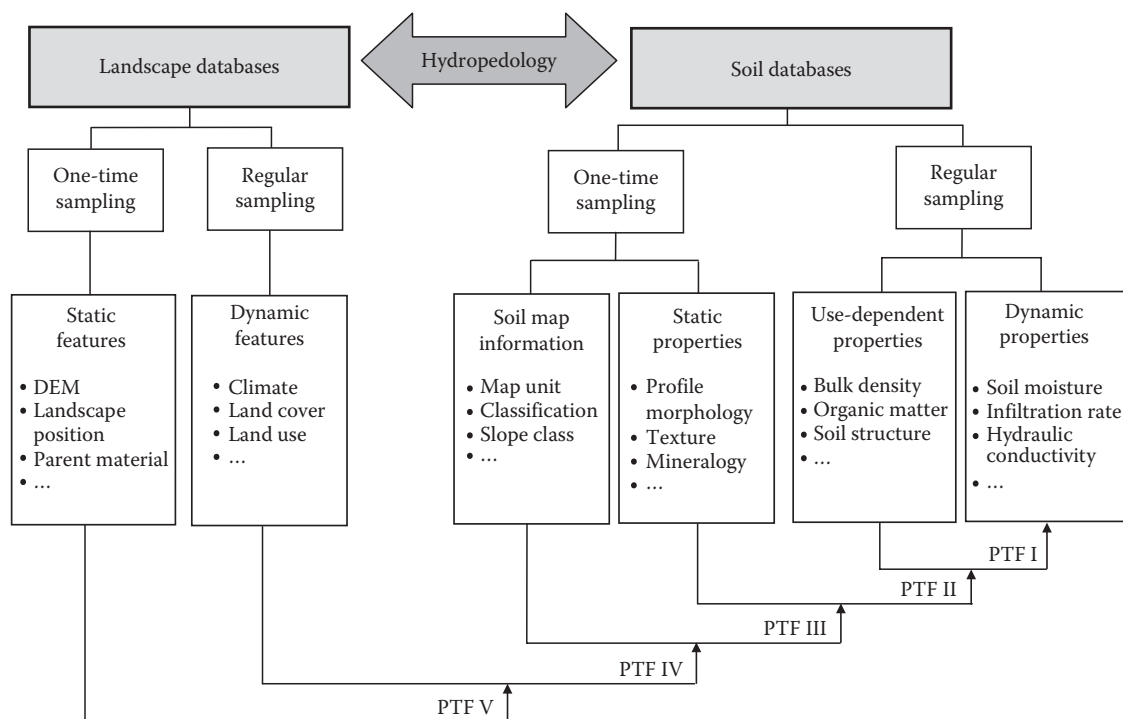

FIGURE 35.7 Five general categories of PTFs utilizing different types of input data for estimating dynamic soil properties. Landscape features such as DEMs, land use/land cover, and others serve as additional inputs to PTFs, hence connecting the pedon and the landscape scales.

soil properties to soil hydraulic information, both of dynamic nature requiring regular sampling. PTF type II includes relatively static soil properties that could be sampled only once into the prediction of dynamic soil properties. PTF type III further considers soil mapping and classification-related information to improve the prediction. Landscape features such as DEMs, land use/land cover, and others could serve as additional inputs in PTF type IV and V, hence connecting the pedon and landscape scales. It is likely that PTFs in combination with routine spatial information from soil survey, topography, and land use could improve the regional estimates of soil hydraulic parameters. In terms of land use and use-dependent soil properties, Droogers and Bouma (1997) suggested the terms "genoform" for genetically defined soil series and "phenoform" for soil types resulting from a particular form of management in a given genoform. Such distinction between major soil management types within the same soil series could potentially enhance PTFs that involve soil series and land use as carriers of soil hydraulic information. Realizing the importance of dynamic and use-dependent soil properties, the NCSS program is now considering the possible development of a dynamic soil properties database, which, once developed, would significantly facilitate the enhancement of PTFs.

Recent development in DSM coupled with earlier work on PTFs (Bouma, 1989; Bouma et al., 1999) has improved significantly the ability of soil science to transform itself from a semiquantitative to a quantitative discipline. This opportunity presents challenges at the same time, especially those related to understanding the scale dependency of hydrologic processes and their relation with spatial distribution of soil properties.

35.9 Digital Soil Mapping with Hydropedologic Concepts

In order to assess hydrologic function, soils and hydrologic processes must be fully understood within the context of the landscapes in which they occur. Soil serves as an interface between the solid and fluid envelopes, which dissipate mass and energy fluxes at the earth's surface (National Research Council, 2001). The majority of United States and World Soil Surveys are built upon soil–landscape model (Hudson, 1992), which recognizes the fact that in many landscapes, soil development occurs in response to the water movement through and over the landscape (Moore et al., 1993). This model limits the analysis of soil development at the scale where all the other soil-forming factors, namely, climate, parent material, organisms, and time, vary less compared to relief or topography. Jenny's model of five main soil-forming factors is expressed in the following equation:

$$S = f(cl, o, r, p, t. . .) \tag{35.3}$$

where S, soil, is considered to be a function of climate (cl), organisms (o), relief (r), and parent material (p) acting through time (t).

Aside from scientific value of the model, its widespread use may have been also a result of technical and practical considerations at the start of the Soil Survey Program related to purpose of the survey, scale, and resolution of the available spatial information. The use of Jenny's conceptual model leads to the development of maps relating soil pedogenesis and properties (Gessler et al., 1995) to landscape positions in a qualitative way (Walker et al., 1968; Furley, 1976; Daniels et al., 1985; Stone et al., 1985; Kreznor et al., 1989; Carter and Ciolkosz, 1991). The conventional soil maps reflect this approach in delineated soil polygon map units that represent the spatial continuum of soil properties with single values and ranges with no clear indication about their spatial distribution within map units. Soil scientists have recognized the inability to characterize the continuum and have compromised by providing polygon values with overlapping ranges for many soil properties. The use of this information in its current format presents a challenge for hydrologic modeling that requires a spatial continuum representation of soils and their properties (Libohova et al., 2010) as they modify material and energy fluxes at the earth's surface (Gessler et al., 1995). The challenge is to use hydropedologic principles to develop numerical and spatially realistic soil–landscape models useful for a variety of purposes beyond taxonomic classification (McSweeney et al., 1994). However, soil–landscape processes and relationships express themselves over a wide range of spatial and temporal scales, which reflect past climates, truncations by overriding processes, and perhaps feedback mechanisms (Malanson et al., 1990), and often exhibit different and complex scale-dependent variations (Butler, 1964; Beckett and Webster, 1971; Burrough, 1993; Wilding et al. 1994; Wilding and Drees, 1983). The scale dependency of soil variation indicates that more soil-forming factors may be at work at a particular location offering flexibility in incorporating other factors, especially location for explaining soil variability. Soil variability often relates well with terrain attributes, thus offering a potential use as spatial predictors of soil properties at specific locations (Gessler et al., 1995). With the dawning of GIS and powerful georeferencing tools, specific georeferenced spatial location has become a key consideration in the understanding of soils and their variability. A new formalization of the factors equation offered by McBratney et al. (2003) gives the following equation:

$$S_a = f(s, c, o, r, p, a, n) \tag{35.4}$$

where S_a is a set of soil attributes at a given location understood as a function of other observed soil attributes at either the present location or other proximal locations (s), climate (c), organisms (o), relief (r), parent materials (p), age or time (a), and spatial location or position (n). The formula offers an expansion of the equation discussed by Jenny (1941) to include soil properties observed at specific locations. It implies that, by using geostatistical tools, formal relationships between soils at measured distances can be developed to strengthen soil models. This is not a new idea; Jenny himself spent a great deal of time validating spatial soil relationships (Hudson, 1992), but it is a new formulation of the importance of spatial relationships and their power to help explain

soil properties. With available soil polygon data and soil point data, quantifiable terrain attributes (now often easily derived from DEMs in a GIS platform), and remotely sensed earth-surface-reflectance data, numerical analysis of soil–landscape relationships are possible to a much higher degree of accuracy and precision than ever before. The SCORPAN (S = soil, C = climate, O = organisms, R = topography, P = parent material, A = age and N = spatial position) model reflects the new emphasis on spatial location as a key attribute capable of providing predictive power in soil models. The increase recently in spatial resolution of DEMs and other spatial data coupled with the emerging of powerful terrain analysis software and geostatistical tools has provided an opportunity for soil scientists and digital soil mappers to numerically characterize soil–landscape models to develop raster-based predicted soil maps and soil property maps (Ryan et al., 2000). Soil–landscape model provides the appropriate scale for DSM to characterize soil variability as attributed to soil–water movement and distribution. This means that when climate, parent material, vegetation, and time are relatively constant, topography can explain the majority of differences in patterns of soil distribution. Topography itself is not the differentiating factor. However, topography controls water movement, which in turn affects pedogenesis. The use of terrain attributes as derived from topography captured by DEM provides the means to quantitatively represent the spatial distribution of soils and their related properties.

As previously discussed in the scale dependency of soils for modeling section, environmental modeling and management require explicit and quantitative spatial prediction of soil and landscape attributes (Gessler et al., 1995). Spatial variability affects hydrologic response over a wide range of scales (Sayfried and Wilcox, 1995; Robinson and Sivapalan, 1997) and is characterized as either stochastic and/or deterministic depending on the scale and the hydrologic process. This, in turn, means that any hydrological model has to represent soil model input parameters related to the scale of the research question, which can only be accomplished through DSM. Physically based distributed hydrologic models utilize spatially distributed inputs such as topography, soil type, and vegetation that in theory could provide a better description of hydrological processes (Bathurst and O'Connell, 1992). However, the amount of data input requirements could constitute a disadvantage for the physically based hydrologic models compared to lumped parameter models (Sayfried and Wilcox, 1995), and it may not always lead to better model predictions of hydrologic processes (Beven, 1987; Loague, 1990; Wicox et al., 1990; Anderson et al., 1992; Grayson et al., 1992).

The development of high-resolution spatial data and increase in computing power has led to demands for higher-resolution input data for hydrological modeling, especially for physically based distributed hydrology models. However, problems related to the limits of detail representation and parameterization of subgrid scale processes (within pixel) will still remain (Beven, 2001) and need to be addressed in the context of multiscale heterogeneities to identify their broad scale or general patterns (Sivapalan, 2005). The discussion on hydrological parameters, multiscale heterogeneities, and their scale dependency pattern

manifestation is closely related to the stochastic versus deterministic representation of spatial variability of hydrological parameters. This definition resonates with the soil–landscape model that could be considered as the "deterministic length scale" for describing the spatial variability of soils and their related properties. Both trends are found in every parameter, and the dominance of one or the other is scale dependent as previously discussed, which means that the future efforts in characterizing spatial variability of soil properties in a raster format as input to hydrological models should consider other scales in addition to the scale of soil–landscape model. This approach utilizing DSM is not limited only to hydrology and can be applied to other areas like ecology and environmental sciences and would lead to the development of soil property maps that emphasize their functions in relation to the process being hydrological, ecological, or other applications (Forman, 1995).

New spatial technologies (including GPS and GIS) have allowed the soil survey to georeference all of their recent soil samples as part of their soil characterization database (Soil Survey Staff, 2009) developed by the Soil Survey Laboratory (SSL), National Soil Survey Center. This database provides access to the raw characterization data, including morphology, saturated hydraulic conductivity, depth of soil, and bulk density. Using the source data to produce continuous maps of hydrologically important soil parameters should more clearly link real measurements with landscape position and provide a better set of soil parameters for hydrologic modeling. Given the recent developments in digital technologies like GIS, GPS, and spatial terrain analysis software coupled with DSM, spatially distributed raster-based predicted soil property maps for modeling can now be produced (McKenzie and Gallant, 2007; McKenzie and Ryan, 1999), validated, and tested with hydrologic models.

35.10 Conclusions

In 2001, the U.S. National Research Council (NRC, 2001) released a report entitled Basic Research Opportunities in Earth Science. This report highlighted the need for understanding the earth system and identified the pedosphere as an area crucial to the critical zone function. Hydropedology can contribute significantly to the holistic study of the earth's critical zone because of the key roles that soil and water play together in the critical zone's evolution and functioning. Understanding and quantifying the interactions between water and soil is critical for understanding multiscale ecosystem functions that sustain life. The future land use planning and use of natural resources under changing climate require multidisciplinary approaches where prediction and forecast based on quantifiable metrics must be achieved. Wilding and Lin (2006) pointed out a common agenda where soil science has been identified as a major stakeholder in the growing interests in bio- and geosciences as highlighted in the NRC (2001) report. This agenda was to (1) broaden constituencies beyond traditional partners; (2) expand focus in the near-surface "critical zone" to include food security, food safety, ecosystem management, biosphere sustainability, environmental protection, and urban

environment; (3) enhance fundamental knowledge of earth systems that is more systematic, interdisciplinary, dynamic, and process oriented; (4) identify early warning systems of natural hazards and resource degradation; and (5) develop joint research and education partnerships that attract coupling of public and private support. As described in this chapter, hydropedology provides one important mechanism where linking disciplines with a common theme can provide the avenue for a greater understanding of complex and dynamic interactions between the pedosphere and the hydrosphere.

References

Addiscott, T.M. 1995. Entropy and sustainability. Eur. J. Soil Sci. 46:161–168.

Alef, K., and D. Kleiner. 1989. Rapid and sensitive determination of microbial activity in soils and in soil aggregates by dimethylsulfoxide reduction. Biol. Fertil. Soils 8:349–355.

Anderson, A.N., A.B. McBratney, and E.A. FitzPatrick. 1996. Soil mass, surface, and spectral fractal dimensions estimated from thin section photographs. Soil Sci. Soc. Am. J. 60:962–969.

Anderson, S.H., and J.W. Hopmans. 1994. Tomography of soil-water-root processes. Soil Science Society of America, Special Pub. No. 36. American Society of Agronomy—Soil Science Society of America, Madison, WI.

Arndt, J.L., and J.L. Richardson. 1988. Hydrology salinity and hydric soil development in a North Dakota prairie pothole wetland system. Wetlands 8:94–107.

Arya, L.M., and J.F. Paris. 1981. A physico-empirical model to predict the soil moisture characteristics from particle-size distribution and bulk density data. Soil Sci. Soc. Am. J. 45:218–227.

Arya, L.M., F.J. Leij, M.Th. van Genuchten, and P.J. Shouse. 1999. Scaling parameter to predict the soil water characteristics from particle-size distribution data. Soil Sci. Soc. Am. J. 63:510–519.

Ashley, G.M., and S.G. Driese. 2000. Paleopedology and paleohydrology of a volcaniclastic paleosol interval: Implications for early pleistocene stratigraphy and paleoclimate record, Olduvai Gorge, Tanzania. J. Sediment. Res. 70:1065–1080.

Atkinson, T.C. 1978. Techniques for measuring subsurface flow on hillslopes. *In* M.J. Kirkby (ed.) Hillslope hydrology. John Wiley & Sons, Chichester, U.K.

Austin, A.T. 2002. Differential effects of precipitation on production and decomposition along a rainfall gradient in Hawaii. Ecology 83:328–338.

Balkwill, D.L., and W.C. Ghiorse. 1985. Characterization of subsurface bacteria associated with two shallow aquifers in Oklahoma. Appl. Environ. Microbiol. 50:580–588.

Balkwill, D.L., F.R. Leach, J.T. Wilson, J.F. McNabb, and D.C. White. 1988. Equivalence of microbial biomass measures based on membrane lipid and cell wall components, adenosine triphosphate and direct counts in subsurface aquifer sediments. Microb. Ecol. 16:73–84.

Bartoli, F., R. Phylippy, M. Doirisse, S. Niquet, and M. Dubuit. 1991. Structure and self-similarity in silty and sandy soils: The fractal approach. J. Soil Sci. 42:167–185.

Basu, N.B., P.S.C. Rao, H.E. Winzeler, S. Kumar, P.R. Owens, and V. Merwade. 2009. Parsimonious modeling of hydrologic responses in engineered Watersheds: Structural heterogeneity vs. functional homogeneity. Water Resour. Res. 46:W04501.

Bathurst, J.C., and P.E. O'Connell. 1992. Future of distributed modeling: The Systeme Hydrologique Europeen. Hydrol. Proc. 6:265–277.

Batjes, N.H. 1996. Development of a world data set of soil water retention properties using pedotransfer rules. Geoderma 71:31–52.

Baveye, P., and C.W. Boast. 1999. Physical scales and spatial predictability of transport processes in the environment, p. 261–280. *In* D.L. Corwin, K. Loague, and T.R. Ellsworth (eds.) Assessment of non-point source pollution in the vadose zone. Geophysics Monograph No. 108. American Geophysical Union, Washington, DC.

Baveye, P., J.-Y. Parlange, and B.A. Stewart. 1999. Fractals in soil science. Advances in Soil Science. CRC Press, Boca Raton, FL.

Beckett, P.H.T., and R. Webster. 1971. Soil variability: A review. Soils Fertil. 34:1–15.

Bell, J.C. 1990. A GIS-based soil-landscape modeling approach to predict soil drainage class. Ph.D. Thesis. Pennsylvania State University. University Park, PA.

Beven, K., and P. Germann. 1982. Macropores and water flow in soils. Water Resour. Res. 18:1311–1325.

Beven, K.J. 1987. Towards a new paradigm in hydrology. IASH Publ. No. 164, pp. 393–403.

Beven, K.J. 2001. Rainfall-runoff modelling. John Wiley & Sons, Chichester, U.K.

Beven, K.J. 2002. Towards an alternative blueprint for a physically-based digitally simulated hydrologic response modelling system. Hydrol. Process. 16:189–206.

Blöschl, G., and M. Sivapalan. 1995. Scale issues in hydrological modelling: A review. Hydrol. Process. 9:251–290.

Blume, E., M. Bischoff, J. Reichert, T. Moorman, A. Konopka, and R. Turco. 2002. Surface and subsurface microbial biomass, community structure and metabolic activity as a function of soil depth and season. Appl. Soil Ecol. 592:1–11.

Botros, F.E., Th. Harter, Y.S. Onsoy, A. Tuli, and J.W. Hopmans. 2009. Spatial variability of hydraulic properties and sediment characteristics in a deep alluvial unsaturated zone. Vadose Zone J. 8:276–289.

Boul, S.W., R.J. Southhard, R.C. Graham, and P.A. McDaniel. 2003. Soil genesis and classification. 5th edn. Iowa State Press, Ames, IA.

Bouma, J., and H.A.J. van Lanen. 1987. Transfer functions and threshold values: From soil characteristics to land qualities, p. 106–110. *In* K.J. Beek, P.A. Burrough and D.E. MC Cormack et al. (eds.) Proceedings of the ISSS/SSSA workshop on quantified land evaluation. Washington, DC.

April 27–May 2, 1986. International Institute for Aerospace Survey and Earth Sciences Publication No. 6. ITC Publications, Enschede, the Netherlands.

Bouma, J., J. Stoorvogel, B.J. van Alphen, and H.W.G. Booltink. 1999. Pedology, precision agriculture, and the changing paradigm of agricultural research. Soil Sci. Soc. Am. J. 63:1763–1768.

Bouma, J. 1989. Using soil survey data for quantitative land evaluation. Adv. Soil Sci. 9:177–213.

Bouma, J. 1990. Using morphometric expressions for macropores to improve soil physical analyses of field soils. Geoderma 46:3–13.

Bouma, J. 1992. Effects of soil structure, tillage, and aggregation upon soil hydraulic properties, p. 1–36. *In* R.J. Wagenet, P. Bavege, and B.A. Stewart (eds.) Interacting processes in soil science. Lewis Publishers, Boca Raton, FL.

Bouma, J. 2006. Hydropedology as a powerful tool for environmental policy research. Geoderma 131:275–286.

Bundt, M., F. Widmer, M. Pesaro, J. Zeyer, and P. Blaser. 2001. Preferential flow paths: Biological "hot spots" in soils. Soil Biol. Biochem. 33:729–738.

Buol, S.W., F.D. Hole, R.J. McCracken, and R.J. Southard. 1997. Soil genesis and classification. 4th edn. Iowa State University Press, Ames, IA.

Buol, S.W., R.J. Southard, R.C. Graham, and P.A. McDaniel. 2001. Soil genesis and classification. 5th edn. Iowa State University Press, Ames, IA.

Burke, I.C., and W.K. Lauenroth. 2002. Ecosystem ecology at regional scales. Ecology 83:305–306.

Burrough, P.A. 1993. Soil variability: A late 20th century view. Soils Fertil. 56:529–562.

Buss, S.R., J. Thrasher, P. Morgan, and J.W.N. Smith. 2006. A review of mecoprop attenuation in the subsurface. Q. J. Eng. Geol. Hydrogeol. 39:283–292.

Butler, B.E. 1964. Assessing the soil factor in agricultural production. J. Aust. Inst. Agric. Sci. 30:232–240.

Calmon, M.A., R.L. Day, E.J. Ciolkosz, and G.W. Peterson. 1998. Soil morphology as an indicator of soil hydrology on a hillslope underlain by a fragipan, p. 129–150. *In* M.C. Rabenhorst, J.C. Bell, and P.A. McDaniel (eds.) Quantifying soil hydromorphology. SSSA Special Publication No. 54. SSSA, Madison, WI.

Carter, B.J., and E.J. Ciolkosz. 1991. Slope gradient and aspect effects on soils developed from sandstone in Pennsylvania. Geoderma 49:199–213.

Castenson, K.L., and M.C. Rabenhorst. 2006. Indicator of reduction in soil (IRIS): Evaluation of a new approach for assessing reduced conditions in soil. Soil Sci. Soc. Am. J. 70:1222–1226.

Clothier, B.E., S.R. Green, and M. Deurer. 2008. Preferential flow and transport in soil: Progress and prognosis. Eur. J. Soil Sci. 59:2–13.

Corwin, D.L., and R.J. Wagenet. 1996. Applications of GIS to the modeling of nonpoint source pollutants in the vadose zone: A conference overview. J. Environ. Qual. 25:403–411.

Corwin, D.L., J. Hopmans, and G.H. de Rooij. 2006. From field-to landscape-scale vadose zone processes: Scale issues, modeling, and monitoring. Vadose Zone J. 5:129–139.

CUAHSI. Consortium of Universities for the Advancement of Hydrologic Science, Inc. 2007. Hydrology of a dynamic earth—A decadal research plan for hydrologic science. CUAHSI, Washington, DC.

Daniels, R.B., and S.W. Buol. 1992. Water table dynamics and significance to soil genesis. *In* J.M. Kimble (ed.) Proceedings of the 8th international soil correlation meeting (VIII ISCOM): Characterization, classification and utilization of wet soils. USDA-SCS, National Soil Survey Center, Lincoln, NE.

Daniels, R.B., E.E. Gamble, and L.A. Nelson. 1971. Relations between soil morphology and water table levels on a dissected North Carolina coastal plain surface. Soil Sci. Soc. Am. Proc. 35:781–784.

Daniels, R.B., J.W. Gilliam, D.K. Cassel, and L.A. Nelson. 1985. Soil erosion class and landscape position in the North Carolina Piedmont. Soil Sci. Soc. Am. J. 49:991–995.

Demas, G.P., and M.C. Rabenhorst. 2001. Factors of subaqueous soil formation: A system of quantitative pedology for submersed environments. Geoderma 102:189–204.

Deurer, M., S.R. Green, B.E. Clothier, J. Böttcher, and W.H.M. Duijnisveld. 2003. Drainage networks in soils: A concept to describe bypass-flow pathways. J. Hydrol. 272:148–162.

Dodds, W.K., M.K. Banks, C.S. Clenan, C.W. Rice, D. Sotomayor, E.A. Strauss, and W. Yu. 1996. Biological properties of soil and subsurface sediments under abandoned pasture and cropland. Soil Biol. Biochem. 28:837–846.

Droogers, P., and J. Bouma. 1997. Soil survey input in exploratory modeling of sustainable soil management practices. Soil Sci. Soc. Am. J. 61:1704–1710.

Dupuy, N.C., and B. Dreyfus. 1992. Bradyrhizobium populations occur in deep soil under the leguminous tree *Acacia albida*. Appl. Environ. Microbiol. 58:2415–2419.

Epstein, H.E., I.C. Burke, and W.K. Lauenroth. 2002. Regional patterns of decomposition and primary production rates in the U.S. Great Plains. Ecology 83:320–327.

Ersahin, S. 2003. Comparing kriging and cokriging to estimate infiltration rate. Soil Sci. Soc. Am. J. 67:1848–1855.

European Communities. 2007. Status and prospect of soil information in southeastern Europe: Soil databases, projects and applications, p. 189. Tomislav Hengl, Panos Panagos, Arwyn Jones and Gregely Tóth (eds.). Workshop Series EUR 22646 EN, ISBN 978-92-79-04972-9, ISSN 1018-5593.

Evans, C.V., and D.P. Franzmeier. 1986. Saturation, aeration, and color patterns in a toposequence of soils in north-central Indiana. Soil Sci. Soc. Am. J. 50:975–980.

Evans, C.V., and D.P. Franzmeier. 1988. Color index values to represent wetness and aeration in some Indiana soils. Geoderma 41:353–368.

Feigum, C.D. 2000. Evaluation of a drained lake basin and catchment utilizing soil, landscape position, and hydrology. Ph.D. Dissertation. North Dakota State University. Fargo, ND.

Fiedler, S., and M. Sommer. 2004. Water and redox conditions in wetland soils—Their influence on pedogenic oxides and morphology. Soil Sci. Soc. Am. J. 68:326–335.

Field, J.A., J.C. Parker, and N.L. Powell. 1984. Comparison of field- and laboratory-measured and predicted hydraulic properties of a soil with macropores. Soil Sci. 138:385–396.

Fierer, N., J.P. Schimel, and P.A. Holden. 2003. Variations in microbial community composition through two soil depth profiles. Soil Biol. Biochem. 35:167–176.

Flury, M., H. Flühler, W.A. Jury, and J. Leuenberger. 1994. Susceptibility of soils to preferential flow of water: A field study. Water Resour. Res. 30:1945–1954.

Flühler, H., W. Durner, and M. Flury. 1996. Lateral solute mixing processes—A key for understanding field-scale transport of water and solutes. Geoderma 70:165–183.

Forman, R.T.T. 1995. Land mosaics. The ecology of landscapes and regions. Cambridge University Press, Cambridge, U.K.

Franzmeier, D.P. 1991. Estimation of hydraulic conductivity from effective porosity data for some Indiana soils. Soil Sci. Soc. Am. J. 55:1801–1803.

Franzmeier, D.P., J.E. Yahner, G.C. Steinhardt, and H.R. Sinclair. 1983. Color patterns and water table levels in some Indiana soils. Soil Sci. Soc. Am. J. 47:1196–1202.

Freer, J., K.J. Beven, and B. Ambroise. 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resour. Res. 32:2161–2173.

Furley, P.A. 1976. Soil–slope–plant relationships in the northern Maya mountains, Belize, Central America. J. Biogeogr. 3:303–319.

Gessler, P.E., I.D. Moore, N.J. McKenzie, and P.J. Ryan. 1995. Soil–landscape modelling and spatial prediction of soil attributes. Int. J. Geogr. Inf. Sci. 9:4, 421–432.

Gish, T.J., C.L. Walthall, C.S.T. Daughtry, and K.-J.S. Kung. 2005. Using soil moisture and spatial yield patterns to identify subsurface flow pathways. J. Environ. Qual. 34:274–286.

Goodchild, C.W. 2007. Modelling the impact of climate change on domestic water demand. Water Environ. J. 17:8–12.

Gottlein, A., and B. Manderscheid. 1998. Spatial heterogeneity and temporal dynamics of soil water tension in a mature Norway spruce stand. Hydrol. Process. 12:417–428.

Graham, R.C., R.B. Daniels, and S.W. Buol. 1990. Soil-geomorphic relations on the Blue Ridge Fronts: 1. Regolith types and slope processes. Soil Sci. Soc. Am. J. 54:1362–1367.

Grayson, R., and G. Blöschl (eds.). 2000. Spatial pattern in catchment hydrology: Observations and modeling. Cambridge University Press, Cambridge, U.K.

Grayson, R.B., G. Blöschl, and I.D. Moore. 1995. Distributed parameter hydrologic modelling using vector elevation data: Thales and TAPES-C, p. 669–695. *In* V.P. Singh (ed.) Computer models of watershed hydrology. Water Resources Publications, Highlands Ranch, CO.

Griffin, R.W., L.P. Wilding, and L.R. Drees. 1992. Relating morphological properties to wetness conditions in the Gulf Coast prairie of Texas, p. 126–134. *In* J.M. Kimble (ed.) Proceedings of the 8th international soil correlation meeting (VIII ISCOM): Characterization, classification, and utilization of wet soils. October 1990. Lincoln, NE. USDA-SCS, National Soil Survey Center, Lincoln, NE.

Hall, G.F., and C.G. Olson. 1991. Predicting variability of soil from landscape models, p. 9–24. *In* M.J. Mausbach and L.P. Wilding (eds.) Spatial variabilities of soils and landforms. Soil Science Society of America Special Publication No. 28, Madison, WI.

Harlan, P.W., and D.P. Franzmeier. 1974. Soil–water regimes in Brookston and Crosby soils. Soil Sci. Soc. Am. Proc. 4:638–643.

Havens, A.V. 1998. Climate and microclimate of the New Jersey pine barrens, p. 113–131. *In* R.T.T. Forman (ed.) Pine barrens: Ecosystem and landscape. Rutgers University Press, Piscataway, NJ.

Haverkamp, R., and J.Y. Parlange. 1986. Predicting the water-retention curve from particle size distribution: 1. Sandy soils without organic matter. Soil Sci. 142:325–339.

Haws, N.W., B. Liu, C.W. Boast, P.S.C. Rao, E.J. Kladivko, and D.P. Franzmeier. 2004. Spatial variability and measurement scale of infiltration rate on an agricultural landscape. Soil Sci. Soc. Am. J. 68:1818–1826.

Hempel, J.W., R.D. Hammer, A.C. Moore, J.C. Bell, J.A. Thompson, and M.L. Golden. 2008. Challenges to digital soil mapping, p. 81–90. *In* A.E. Hartemink, A.B. McBratney, and M. de Lourdes Mendonca-Santos (eds.) Digital soil mapping with limited data. Springer, Berlin, Germany.

Hendrickx, J.M.H., and M. Flury. 2001. Uniform and preferential flow mechanisms in the vadose zone, p. 149–187. *In* Conceptual models of flow and transport in the fractured vadose zone. National Academy Press, Washington, DC.

Hopkins, D.G. 1996. Hydrologic and abiotic constraints on soil genesis and natural vegetation patterns in the Sand Hills of North Dakota. Ph.D. Dissertation. North Dakota State University. Fargo, ND.

Hopmans, J.W., and G.H. Schoups. 2005. Soil water flow and different spatial scales, p. 999–1010. *In* M.G. Anderson (ed.) Encyclopedia of hydrologic sciences. John Wiley & Sons Ltd., Chichester, U.K.

Hudson, B.D. 1992. The soil survey as paradigm-based science. Soil Sci. Soc. Am. J. 56:836–841.

Hurlbert, S.H. 1971. The non-concept of species diversity: A critique and alternative parameters. Ecology 52:577–586.

Ibanez, J.J., S. De-Alba, F.F. Bermuidez, and A. Garcia-Alvarez. 1995. Pedodiversity: Concepts and measures. Catena 24:215–232.

Indorante, S.J., R.L. McLeese, R.D. Hammer, B.W. Thompson, and D.L. Alexander. 1996. Positioning soil survey for the 21st century. J. Soil Water Conserv. 51:21–28.

Iqbal, J., J.A. Thomasson, J.N. Jenkins, P.R. Owens, and F.D. Whisler. 2005. Spatial variability analysis of soil physical properties of alluvial soils. Soil Sci. Soc. Am. J. 69:1338–1350.

Jacobs, P.M., L.T. West, and J.N. Shaw. 2002. Redoximorphic features as indicators of seasonal saturation, Loundes County, Georgia. Soil Sci. Soc. Am. J. 66:315–323.

Jenkinson, B.J. 1998. Wet soil monitoring project on two till plains in south and west-central Indiana. M.S. Thesis. Purdue University. West Lafayette, IN.

Jenkinson, B.J. 2002. Hydrology of sandy soils in northwest Indiana and iron oxide indicators to identify hydric soils. Ph.D. Thesis. Purdue University. West Lafayette, IN.

Jenkinson, B.J., D.P. Franzmeier, and W.C. Lynn. 2002. Soil hydrology on an end moraine and a dissected till plain in west–central Indiana. Soil Sci. Soc. Am. J. 66:1367–1376.

Jenkinson, B.J., and D.P. Franzmeier. 2006. Development and evaluation of Fe coated tubes that indicate reduction in soils. Soil Sci. Soc. Am. J. 70:183–191.

Jenny, H. 1941. Factors of soil formation: A system of quantitative pedology. McGraw-Hill Book Company, New York.

Kavdir, Y., and A.J.M. Smucker. 2005. Soil aggregate sequestration of cover crop root and shoot derived soil nitrogen. Plant Soil 272:263–276.

Kempton, R.A. 1979. Structure of species abundance and measurement of diversity. Biometrics 35:307–322.

Khan, F.A., and T.E. Fenton. 1994. Saturated zones and soil morphology in a Mollisol catena of Central Iowa. Soil Sci. Soc. Am. J. 58:1457–1464.

Khan, F.A., and T.E. Fenton. 1996. Secondary iron and manganese distributions and aquic conditions in a Mollisol catena of Central Iowa. Soil Sci. Soc. Am. J. 60:546–551.

Kie, J.G., R.T. Bowyer, M.C. Nicholson, B.B. Boroski, and E.R. Loft. 2002. Landscape heterogeneity at differing scales: Effects on spatial distribution of Mule deer. Ecology 83:530–544.

Kieft, T.L., D.B. Ringelberg, and D.C. White. 1994. Changes in ester-linked phospholipid fatty acid profiles of subsurface bacteria during starvation and desiccation in a porous medium. Appl. Environ. Microbiol. 60:3292–3299.

King, J.J., and D.P. Franzmeier. 1981. Estimation of saturated hydraulic conductivity from soil morphological and genetic information. Soil Sci. Soc. Am. J. 45:1153–1156.

Kirchner, J.W. 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resour. Res. 42:W03S04. doi:10.1029/2005WR004362

Konopka, A., and R.F. Turco. 1991. Biodegradation of organic compounds in vadose zone and aquifer sediments. Appl. Environ. Microbiol. 57:2260–2268.

Kotz, J.C., M.D. Joesten, J.L. Wood, and J.W. Moore. 1994. The chemical world. Harcourt Brace and Co., Orlando, FL.

Kozak, J.A., and L.R. Ahuja. 2005. Scaling of infiltration and redistribution of water across soil textural classes. Soil Sci. Soc. Am. J. 69:816–827.

Krasovskaia, I., L. Gottschalk, E. Leblois, and E. Sauquet. 2003. Dynamics of river flow regimes viewed through attractors. Nordic Hydrol. 34:461–476.

Kreznor, W.R., K.R. Olson, W.L. Banwart, and D.L. Johnson. 1989. Soil, landscape, and erosion relationships in a northwest Illinois watershed. Soil Sci. Soc. Am. J. 53:1763–1771.

Kubiena, W.L. 1938. Micropedology. Collegiate Press, Ames, IA.

Kung, K.-J.S. 1990. Preferential flow in a sandy vadose zone: 1. Field observation. Geoderma 46:51–58.

Kutílek, M. 2004. Soil hydraulic properties as related to soil structure. Soil Till. Res. 79:175–184.

Latshaw, G.J., and R.F. Thompson. 1968. Water table study verifies soil interpretations. J. Soil Water Conserv. 23:65–67.

Laurenson, E.M. 1974. Modeling of stochastic-deterministic hydrologic systems. Water Resour. Res. 10(5):955–961.

Lehmann, P., C. Hinz, G. McGrath, H.J. Tromp-van Meerveld, and J.J. McDonnell. 2007. Rainfall threshold for hillslope outflow: An emergent property of flow pathway connectivity. Hydrol. Earth Syst. Sci. 11:1047–1063.

Leij, F., W.J. Alves, M.Th. van Genuchten, and J.R. Williams. 1996. The UNSODA unsaturated soil hydraulic database. User's manual version 1.0. EPA/600/R-96/095. National Risk Management Laboratory, Office of Research and Development, Cincinnati, OH.

Li, K., A. Amoozegar, W.P. Robarge, and S.W. Buol. 1997. Water movement and solute transport through saprolite. Soil Sci. Soc. Am. J. 61:1738–1745.

Li, X.Y., Z.P. Yang, X.Y. Zhang, and H.S. Lin. 2009. Connecting ecohydrology and hydropedology in desert shrubs: Stemflow as a source of preferential water flow in soils. Hydrol. Earth Syst. Sci. 13:1133–1144.

Libohova, Z. 2010. Terrain attribute soil mapping for predictive continuous soil property maps. Ph.D. Dissertation. Purdue University. West Lafayette, IN.

Lilly, A. 1997. A description of the HYPRES database (Hydraulic Properties of European Soils). *In* Bruand, A. et al. (eds.) The use of pedotransfer in soil hydrology research in Europe. INRA Orleans and EC/JRC Ispra, Orleans, France.

Lin, H.S., and X. Zhou. 2008. Evidence of subsurface preferential flow using soil hydrologic monitoring in the Shale Hills catchment. Eur. J. Soil Sci. 59:34–49.

Lin, H.S., E. Brook, P. McDaniel, and J. Boll. 2008b. Hydropedology and surface/subsurface runoff processes. *In* M.G. Anderson (ed.) Encyclopedia of hydrologic sciences. John Wiley & Sons, Ltd., New York.

Lin, H.S., J. Bouma, L. Wilding, J. Richardson, M. Kutilek, and D. Nielsen. 2005. Advances in hydropedology. Adv. Agron. 85:1–89.

Lin, H.S., J. Bouma, P. Owens, and M. Vepraskas. 2008a. Hydropedology: Fundamental issues and practical applications. Catena 73:151–152.

Lin, H.S., J. Bouma, Y. Pachepsky, A. Western, J. Thompson, M.Th. van Genuchten, H. Vogel, and A. Lilly. 2006. Hydropedology: Synergistic integration of pedology and hydrology. Water Resour. Res. 42:W05301. doi:10.1029/2005WR004085

Lin, H.S., K.J. McInnes, L.P. Wilding, and C.T. Hallmark. 1996. Effective porosity and flow rate with infiltration at low tensions into a well-structured subsoil. Trans. ASAE 39:131–135.

Lin, H.S., K.J. McInnes, L.P. Wilding, and C.T. Hallmark. 1997. Low tension water flow in structured soils. Can. J. Soil Sci. 77:649–654.

Lin, H.S., K.J. McInnes, L.P. Wilding, and C.T. Hallmark. 1998. Macroporosity and initial moisture effects on infiltration rates in vertisols and vertic intergrades. Soil Sci. 163:2–8.

Lin, H.S., K.J. McInnes, L.P. Wilding, and C.T. Hallmark. 1999a. Effects of soil morphology on hydraulic properties: I. Quantification of soil morphology. Soil Sci. Soc. Am. J. 63:948–954.

Lin, H.S., K.J. McInnes, L.P. Wilding, and C.T. Hallmark. 1999b. Effects of soil morphology on hydraulic properties: II. Hydraulic pedotransfer functions. Soil Sci. Soc. Am. J. 63:955–961.

Lin, H.S., K. Singha, D. Chittleborough, H.-J. Vogel, and S. Mooney. 2008c. Advancing the emerging field of hydropedology. EOS Trans. Am. Geophys. Union 89:490.

Lin, H.S. 2003. Hydropedology: Bridging disciplines, scales, and data. Vadose Zone J. 2:1–11.

Lin, H.S. 2010. Earth's critical zone and hydropedology: Concepts, characteristics, and advances. Hydrol. Earth Syst. Sci. 14:25–45.

Lin, H.S. 2010. Linking principles of soil formation and flow regimes. J. Hydrol. doi:10.1016/j.jhydrol.2010.02.013. 393: 3–19.

Loague, K. 1990. Changing ideas in hydrology—The case of physically based models—Comment. J. Hydrol. 120:405–407.

Logsdon, S.D., E. Perfect, and A.M. Tarquis. 2008. Multiscale soil investigations: Physical concepts and mathematical techniques. Vadose Zone J. 7:453–455.

Lohse, K.A., and W.E. Dietrich. 2005. Contrasting effects of soil development on hydrological properties and flow paths. Water Resour. Res. 41:W12419.

Lohse, K.A., P.D. Brooks, J.C. McIntosh, T. Meixner, and T.E. Huxman. 2009. Interactions between biogeochemistry and hydrologic systems. Annu. Rev. Environ. Resour. 34:1.1–1.32.

Lu, H., C.J. Moran, and M. Sivapalan. 2005. A theoretical exploration of catchment-scale sediment delivery. Water Resour. Res. 41:W09415. doi:10.1029/2005WR004018

Luo, L.F., H.S. Lin, and P. Halleck. 2008. Quantifying soil structure and preferential flow in intact soil using X-ray computed tomography. Soil Sci. Soc. Am. J. 72:1058–1069.

MacMillan, R.A. 2008. Chapter 10. Experiences with applied DSM: Protocol, availability, quality and capacity building. *In* A.E. Hartemink, A.B. McBratney, and M. de Lourdes Mendonca-Santos (eds.) Digital soil mapping with limited data. Springer, Berlin, Germany.

MacMillan, R.A. 2009. Mapping the world's soil. V1 Magazine (accessed online on January 30, 2010, at http://www.vector1media.com/dialogue/interviews/ 9388-globalsoilmap-mapping-the-worlds-soil).

Malanson, G.P., D.R. Butler, and S.J. Walsh. 1990. Chaos theory in physical geography. Phys. Geogr. 11:293–304.

Mausbach, M.J., and L.P. Wilding (eds.) 1991. Spatial variabilities of soils and landforms. SSSA Special Publications No. 28. SSSA, Madison, WI.

McBratney, A.B., and R. Webster. 1983. Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. J. Soil Sci. 34:137–162.

McBratney, A.B., B. Minasny, S.R. Cattle, and R.W. Vervoort. 2002. From pedotransfer functions to soil inference systems. Geoderma 109:41–73.

McBratney, A.B., I.O.A. Odeh, Th.F.A. Bishop, M.S. Dunbar, and T.M. Shatar. 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97:293–327.

McBratney, A.B., M.L. Mendonca Santos, and B. Minasny. 2003. On digital soil mapping. Geoderma 117:3–52.

McClain, M.E., E.W. Boyer, C.L. Dent, S.E. Gergel, N.B. Grimm, P.M. Groffman, S.C. Hart et al. 2003. Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. Ecosystems 6:301–312.

McDonnell, J.J., M. Sivapalan, K. Vaché, S. Dunn, G. Grant, R. Haggerty, C. Hinz et al. 2007. Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. Water Resour. Res. 43:W07301.

McGlynn, B., J.J. McDonnell, and D. Brammer. 2002. A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchment, New Zealand. J. Hydrol. 257:1–26.

McKenzie, N., and J. Gallant. 2007. Digital soil mapping with improved environmental predictors and models of pedogenesis, p. 327–352. *In* P. Lagacherie, A.B. McBratney, and M. Voltz (eds.) Digital soil mapping, an introductory perspective. Developments in soil science. Vol. 31, Chapter 24. Elsevier, Amsterdam, the Netherlands.

McKenzie, N.J., and D.W. Jacquier. 1997. Improving the field estimation of saturated hydraulic conductivity in soil survey. Aust. J. Soil Res. 35:803–825.

McKenzie, N.J., and P.J. Ryan. 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89:67–94.

McSweeney, K., P.E. Gessler, D. Hammer, J. Bell, and G.W. Petersen. 1994. Towards a new framework for modeling the soil–landscape continuum, p. 127–145. *In* R.G. Amundsen et al. J.W. Harden and M.J. Singer (eds.) Factors of soil formation: A 50th anniversary perspective. SSSA Special Publications No. 33. SSSA, Madison, WI.

Megonigal, J.P., W.H. Patrick, Jr., and S.P. Faulkner. 1993. Wetland identification in seasonally flooded forest soils: Soil morphology and redox dynamics. Soil Sci. Soc. Am. J. 57:140–149.

Miller, E.E., and R.D. Miller. 1956. Physical theory for capillary flow phenomena. J. Appl. Phys. 4:324–332.

Moore, I.D., A. Lewis, and J.C. Gallant. 1994. Terrain attributes: Estimation methods and scale effects, p. 189–214. *In* A.J. Jakeman, M.B. Beck, and M. McAleer (eds.) Modelling change in environmental systems. Wiley, London, U.K.

Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Petersen. 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57:43–452.

Moore, I.D., R.B. Grayson, and A.R. Ladson. 1991. Digital terrain modelling: Review of hydrological, geomorphological, and biological applications. Hydrol. Process. 5:3–30.

Mualem, Y. 1986. Hydraulic conductivity of unsaturated soils: Predictions and formulas, p. 799–823. *In* A. Klute (ed.) Methods of soil analysis. Part 1. 2nd edn. ASA, Madison, WI.

Murphy, D.V., I.R.P. Fillery, and G.P. Sparling. 1997. Method to label soil cores with $^{15}NH_3$ gas as a prerequisite for ^{15}N isotopic dilution and measurement of gross N mineralization. Soil Biol. Biochem. 29:1731–1741.

Netto, M.A., R.A. Pieritz, and J.P. Gaudet. 1999. Field study of local variability of soil water content and solute concentration. J. Hydrol. 215:22–37.

Noguchi, S., Y. Tsuboyama, R.C. Sidle, and I. Hosoda. 1999. Morphological characteristics of macropores and the distribution of preferential flow pathways in a forested slope segment. Soil Sci. Soc. Am. J. 63:1413–1423.

NRC (National Research Council). 1991. Opportunities in the hydrologic sciences. National Academy Press, Washington, DC.

NRC (National Research Council). 2001. Basic research opportunities in earth science. National Academy Press, Washington, DC.

Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1992. Soil pattern recognition with fuzzy-c-mean: Application to classification and soil–landform interrelationships. Soil Sci. Soc. Am. J. 56:499–504.

Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63:197–214.

Owens, P.R., L.P. Wilding, W.M. Miller, and R.W. Griffin. 2008. Using iron metal rods to infer oxygen status in seasonally saturated soils. Catena 73:197–203.

Owens, P.R. 2001. Inferring oxygen status in soils using iron metal rods. Ph.D. Dissertation. Texas A&M University. College Station, TX.

Pachepsky, Y.A., and W.J. Rawls (eds.). 2004. Development of pedotransfer functions in soil hydrology. Elsevier, Amsterdam, the Netherlands.

Pachepsky, Y.A., and W.J. Rawls. 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. Soil Sci. Soc. Am. J. 63:1748–1757.

Pachepsky, Y.A., and W.J. Rawls. 2003. Soil structure and pedotransfer functions. Eur. J. Soil Sci. 54:443–452.

Pachepsky, Y., D. Gimenez, A. Lilly, and A. Nemes. 2008. Promises of hydropedology. CAB Rev. 3:1–19.

Pachepsky, Y.A., W.J. Rawls, and D.J. Timlin. 1999. The current status of pedotransfer functions: Their accuracy, reliability, and utility in field- and regional-scale modeling, p. 223–234. *In* D. Corwin, K. Loague, and T. Ellsworth (eds.) Assessment of non-point source pollution in the vadose zone. American Geophysical Union, Washington, DC.

Pachepsky, Y.A., W.J. Rawls, and H.S. Lin. 2006. Hydropedology and pedotransfer functions. Geoderma 131:308–316.

Peck, A.J. 1983. Field variability of soil physical properties, p. 189–221. *In* D. Hillel (ed.) Advances in irrigation. Academic Press, San Diego, CA.

Perret, J.S., S.O. Prasher, and A.R. Kacimov. 2003. Mass fractal dimension of soil macropores using computed tomography: From the box-counting to the cube-counting algorithm. Eur. J. Soil Sci. 54:569–579.

Perret, J., S.O. Prasher, A. Kantzas, and C. Langford. 1999. Three-dimensional quantification of macropore networks in undisturbed soil cores. Soil Sci. Soc. Am. J. 63:1530–1543.

Perrier, E., Bird, N., and Rieu, M. 1999. Generalizing the fractal model of soil structure: The PSF approach. Geoderma 88:137–164.

Peyton, R.L., C.J. Gantzer, S.H. Anderson, B.A. Haeffner, and P. Pfeifer. 1994. Fractal dimension to describe soil macropore structure using X ray computed tomography. Water Resour. Res. 30:691–700.

Phillips, J.D. 2001. The relative importance of intrinsic and extrinsic factors in pedodiversity. Ann. Assoc. Am. Geogr. 91:609–621.

Ponnamperuma, F.N. 1972. The chemistry of submerged soils. Adv. Agron. 24:29–96.

Quisenberry, V.L., B.R. Smith, R.E. Phillips, H.D. Scott, and S. Nortcliff. 1993. A soil classification system for describing water and chemical transport. Soil Sci. 156:306–315.

Rabenhorst, M.C., J.C. Bell, and P.A. McDaniel. 1998. Quantifying soil hydromorphology. SSSA Special Publication No. 54. SSSA, Madison, WI.

Rabenhorst, M.C., R.R. Bourgault, and B.R. James. 2008. Iron oxyhydroxide reduction in simulated wetland soils; effects of mineralogical composition of IRIS paints. Soil Sci. Soc. Am. J. 72:1838–1842.

Rabenhorst, M.C., W.D. Hively, and B.R. James. 2009. Measurements of soil redox potential. Soil Sci. Soc. Am. J. 73:668–674.

Richardson, J.L., and D.A. Lietzke. 1983. Weathering profiles in fluvial sediments of the Middle Coastal Plain of Virginia. Soil Sci. Soc. Am. J. 47:301–304.

Richardson, J.L., J.L. Arndt, and J.A. Montgomery. 2001. Hydrology of wetland and related soils, p. 35–84. *In* J.L. Richardson and M.J. Vepraskas (eds.) Wetland soils: Genesis, hydrology, landscapes, and classification. Lewis, CRC Press, Boca Raton, FL.

Richardson, J.L., L.P. Wilding, and R.B. Daniels. 1992. Recharge and discharge of groundwater in aquic conditions illustrated with flownet analysis. Geoderma 53:65–78.

Richter, D.D., and D. Markewitz. 2001. Understanding soil change: Soil sustainability over millennia, centuries, and decades. Cambridge University Press, Cambridge, U.K.

Robinson, J.S., and M. Sivapalan. 1997. An investigation into the physical causes of scaling and heterogeneity of regional flood frequency. Water Resour. Res. 33:1045–1059.

Rodriguez-Iturbe, I., and A. Rinaldo. 1997. Fractal river basins, chance and self-organization. Cambridge University Press, Cambridge, U.K.

Rodriguez-Cruz, M.S., M.J. Sanchez-Martin, and M. Sanchez-Camazano. 2006. Surfactant-enhanced desorption of Atrazine and Linuron residues as affected by aging of herbicides in soil. Arch. Environ. Contam. Toxicol. 50:128–137.

Runge, E.C.A. 1973. Soil development sequences and energy models. Soil Sci. 112:183–193.

Ryan, P.J., N.J. McKenzie, D. O'Connell, A.N. Loughhead, P.M. Leppert, D. Jacquier, and L. Ashton. 2000. Integrating forest soils information across scales: Spatial prediction of soil properties under Australian forests. For. Ecol. Manage. 138:139–157.

Sayfried, M.S., and B.P. Wilcox. 1995. Scale and the nature of spatial variability: Field examples having implications for hydrological modeling. Water Resour. Res. 31:173–184.

Schaap, M.G., F.J. Leij, and M.Th. van Genuchten. 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. 251:163–176.

Schoeneberger, P.J., and D.A. Wysocki. 2005. Hydrology of soils and deep regolith: A nexus between soil geography, ecosystems and land management. Geoderma 126:117–128.

Schoeneberger, P.J., D.A. Wysocki, E.C. Benham, and W.D. Broderson (eds.). 2002. Field book for describing and sampling soils. Version 2.0. USDA–NRCS, National Soil Survey Center, Lincoln, NE.

Scott, H.D. 2000. Soil physics: Agricultural and environmental applications. Iowa State University Press, Ames, IA.

Shaw, J.N., L.T. West, D.E. Radcliffe, and D.D. Bosch. 2000. Preferential flow and pedotransfer functions for transport properties in sandy Kandiudults. Soil Sci. Soc. Am. J. 64:670–678.

Shepard, J.C. 1993. Using a fractal model to compute the hydraulic conductivity function. Soil Sci. Soc. Am. J. 57:300–306.

Shirazi, M.A., C.B. Johnson, J.M. Omernik, D. White, P.K. Haggerty, and G.E. Griffith. 2003. Quantitative soil descriptions for ecoregions of the United States. J. Environ. Qual. 32:550–561.

Shovic, H.F., and C. Montagne. 1985. Application of a statistical soil-landscape model to an order III wildland soil survey. Soil Sci. Soc. Am. J. 49:961–968.

Simmons, B.L., R.K. Niles, and D.H. Wall. 2008. Distribution and abundance of alfalfa-field nematodes at various spatial scales. Appl. Soil Ecol. 38:211–222.

Simonson, G.H., and L. Boersma. 1972. Soil morphology and water table relations: Correlation between annual water table fluctuations and profile features. Soil Sci. Soc. Am. Proc. 36:649–653.

Simonson, R.W. 1959. Outline of generalized theory of soil genesis. Soil Sci. Soc. Am. Proc. 23:152–156.

Simunek, J., N.J. Jarvis, M.Th. van Genuchten, and A. Gärdenäs. 2003. Review and comparison of models for describing non equilibrium and preferential flow and transport in the vadose zone. J. Hydrol. 272:14–35.

Sivapalan, M. 2005. Pattern, process and function: Elements of a unified theory of hydrology at the catchment scale, p. 27. *In* M.G. Anderson (ed.) Encyclopedia of hydrological sciences. John Wiley & Sons, Ltd., Chichester, U.K.

Soil Survey Division Staff. 1993. Soil survey manual. USDA Handbook 18. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1999. Soil taxonomy—A basic system of soil classification for making and interpreting soil surveys. 2nd edn. USDA–NRCS. Agricultural Handbook No. 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2003. Key to soil taxonomy. 9th edn. USDA–NRCS. US Government Printing Office, Washington, DC.

Soil Survey Staff. 2004. Soil survey geographic (SSURGO) database for the United States, USDA–NRCS. US Government Printing Office, Washington, DC.

SoilVision Systems Ltd. 2002. SoilVision® Soils dataset (available online at http://www.soilvision.com/Additional_domains/soildatabase.com/index.shtml). Soil Vision Systems, Saskatoon, Saskatchewan, Canada.

Sommer, M. 2006. Influence of soil pattern on matter transport in and from terrestrial biogeosystems—A new concept for landscape pedology. Geoderma 133:107–123.

Sposito, G. (ed.). 1998. Scale dependence and scale invariance in hydrology. Cambridge University Press, Cambridge, U.K.

Sposito, G., and R.J. Reginato (eds.). 1992. Opportunities in basic soil science research. SSSA, Madison, WI.

Stone, J.R., J.W. Gilliam, D.K. Cassel, R.B. Daniels, L.A. Nelson, and H.J. Kleiss. 1985. Effect of erosion and landscape position on the productivity of piedmont soils. Soil Sci. Soc. Am. J. 49:987–991.

Taylor, J.P., B. Wilson, M.S. Mills, and R.G. Burns. 2002. Comparison of microbial numbers and enzymatic activities in surface soils and subsoils using various techniques. Soil Biol. Biochem. 34:387–401.

Thompson, J.A., and J.C. Bell. 1996. Color index for identifying hydric conditions for seasonally saturated Mollisols in Minnesota. Soil Sci. Soc. Am. J. 60:1979–1988.

Thompson, J.A., J.C. Bell, and C.A. Butler. 1997. Quantitative soil-landscape modeling for estimating the areal extent of hydromorphic soils. Soil Sci. Soc. Am. J. 61:971–980.

Tietje, O., and V. Hennings. 1996. Accuracy of the saturated hydraulic conductivity prediction by pedo-transfer functions compared to the variability within FAO textural classes. Geoderma 69:71–84.

Tillotson, P.M., and D.R. Nielsen. 1984. Scale factors in soil science. Soil Sci. Soc. Am. J. 48:953–959.

Tisdall, J.M. 1995. Formation of soil aggregates and accumulation of soil organic matter, p. 57–96. *In* M. Carter and B. Stewart (eds.) Structure and organic matter storage in agricultural soils. Advances in Soil Science, CRC Press, New York.

Tisdall, J.M., and Oades, J.M. 1982. Organic matter and water-stable aggregates in soils. J. Soil Sci. 62:141–163.

Tugel, A.J., J.E. Herrick, J.R. Brown, M.J. Mausbach, W. Puckett, and K. Hipple. 2005. Soil change, soil survey, and natural resources decision making: A blueprint for action. Soil Sci. Soc. Am. J. 69:738–747.

USDA–NRCS, National Geospatial Development Center. 2006. Assessment and implementation of digital soil mapping in the National Cooperative Soil Survey: A challenge dialogue with the Soil Survey Community. USDA–NRCS, Lincoln, NE.

van Genuchten, M.Th., M.G. Schaap, B.P. Mohanty, J. Simunek, and F.J. Leij. 1999b. Modeling flow and transport processes at the local scale, p. 23–45. *In* J. Feyen and K. Wiyo (eds.) Modelling of transport process in soils at various scales in time and space. Wageningen Press, Wageningen, the Netherlands.

van Genuchten, M.Th., R.L. Feike, and L. Wu (eds.). 1999a. Proc. Int. Workshop Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media. University of California, Riverside, CA.

Van Gestel, M., J. Ladd, and M. Amato. 1992. Microbial biomass responses to seasonal change and imposed drying regimes at increasing depths of undisturbed topsoil profiles. Soil Biol. Biochem. 24:103–111.

Vaughan, K.L., M.C. Rabenhorst, and B.A. Needelman. 2009. Saturation and temperature effects on the development of reducing conditions in soils. Soil Sci. Soc. Am. J. 73:663–667.

Veneman, P.L.M., D.L. Lindbo, and L.A. Spokas. 1998. Soil moisture and redoximorphic features: A historical perspective, p. 1–23. *In* M.C. Rabenhorst, J.C. Bell, and P.A. McDaniel (eds.) Quantifying soil hydromorphology. SSSA Special Publication No. 54. SSSA, Madison, WI.

Veneman, P.L.M., M.J. Vepraskas, and J. Bouma. 1976. The physical significance of soil mottling in a Wisconsin toposequence. Geoderma 15:103–118.

Vepraskas, M.J., and J. Bouma. 1976. Model experiments on mottle formation simulating field conditions. Geoderma 15:217–230.

Vepraskas, M.J., and L.P. Wilding. 1983a. Albic neoskeletans in argillic horizons as indices of seasonal saturation and iron reduction. Soil Sci. Soc. Am. J. 47:1202–1208.

Vepraskas, M.J., and L.P. Wilding. 1983b. Aquic moisture regimes in soils with and without low chroma colors. Soil Sci. Soc. Am. J. 47:280–285.

Vepraskas, M.J., and S.P. Faulkner. 2001. Redox chemistry of hydric soils, p. 85–105. *In* J.L. Richardson and M.J. Vepraskas (eds.) Wetland soils. CRC Press, Boca Raton, FL.

Vepraskas, M.J. 1995. Redoximorphic features for identifying aquic conditions. Technical Bulletin No. 301. North Carolina State University, Raleigh, NC.

Vereecken, H., J. Maes, and J. Feyen. 1990. Estimating unsaturated hydraulic conductivity from easily measured soil properties. Soil Sci. 149:1–12.

Vereecken, H., R. Kasteel, J. Vanderborght, and T. Harter. 2007. Upscaling hydraulic properties and soil water flow processes in heterogeneous soils: A review. Vadose Zone J. 6:1–28.

Verhagen, J., and J. Bouma. 1998. Defining threshold values for residual N-levels. Geoderma 85:199–213.

Vervoort, R.W., and S.R. Cattle. 2003. Linking hydraulic conductivity and tortuosity parameters to pore space geometry and pore-size distribution. J. Hydrol. 272:36–49.

Vervoort, R.W., D.E. Radcliffe, and L.T. West. 1999. Soil structure development and preferential solute flow. Water Resour. Res. 35:913–928.

Vogel, H.J., and K. Roth. 2003. Moving through scales of flow and transport in soil. J. Hydrol. 272:95–106.

Vogel, H.J., I. Cousin, and K. Roth. 2002. Quantification of pore structure and gas diffusion as a function of scale. Eur. J. Soil Sci. 53:465–473.

Vogel, T., H.H. Gerke, R. Zhang, and M.Th. Van Genuchten. 2000. Modeling flow and transport in a two-dimensional dual-permeability system with spatially variable hydraulic properties. J. Hydrol. 238:78–89.

Wagenet, R.J., J. Bouma, and R.B. Grossman. 1991. Minimum data sets for use of soil survey information in soil interpretive models, p. 161–182. *In* M.J. Mausbach and L.P. Wilding (eds.) Spatial variabilities of soils and landforms. SSSA Special Publication No. 28. SSSA, Madison, WI.

Wagner, W., V. Naeimi, K. Scipal, R. de Jeu, and J. Martınez-Fernandez. 2007. Soil moisture from operational 590 meteorological satellites. Hydrogeol. J. 15:121–131.

Walker, P.H., G.F. Hall, and R. Protz. 1968. Relation between landform parameters and soil properties. Soil Sci. Soc. Am. Proc. 32:101–104.

Warrick, A.W. 1998. Spatial variability, p. 655–676. *In* D. Hillel (ed.) Environmental soil physics. Academic Press, San Diego, CA.

Weiler, M., and F. Naef. 2003. Simulating surface and subsurface initiation of macropore flow. J. Hydrol. 273:139–154.

Wilcox, B.P., W.J. Rawls, D.L. Brakensiek, and J.R. Wight. 1990. Predicting runoff from rangeland catchments: A comparison of two models. Water Resour. Res. 26:2401–2410.

Wilding, L.P., and H.S. Lin. 2006. Advancing the frontiers of soil science towards a geoscience. Geoderma 131:257–274.

Wilding, L.P., and L.R. Drees. 1983. Spatial variability and pedology, p. 83–117. *In* L.P. Wilding, N.E. Smeck, and G.F. Hall (eds.) Pedogenesis and soil taxonomy. I. Concepts and interactions. 1st edn. Elsevier, Amsterdam, the Netherlands.

Wilding, L.P., J. Bouma, and D. Goss. 1994. Impact of spatial variability on modeling, p. 61–75. *In* R. Bryant and M.R. Hoosbeek (eds.) Quantitative modeling of soil forming processes. SSSA Special Publication No. 39. SSSA, Madison, WI.

Wilding, L.P. 2000. Pedology, p. E1–E4. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Winzeler, H.E., P.R. Owens, B.C. Joern, J.J. Camberato, B.D. Lee, D.E. Anderson, and D.R. Smith. 2008. Potassium fertility and terrain attributes in a Fragiudalf drainage catena. Soil Sci. Soc. Am. J. 72:1311–1320.

Wood, E.F., K. Beven, M. Sivapalan, and L. Band. 1988. Effects of spatial variability and scale with implication to hydrology modeling. J. Hydrol. 102:29–47.

Wösten, J.H.M., A. Lilly, A. Nemes, and C. Le Bas. 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90:169–185.

Wösten, J.H.M., Ya.A. Pachepsky, and W.J. Rawls. 2001. Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics. J. Hydrol. 251:123–150.

Yakovleva, L.V. 1980. Change in the physical properties of weakly podzolic soils during gleying. Sov. Soil Sci. 3:331–338.

Zeleke, T.B., and B.C. Si. 2005. Scaling relationships between saturated hydraulic conductivity and soil physical properties. Soil Sci. Soc. Am. J. 69:1691–1702.

Zhu, A., Q. Feng, A. Moore, and J.E. Burt. 2010. Prediction of soil properties using fuzzy membership values. Geoderma 158:199–206.

Zhu, A.X., B. Hudson, J. Burt, K. Lubich, and D. Simonson. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Sci. Soc. Am. J. 65:1463–1472.

Zhu, A.X., L.E. Band, B. Dutton, and T.J. Nimlos. 1996. Automated soil inference under fuzzy logic. Ecol. Model. 90:123–145.

Zhu, A.X., L.E. Band, R. Vertessy, and B. Dutton. 1997. Derivation of soil properties using a soil land inference model (SoLIM). Soil Sci. Soc. Am. J. 61:523–533.

Zhu, A.X. 2000. Mapping soil landscape as spatial continua: The neural network approach. Water Resour. Res. 36:663–677.

Zobeck, T.M., and A. Richie. 1984. Relation of water table depth and soil morphology in two clay-rich soils of northwestern Ohio. OH J. Sci. OH Acad. Sci. 84:228–236.

36

Subaqueous Soils

Mark H. Stolt
University of Rhode Island

Martin C. Rabenhorst
University of Maryland

36.1 Introduction and Historical Development of Subaqueous Soil Concepts

One of the new frontiers in soil science that has come into focus over the past two decades has been the study of subaqueous soils. Although the concept has appeared in the literature at times (Kubiëna, 1953; Muckenhausen, 1965), only recently have these substrates received recognition in the United States and as such are now accommodated under the definition of soils (Soil Survey Staff, 1999). Under the new definition, soils may occur in permanently flooded or ponded environments with water depths up to approximately 2.5 m (Soil Survey Staff, 1999).

Previous opposition to the concept of subaqueous soils was primarily focused upon the idea that subaqueous substrates are not, in fact, soils but sediments. The ruling dogma was that by definition a soil must be able to support the growth of plants (Soil Survey Staff, 1975). Thus, the essential absence of higher plants in many subaqueous environments excluded these substrates from the pedologic realm. A secondary issue was related to the defining boundaries of a soil. The first edition of *Soil Taxonomy* (1975) stated that the upper limit of soils is "…. air or shallow water." In this sense, "shallow water" was meant to exclude soils permanently under water. Thus, these materials were also excluded from being soil by their permanent inundation. Over the 25 years that spanned the development of the second edition of *Soil Taxonomy* (Soil Survey Staff, 1999), pedological thinking continued to evolve such that pedologists began distancing themselves from their agricultural roots with a loosening of the link between the definition of soils and the growth of plants. Rather, pedologists began to emphasize what had already become deeply entrenched as the foundation to the taxonomy itself, namely, the formation of horizons resulting from those generalized pedogenic processes described by Simonson (1959). For example, Bockheim (1990, 1997) and Campbell and Claridge (1987) showed that in the harsh-cold climate of Antarctica, where higher plants are not able to grow, soil horizons were still observed as a result of pedogenic processes (i.e., additions, losses, transfers, and transformations). Thus, the idea that these areas should be recognized as soils was gaining support among the pedologic community, even though they were not necessarily capable of supporting the growth of higher plants.

Much of the credit for the emergence of subaqueous soils as a field of soil science has to be given to Dr. George P. Demas. The story goes that the concept formed in George's mind as he was standing on the edge of the marsh in Maryland that he was mapping, and looking down into the shallow tidal water of Sinepuxent Bay, he posed the question "Why should we stop mapping here?" He began to consider that such submersed aquatic vegetation as eelgrass (*Zostera marina*) and widgeongrass (*Ruppia maritima*) was rooted in these substrates (Figure 36.1) and as he began to closely examine them observed what could be construed as pedogenic horizons. Soon afterward he published his paper "Submerged Soils: A New Frontier in Soil Survey" in *Soil Survey Horizons* (Demas, 1993). Over the next 6 years under the guidance of Dr. Martin Rabenhorst, Demas further developed the ideas and concepts for the characterization, formation, and mapping of subaqueous soils in his PhD dissertation (Demas, 1998).

The works of Demas led to a number of additional studies that form the basis for most of the discussion in this chapter.

FIGURE 36.1 Eelgrass meadow growing on a shoal in 1.5 m of water. The soil supporting the eelgrass is a sandy skeletal Typic Haplowassent. (Photo Jim Turenne.)

In particular, the most important findings or accomplishments included in the work and dissertation were as follows:

- Subaqueous soils form as a result of the generalized processes of additions, losses, transfers, and transformations (Demas and Rabenhorst, 1999). This led to a change in the definition of soil to include substrates that are permanently under significant water (approximately 2.5 m) and that show evidence of pedogenesis (Soil Survey Staff, 1999).
- Bathymetric maps could be constructed to use as a soil survey base map and identify subaqueous landforms in a manner analogous to the subaerial landforms that soil scientists had been studying for most of the past century (Demas and Rabenhorst, 1998).
- Similar soils occurred or formed on similar landforms. Therefore, the "soil–landscape" paradigm typically used to map subaerial environments could be applied to the mapping subaqueous soils, and, thus, specific soil–landscape relationships began to be documented for the coastal lagoons of the Mid-Atlantic United States.
- The establishment of the first official soil series for subaqueous soils: Demas,* Sinepuxent, Southpoint, Tizzard, Trappe, and Whittington.

Kubiëna (1953) was the first to use the term "subaqueous soils" to describe permanently inundated soils. Those soils composed of layers and forming in low-energy subaqueous environs were classified into four groups (Kubiëna, 1953). Most of the focus was on soils having considerable soil organic matter: dy, gyttja, and sapropel; terms often applied to substrates in limnological studies (Saarse, 1990). Horizon sequences were typically A, AC, and C regardless of the soil type. Dy soils formed below water columns that were acidic, nutrient poor, and having high concentrations of soluble organic compounds. These soil materials have a gel-like form indicative of amorphous organic matter. Gyttja forms in subaqueous soils rich in nutrients. The majority of the materials are coprogenic in origin, having a loose arrangement

(Jongerius and Rutherford, 1979) typical of high n-value soil materials (low bearing capacity; Soil Survey Staff, 2010). These subaqueous soil materials have also been referred to as sedimentary peat, coprogenous earth, or limnic materials (see Fox, 1985; Soil Survey Staff, 2010). Subaqueous layers that contain various amounts of more or less unrecognizable organic debris that are enriched in sulfides were termed sapropels. Most of these sulfides are Fe-monosulfides or pyrite in the solid form or hydrogen sulfide gas as recognized by the rotten egg smell. Colors are typically black that changes to gray upon drying.

The classification of subaqueous soils by Kubiëna (1953) was cited in the national soil classification system in Germany (Muckenhausen, 1965), but there is no evidence that these soils were the focus of any serious investigation. A decade later, although he did not elaborate nor focus much on what he called "subaquatic soils," Ponnamperuma (1972) did affirm that these soils forming under water reflected "horizon differentiation distinct from sedimentation." Nevertheless, between the publication of this paper and the time when Demas began to focus on these systems more than two decades later, apparently little attention was paid to subaqueous soils.

36.2 Soil Genesis in Subaqueous Environments

In addition to the generalized model of soil genesis (additions, losses, transfers, and transformations) described by Simonson (1959), pedologists have often invoked the state factor equation of Jenny (1941) to describe and conceptually model the formation of soils. While Jenny's model has limitations, it recognizes the contributions of various soil-forming "factors." In considering the genesis of subaqueous soils, similarities to the processes and factors described by Jenny (1941) were recognized, but significant differences were also noted. The generalized model for estuarine sediments of Folger (1972) was noted where he described their origin as being derived from source geology (G), bathymetry (B), and hydrologic condition (H) (flow regime).

$$S_e = f(G, H, B) \qquad (36.1)$$

The concepts of both of these previous equations were joined, with some further modifications, into a state factor equation to describe the formation of subaqueous soils (Demas and Rabenhorst, 2001).

$$S_s = f(C, O, B, F, P, T, W, E) \qquad (36.2)$$

where
    S_s is subaqueous soil
    C is the climatic regime
    O is organisms
    B is bathymetry
    F is flow regime
    P is parent material
    T is time
    W is water column attributes
    E is catastrophic events

---

* Originally proposed as Wallops but posthumously named Demas following the untimely death of innovator Dr. George P. Demas in December 1999.

Climatic regime (C) was not included in Folger's equation and does not include precipitation as in Jenny's model. The climatic component in this model primarily represents temperature. Temperature, for example, will affect the rate of organic matter decomposition (and other biogeochemical reactions).

Organisms (O) was also not included by Folger and represents the role that biota play in subaqueous pedogenesis. As an example, the burrowing of benthic organisms (essentially irrigating their burrows with oxygenated water) often contribute to the development of light-colored, surface horizons, as well as the obvious contributions of plant carbon to the upper soil horizons.

Bathymetry and flow regime (B and F) replace relief (R) in Jenny's equation. The catena concept per se is not applicable in a permanently submersed environment because relief or topograph do not play the same role as in subaerial soils. Bathymetry contributes to the effects of internal and wind-generated waves on the subaqueous soil surface. Flow regime helps to shape underwater topography and accounts for differences in the energies associated with currents and tides. Together, these two factors (B, F) essentially play the same genetic role as relief does in subaerial soil environments.

Parent material (P) was a factor in both the equations of Folger and Jenny and explains the effect of the source material on subaqueous soil profile attributes. For example, subaqueous soils that form in areas where they receive barrier island washover materials are predictably sandy textured.

Time (T) of course represents the amount of time available for the expression of subaqueous soil attributes.

Water column attributes (W) was not included in either Jenny's or Folger's equations and has been added to include variations in the chemical composition of the water column that could have an impact on subaqueous soil characteristics. Those subaqueous soil profiles developed in freshwater regions or fresh portions of estuaries will likely be significantly different than those formed in more saline or brackish environments where sulfate is available for reduction to sulfide and the potential formation of solid phase sulfide minerals. Similarly, the dissolved oxygen levels in the water column could dramatically impact the formation or the thickness of light-colored, oxidized, surface horizons.

Catastrophic events (E) is included in this equation to account for the possibility that subaqueous soil profiles may be dramatically impacted by major storm events or other uncontrollable or unknown factors. The effects of storms or modest hurricanes, however, do not seem to cause wholesale alterations to large areas of subaqueous soils.

36.3 Mapping, Research, and Agency Efforts

Subaqueous soils research and mapping projects have been completed or are in progress along the eastern seaboard of the United States from Maine to Florida and along the Gulf Coast in Florida and Texas (Table 36.1). Projects have covered a range of topics including mapping protocols, soil–landscape relationships, carbon sequestration, soil variability, pedogenesis, use and

TABLE 36.1 Summary of Current or Completed Subaqueous Soils Projects

| Investigators	Affiliation	Location	Project Focus	Publications
Demas	UMD/NRCS	Sinepuxent Bay, MD	Soil survey methods and pedogenesis	Dissertation, Demas (1993); Demas et al. (1996); and Demas and Rabenhorst (1998, 1999, 2001)
Bradley	URI	Ninigret Pond, RI	Soil survey and eelgrass methods	Thesis, Bradley and Stolt (2002, 2003, 2006)
Flannagan	UME	Taunton Bay, ME	Soil survey	Thesis, Osher and Flannagan (2007)
Jespersen	UME	Taunton Bay, ME	Carbon accounting	Thesis, Jespersen and Osher (2007)
Angell	UMA	Freshmeadow Pond, MA	Soil survey	Report
Ellis	UFL	Cedar Key, FL	Soil survey	Dissertation
Fischler	UFL	Indian River Inlet, FL	SAV	Thesis
Casby-Horton/Brezina	NRCS	Padre Island, TX	Soil survey and ecological site descriptions	—
Payne	URI	Greenwich Bay, RI	Water quality methods	Thesis
Coppick	UMD	Rehoboth Bay, DE	Soil survey	Thesis in progress
Balduff	UMD	Chincoteague Bay, MD	Soil survey methods	Dissertation
Keirstead/Hundly	NRCS	Little Bay, NH	Soil survey	—
Surabian/Parizek/McVey	NRCS	Little Narragansett Bay, CT, RI	Soil survey and mooring interpretations	Report, Surabian (2007)
MapCoast	MapCoast	Rhode Island estuaries	Soil survey methods	Web available data
Salisbury	URI	Quonochontaug Pond, RI	Shellfish and dredging interpretations	Thesis in progress
Pruett	URI	Point Judith Pond, RI	Eelgrass and carbon accounting	Thesis in progress
Wong	NCSU/NRCS	Jamaica Bay, NY	Soil survey and eelgrass	Thesis
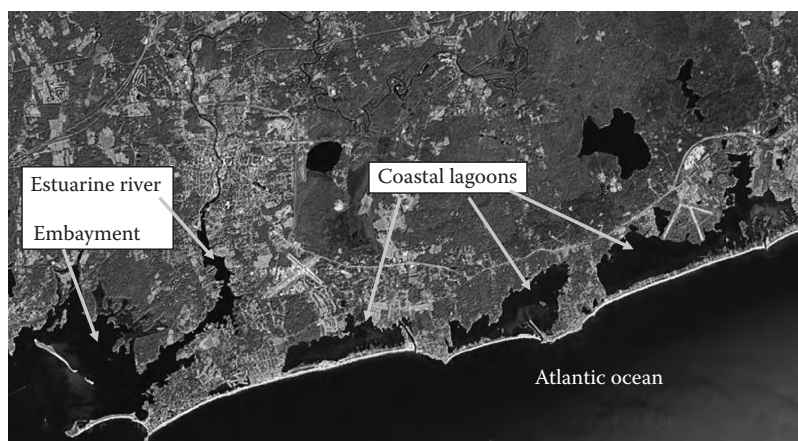
FIGURE 36.2 Examples of embayment, coastal lagoon, and estuarine river systems typically examined in subaqueous soil studies. Water depths generally average less than 2 m at low tide. Image was taken from a 2009 Google map of the southern Rhode Island and Connecticut shoreline.

management interpretations, and relationships between subaqueous soils and submerged aquatic vegetation (SAV) and water quality. Study and mapping areas have primarily concentrated on estuarine areas such as coastal lagoons and shallow water embayments (Figure 36.2). In response to mapping efforts, regional and national subcommittees within the National Cooperative Soil Survey have been established to advance national mapping standards and procedures for subaqueous soils. New parent material and landscape–landform terms have been added to the *National Soil Survey Handbook* (NCSS, 2005), and *Soil Taxonomy* has been revised to accommodate the classification of subaqueous soils. In Rhode Island, a partnership was developed (MapCoast, 2009) among Federal agencies such as US. Environmental Protection Agency (USEPA), USDA-Natural Resources Conservation Service (NRCS), National Oceanic and Atmospheric Administration (NOAA), state-level agencies, and university researchers and scientists to apply the information provided in subaqueous soil investigations to coastal resource issues and problems. The MapCoast partnership is a consortium dedicated to multidisciplinary mapping of coastal underwater resources, including bathymetry, habitat, geology, soils/sediment, and archeological resources in shallow waters (August and Costa-Pierce, 2007). NOAA is redesigning their classification system for shallow subtidal habitats (similar to the Cowardin et al. (1979) system for wetlands) and including a subaqueous soils component (Madden et al., 2008).

36.4 Methods for Characterizing Subaqueous Soils and Mapping Their Distribution

One of the big hurdles in investigations of subaqueous soils and mapping their distribution was the lack of methodologies for identifying, sampling, characterizing, and mapping subaqueous soil properties and their distribution. Discussions on topics such as collecting bathymetric data, using a vibracore, collecting soil samples under water, and handling subaqueous soil samples for subsequent laboratory analysis were essentially absent from the soil literature.

36.4.1 Creating a Subaqueous Terrain Map

Landscape units provide a first approximation of the distribution of soils on the landscape and offer an objective delineation of soil types. Numerous studies have emphasized the importance of landscape components for predicting and explaining soil distributions (Jenny, 1941; Ruhe, 1960; Huddleston and Riecken, 1973; Stolt et al., 1993; Scull et al., 2005). Subaqueous landscapes are fundamentally the same as subaerial systems and have a discernable topography from which subaqueous landforms and landscape units may be identified. However, because of the overlying water, submerged landscapes and landforms cannot be identified easily using standard methods such as stereo photography or visual assessment of the landscape. Therefore, identification and delineation of subaqueous landscape units is somewhat more complicated than that of terrestrial landscapes, and development of subaqueous topographic maps is a critical first step toward delineating subaqueous landscape units.

Bathymetric data (water depths) are used to produce a contour or bathymetric map from which subaqueous landforms can be identified and delineated to begin the soil survey. Thus, x, y, and z coordinates are necessary to create a contour map. XY locations are obtained with a differential GPS receiver (DGPS) with submeter accuracy.

Water depths can be determined in a number of ways. The quickest approach is to use a fathometer or bottom profiler. The transducer portion of the profiler is attached to the boat just below the water line. As the boat moves along the water, depths are obtained and stored in the fathometer computer or a device such as laptop. Demas and Rabenhorst (1998) reported that soundings were collected at approximately $4 \, km^2 \, h^{-1}$. Soundings should be collected essentially along fairly evenly spaced transects that are perpendicular to the shoreline. The depths are corrected by adding the depth between the water surface and bottom of the transducer and correcting for changes in the tide while the data are being collected.

Tide corrections are made from data collected from tide gauges operating at the same time that the bathymetric data are being

collected. One to three tide gauges are generally required depending upon the size of the area of interest and the complexity of the tidal cycle within the estuary. Tide gauges should be surveyed in from United States Geological Survey (USGS) benchmarks. Tide corrections can be made in a number of ways. Most simple tidal fluctuations can be corrected using equations developed from the tidal cycles and applied via a spreadsheet. Complicated corrections may require the use of software designed for the purpose.

There has been some discussion and some attempts to use LIDAR (light detection and ranging) to obtain bathymetric data more rapidly. The SHOALS (U.S. Army Corps of Engineers) bathymetry system uses a scanning, pulsed, infrared (1064 nm), and blue–green (532 nm) laser transmitter where the depth of water is determined from the difference in return times of the two beams and knowing the speed of light in water. Optimally, this system can be used to measure water depth from 0 to 40 m with a vertical accuracy of 20 cm. While this may have promise, there are a number of limitations that can be especially problematic including the following: (1) water clarity limits the ability of light to penetrate to the bottom; (2) high surface waves and heavy fog decrease the depth penetration of the lasers; and (3) heavy vegetation and fluid mud influence the depth penetration of the lasers. Also, in many systems where the maximum water depth is only a few meters, a resolution of 20 cm may not be adequate. Hopefully, advances will continue, so that more rapid acquisition of bathymetry becomes possible.

Bathymetric data should be reviewed to remove aberrant depths and aberrant XY locations. A number of software programs are available to construct topographic maps. As an example, an ArcView TIN model was created using the bathymetric soundings and a hard break line (depth = 0) consisting of the wetline from recent orthophoto. The TIN was converted to a GRID (10 m pixel size). The land–water interface observed from the orthophoto wetline was used as a mask to set all land-based pixels to NODATA. The bathymetric GRID was smoothed by applying a 3 pixel radius averaging filter, and contours were created from the smoothed bathymetric GRID. Although the TIN model was used in our example here, other modeling approaches such as kriging have been applied to bathymetric data to create topographic maps.

Using the fathometer from a boat is limited to water depths greater than 50 cm. In areas where there is considerable tidal fluctuation (a meter or more tidal fluctuation), shallow water may be profiled at high tides. If tidal fluctuations are less, surveying of the shallow water may be necessary. This can be done with a survey grade GPS that records elevations real-time kinetic (RTK), a total station, or an all-purpose elevation rod and level. This approach can also be used to validate contour maps created from bathymetric data.

36.4.2 Landscape and Soil Delineation

Landscape unit boundaries provide the first approximation of the distribution of soils over the landscape. Landscape attributes such as slope, microrelief, surface shape, and geographic proximity to other features, and location help define landforms and landscape unit boundaries. Landscape unit boundaries are hidden beneath the water cover in the subaqueous environment and need to be deciphered from contour maps. Landforms and units such as coves, submerged beaches, shoals, and washover fan flats and slopes are some common examples found in many estuaries (Figure 36.3; NCSS, 2005). In some cases, these features can be observed in aerial photographs that penetrate the water, but in general, a contour map illustrating slope breaks is necessary to define the boundaries on each unit.

Identifying the soil types within a landscape unit is done through reconnaissance efforts and transects. The same criteria used to separate mapping units in subaerial soils can be used in subaqueous soils. Tools such as a Macaulay peat sampler, bucket auger, and tile probe are effective in providing soil samples and data for identifying soil types. Peat samplers work well in high n-value (soft, low-density) soil materials, low-energy environments. In areas where low n-value mineral soils dominate, a bucket auger can be used to sample the upper 75 cm of the soil. Some soil scientists use a sleeve with an inside diameter slightly larger than the diameter of the auger to maintain an auger boring location and to minimize slumping of sandy materials into the auger hole. Samples collected in the bucket auger are often transferred to a vinyl tray or gutter on the boat for description and possible sampling. A floating tube with the tray strapped to the tube can work in the case of wading and mapping. Tile probes can be used to find depth to bedrock or similar consolidated or semiconsolidated materials. Percentages of boulders and large stones can also be estimated with the tile probe. Side-scan sonar images, or video footage across proposed mapping units, may aid in proper placement of boundaries and provide spatial data regarding the distribution of stones and boulders extruding from the soil and into the water column.

Just as in subaerial soil mapping, map unit purity and variability need to be addressed. The most common approaches are using random points or transects to assess variability. Studies of soil variability within the landscape units (Demas, 1998; Bradley and Stolt, 2003; Osher and Flannagan, 2007) demonstrate that the concept (common to subaerial landscapes) that soil type follows landscape form (Hudson, 1992) also holds for subaqueous soils (Table 36.2). For example, Bradley and Stolt (2003) reported that 11 of the 12 map units that were used to map the subaqueous soils in a coastal lagoon had a taxonomic purity (based on the subgroup taxonomic level) above criteria used for delineation of the traditional consociation map units used in most USDA-NRCS soil surveys (Soil Survey Division Staff, 1993).

One of the criticisms of inventorying subaqueous soils is that these environments are considered "ever-changing, unstable, shifting sands." Although some areas such as flood–tidal delta landscapes are quite dynamic, Bradley and Stolt (2002) showed that a detailed 1950s NOAA bathymetric map (NGDC, 1996) of the coastal lagoon (Ninigret Pond) was essentially no different than a bathymetric map created 50 years later. With the lifespan of a soil survey on the order of 25–30 years, subaqueous soil surveys should provide descriptions and interpretations for two to three decades of most areas having subaqueous soils (Table 36.3).
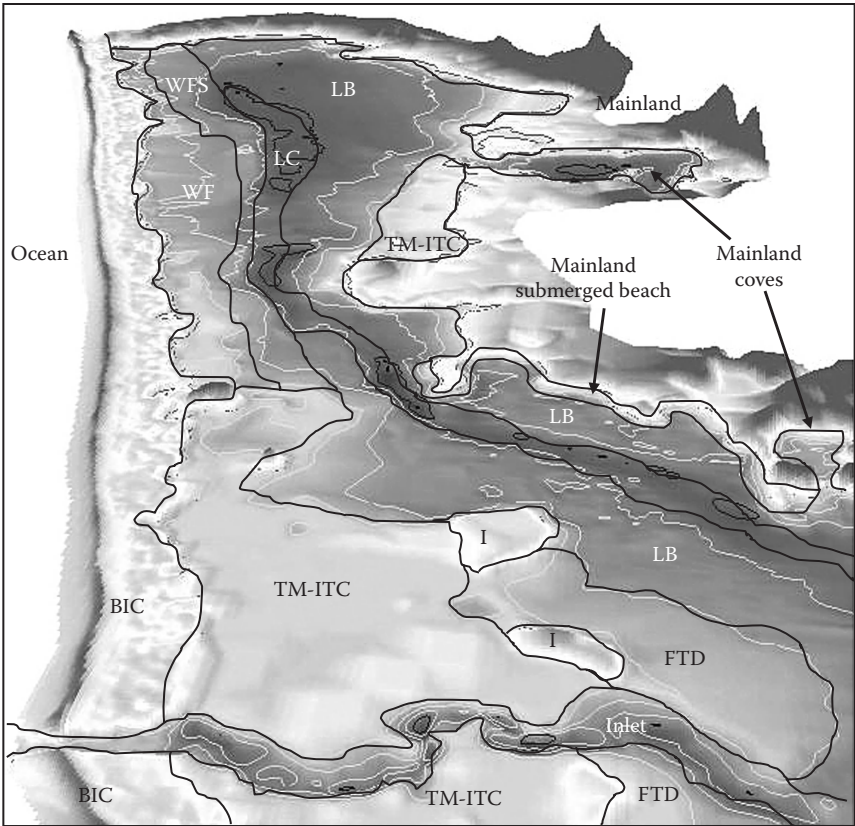
FIGURE 36.3 Examples of landscape units within a coastal lagoon. The barrier island complex (BIC), islands (I), and mainland are subaerial system. The tidal marsh-intertidal complex is sometimes subaqueous and sometimes subaerial. The flood-tidal delta (FTD), lagoon bottom (LB), washover fan (WF), washover-fan slope (WFS), lagoon channel (LC), mainland cove, and mainland submerged beach landscape units are subaqueous systems. The inlet brings in water from the open ocean on high tides and flushes the lagoon during out-going tide. Contour lines are for the subaqueous areas only have a 50 cm interval.

TABLE 36.2 Examples of Landscape Unit, Parent Material, and Soil Type Relationships from a Rhode Island Coastal Lagoon

Landscape Unit	Parent Materials	Typical Soil Subgroup Classification[a]
Lagoon bottom	Silt, fine sand, and organic material	Typic Sulfiwassents
Washover fan flat	Holocene sand	Typic Fluviwassents
Flood–tidal delta flat	Holocene sand	Typic Psammowassents
Washover fan slope	Holocene sand	Sulfic Fluviwassents
Flood–tidal delta slope	Holocene sand	Fluventic Psammowassents
Mainland submerged beach	Glacial fluvial sand and gravel	Typic Haplowassents
Barrier cove	Silt, fine sand, and organic material over glacial fluvial sand and gravel or Holocene sand	Typic Sulfiwassent
Mainland shallow cove	Holocene sand over glacial fluvial sand and gravel	Haplic Sulfiwassents
Midlagoon channel	Glacial fluvial sand and gravel	Typic Haplowassents
Barrier submerged beach	Glacial fluvial sand and gravel	Typic Haplowasssents
Shoal	Glacial fluvial sand and gravel	Typic Haplowassents
Mainland cove	Silts, fine sand, and organic material over buried organic material	Thapto-histic Sulfiwassents

Source: After Bradley, M.P., and M.H. Stolt. 2003. Subaqueous soil-landscape relationships in a Rhode Island estuary. Soil Sci. Soc. Am. J. 67:1487–1495.

[a] Classification based on Keys to Soil Taxonomy (Soil Survey Staff, 2010).

TABLE 36.3 Summary of Selected Subaqueous Soils Interpretations Identified by Federal and Regional Resource Managers for Managing Shallow Subtidal Coastal Areas in the New England (MapCoast, 2009) and Mid-Atlantic States (King, 2004)

Specific resource-based soil interpretation
Submerged aquatic vegetation restoration
Crab habitat
Shellfish stocking
Sustainable shellfish production
Mooring and dock locations
Identifying anthropogenic sites
Nutrient reduction
Pfesteria cyst residence sites
Benthic preservation site identification
Wildlife management
Waterfowl, nurseries, and spawning areas
Habitat protection for horseshoe crab
Tidal marsh protection and creation
Bathymetric maps and navigation
Dredging island creation
Effects of dredging on Benthic ecology
Dune and beach maintenance/replenishment

Sources: Mapping Partnership for Coastal Soils and Sediments. 2009. Soil survey data for Ninigret Pond. http://www.ci.uri.edu/projects/mapcoast/data/default.htm; accessed December 17, 2009; King, P. 2004. Subaqueous soil interpretations. Northeast Region National Cooperative Soil Survey Subaqueous Soils Working Group. *In* M.C. Rabenhorst, M.H. Stolt, P. King, and L. Osher (eds.) National workshop on subaqueous soils. Georgetown, DE.

36.4.3 Sample Collection for Characterization of Typifying Pedons

In high n-value materials, a relatively undisturbed half-core (5 cm diameter) can be collected using a Macaulay peat sampler, which is an excellent tool for providing samples for description and characterization. One limitation is that samples collected with a peat sampler are a bit small, which when coupled with the fact that the materials that can be sampled with the peat sampler have a low density, their dry mass is quite small. Thus, for characterization and descriptive purposes it may be necessary to take multiple adjacent samples if a Macaulay peat sampler is employed. High n-value materials can also be collected in a core barrel. A handle for pushing the core barrel in and pulling it out is attached to the barrel and weight (usually one or two persons lean on the handle) is added to push the core barrel into the soft materials. Several people are usually needed to pull the sample out.

Vibracore sampling is the most effective approach to obtain minimally disturbed samples for detailed description and sampling of typical pedons. A vibracore rig consists of an engine, a cable, and a vibracore head (Lanesky et al., 1979). The engine creates a high frequency, low amplitude vibration. The vibration is transferred through a cable to the vibracore head that is bolted to the top of core barrel or tube. This vibration essentially liquefies the soil materials in a thin zone immediately adjacent to the

tube, enabling the core barrel to penetrate into the soil materials. Weight is often added to the top of the core barrel to assist in pushing the core barrel into the soils. Vibracores come in a variety of forms from small, lightweight, and portable to large heavy machines.

Core barrels are generally made out of 7.5–10 cm inside diameter aluminum pipe (irrigation pipe). Some barrels are made out of polycarbonate (these are clear and light but also six to seven times more expensive than aluminum). Core barrel lengths should be as long as the soil to be sampled, plus water depth, and an extra 50 or 60 cm.

In many sampling systems, a core catcher is attached to the cutting end of the barrel. The catcher keeps the soil from sliding out of the barrel when the core is removed from the soil. Other systems rely on a rubber plug inside the barrel that rises up as the barrel is pushed into the soil. The plug maintains a short, nearly airtight space just above the sample to minimize disturbance. This approach also minimizes the suction of the sample out the bottom of the core when the core is removed.

A 2 m core of subaqueous soil that is collected in several meters of water is generally quite heavy and difficult to pull out of the soil. Thus, in most vibracore systems a tripod is set up with a winch or chainfall on the top to extract the core from the soil. The tripod is set up over an opening in the bottom of the sampling vessel. In most cases, a pontoon boat is used for sampling with a vibracore. Sampling is done through an opening (60 cm × 60 cm) in the deck ("moon pool"). The tripod setup can also be placed over a 2 m × 2 m barge that is towed behind the boat. An opening is cut in the middle of the barge. These barges are built from floating dock Styrofoam that is encased in marine plywood.

36.4.4 Subaqueous Soil Sampling and Description

Standard descriptive terminology as outlined in the *Soil Survey Manual* (1993), and by the National Soil Survey Lab (Schoeneberger et al., 2002), and horizon designations outlined in the *Keys to Soil Taxonomy* (2010) should be used to describe subaqueous soils. Samples collected with a Macaulay or bucket auger can be described and sampled immediately.

Cores collected with a vibracore can be sampled and described on the boat or returned to the lab and kept in cold storage prior to sampling. Keeping the samples in cold storage may not be adequate enough to inhibit sulfide oxidation entirely. Cores are extracted from the barrels by cutting the barrels lengthwise on both sides and removing the half barrel. Electric shears designed to cut metal are the best option. A circular saw can also be used to cut through most of the thickness of the barrels and a razor knife used to complete the cut. Prior to using the razor knife, the shards of aluminum or plastic should be whisked away.

36.4.5 Sample Analysis

Most sample analysis can be made following standard procedures outlined in the *Soil Survey Laboratory Methods Manual* (Burt, 2004). Certain soil properties will be affected by the

subaqueous environment, and laboratory procedures should be conducted with this in mind. Many subaqueous soil horizons (especially those collected from brackish or coastal environments) contain sulfides that may oxidize upon exposure to air. If samples are meant to be collected for classification purposes, treatment of the samples to avoid oxidation of the sulfides is critical. The most common approach is to immediately transfer the sample to a labeled bag, press out an air trapped in the bag, and put the sample on ice in a cooler. If the soil materials appear to be very reactive (unstable), or the amount of time between sampling and return to the lab is extended, pressurized nitrogen gas can be used to sparge the bags to remove oxygen prior to sealing. If deemed necessary, liquid nitrogen can be applied to the bag in the field to immediately freeze the sample after which it should be placed on ice.

Sulfides, salts, and shell fragments are the most important to consider when analyzing the soil and are worth noting. The presence of sulfides in the soils has been noted above and should be accommodated. Measurements of sulfides are not well documented in soil survey publications. Thus, if this characteristic is to be measured, consideration should be given to the numerous methods to measure chromium reducible and acid volatile sulfides (see Bradley and Stolt, 2006; Payne, 2007). Particle-size distribution analysis may need to be altered to accommodate for the weight and flocculation capability of salts. Samples can be washed to remove salts using dialysis tubing. Carbonates in shell fragments can be an issue in measuring organic carbon and should be considered when organic and calcium carbonate carbon is being determined.

36.5 Classification of Subaqueous Soils

Soil classification is much different than traditional sediment classification, where the substrate is termed mud, silty sand, muddy sand (Fegley, 2001), or other somewhat subjective class, and the focus is often on the upper portions of the profile. The soil classification approach offered in *Soil Taxonomy* is more comprehensive and addresses the larger soil profile. For example, a sediment classification of silty sand (e.g., see Fegley, 2001) could be better described using *Soil Taxonomy* as a coarse-silty over sandy skeletal, mixed, Typic Sulfaquent soil. This soil classification conveys that the upper portion of the soil has <18% clay and >70% silt-sized particles; the lower soil materials (to a meter depth or more) are sandy with >35% gravels or larger particles; the silt and sand-sized particles are not dominated by a particular mineral; and that there are enough sulfides within 50 cm of the soil surface that when the soil materials are allowed to oxidize the pH drops to below 4. Such additional knowledge could be important for decisions regarding the use and management of a portion of the estuary.

The latest version of *Keys to Soil Taxonomy* (Soil Survey Staff, 2010) includes taxa within Entisols and Histosols ("Wassents" and "Wassists") to accommodate subaqueous soils. The formative element "wass" is from the German word for water, "wasser" (Ditzler et al., 2008). Criterion for identifying both suborders is a

positive water pressure at the soil surface for at least 21 h each day in all years. The intent of the definition is to identify soils that are inundated everyday, every year, with no exceptions for periodic short- or long-term drought cycles. In certain areas with large tidal fluctuations, such as northern Maine in the United States, soils are inundated with 1–2 m of water everyday with the exception of a few hours at low tide. The 21 h minimum is proposed to allow for short daily exposure of these subaqueous soils.

Six great groups within Wassents are included keying out in the order: Frasiwassents, Psammowassents, Sulfiwassents, Hydrowassents, Fluviwassents, and Haplowassents (see Figure 36.4, e.g., profiles). Freshwater subaqueous soils key out as Frasiwassents based on an electrical conductivity of a 5:1 ratio of water to soil of <0.2 dS m^{-1}. Subaqueous soils that have sandy textures throughout the upper meter are Psammowassents. Sulfiwassents have at least 15 cm of sulfidic materials within the upper 50 cm of the soil. Soils with high n values (low bearing capacity) classify as Hydrowassents. Those soils with an irregular decrease in organic carbon with depth key out as Fluviwassents. Subgroup taxa include Sulfic, Lithic, Thapto-histic, Aeric, Psammentic, Fluventic, Grossic, Haplic, and Typic. With the exception of Grossic, all of these apply in a similar manner to previous applications in other taxa. Grossic is used to identify subaqueous soils that have very thick layers with high n values.
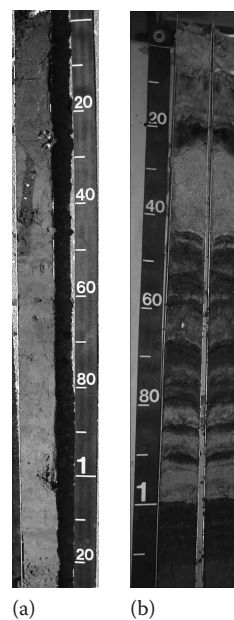

(a)   (b)

FIGURE 36.4 (a) Vibracore profile of a Haplic Sulfiwassent from a shallow mainland cove. The tape shows 10 cm increments. Note the clam krotovina from 30 to 50 cm (C/A horizon). The white pieces in the krotovina are shell fragments. At 120 cm, there is a change in parent material from the Holocene aged estuarine sediments to the Pleistocene outwash sand and gravels. (b) Vibracore profile of a Sulfic Fluviwassent. The soil was collected from a washover-fan slope. Note the many buried A horizons that represent storm surges. The buried A horizon starting at a meter represents the upper part of a previous Typic Sulfiwassent prior to encroachment of the washover fan over the lagoon bottom landscape unit.

Subaqueous Histosols are classified as Wassists. There are three great groups: Frasi-, Sulfi-, and Haplo-wassists. The Frasiwassists have low electrical conductivity (<0.2 dS m⁻¹); Sulfiwassists have >15 cm of sulfidic materials within 50 cm; and Haplowassists are all other Wassists. Three subgroups are proposed: Fibric, Sapric, and Typic, depending on the dominant type of organic materials present. Examples of Wassent–landscape relationships are shown in Table 36.2. The *World Reference Base* (WRB) FAO (2006) has taxa similar to those proposed for *Soil Taxonomy*. WRB's Subaquatic Fluvisols correlate to Wassents, and the Subaquatic Histosols correlate to Wassists.

36.6 Applications of Subaqueous Soil Information

Shallow-water coastal habitats, including coastal lagoons, shallow bays, and estuarine areas, are highly valued and heavily used resources. Almost two-thirds of the worldwide population currently lives in coastal areas (Trenhaile, 1997), and recent demographic studies suggest that in the next 25 years 75% of the U.S. population will live in close proximity to the coast (Bush, 2004). As subaqueous soil science progresses, a wide range of use and management interpretations are expected to be developed for use with estuarine subaqueous soil maps (August and Costa-Pierce, 2007; Surabian, 2007; Payne and Turenne, 2009). These interpretations will aid in coastal, estuarine, and marine restoration, ecosystem management, and conservation efforts. For example, subaqueous soils information can be used to assist in the restoration of SAV and shellfish; identifying shellfish aquaculture sites; design and placement of shoreline protection, docks, and moorings; and identifying subaqueous soils that are of beneficial use from dredging (Demas and Rabenhorst, 1999; Bradley and Stolt, 2003; Bradley and Stolt, 2006; Surabian, 2007). Since subaqueous soil investigations are a relatively new focus in pedology, and most of the subaqueous soil efforts have concentrated on developing field and laboratory methodology, few studies have concentrated on interpretations. Thus, the breadth of information relating soil type with the use and management of these resources is quite limited.

36.6.1 Dredging and Dredge Placement

Dredging of subaqueous soils is a common practice to deepen navigable waterways and to replenish beaches. Subaqueous soils often have layers or horizons where sulfides have accumulated in subaqueous soils as a result of sulfidization. When the sulfide-bearing soils are dredged and placed in the subaerial environment, sulfides oxidize releasing sulfuric acid, lowering the pH, and creating acid sulfate soils (Fanning and Fanning, 1989). Acid sulfate soils may persist for a number of years and are uninhabitable for plants and animals. If deposited near water, these acid sulfate soils can also create runoff that is toxic to aquatic systems (Demas et al., 2004).

To test for sulfides and for taxonomic purposes, subaqueous soils are allowed to oxidize in a moist environment. In general,

those soil materials that after at least 16 weeks of moist incubation reach pH values <4 are considered to have sulfidic materials (Soil Survey Staff, 2010) and are potential acid sulfate soils. Whether these soils reach the incubation pH of potential acid sulfate soils is dependent upon a number of factors such as the buffering capacity, rate and extent of acid production, weathering, and leaching due to environmental factors are not considered in the laboratory approach. For example, during a moist incubation, a small amount of sulfide would lower the pH in a sandy soil with low organic carbon because of the limited amount of buffering capacity. In such soils, the amount of acid produced would be much less than a soil with a similar incubation pH but a higher buffering capacity because of a finer texture or greater organic matter levels. In a natural setting, a small amount of acid could potentially leave a sandy, minimally buffered soil very quickly as the acidity generated by oxidation would be washed out of the system as a result of precipitation and leaching. In contrast, acid sulfate conditions may remain for decades in a fine-textured, buffered soil. Thus, understanding a number of soil parameters is critical to identifying the subaqueous soils that can be deposited in a subaerial environment as dredged materials and maintain conditions conducive for plant growth and a safe environment.

36.6.2 Water Quality

Estuarine ecosystem integrity and sustainability has received tremendous interest in recent years. These interests are being driven by concerns over the negative effects of rapid urbanization and related anthropogenic activities on the coastal environment. As the use of these natural resources has increased, the most urbanized estuaries have been ecologically compromised and common ecosystem functions and values are being lost. Obvious signs of these degraded environmental conditions are accumulations of metals and other contaminants, an increase in emerging diseases and algal blooms (Harvell et al., 1999), and anoxia related fin and shellfish kills (RIDEM, 1998, 2003). Most of these issues are related to poor water quality.

Water quality has traditionally been used in coastal areas as an indicator of the overall health of an estuary (Stevenson et al., 1993; Glasgow and Burkholder, 2000; Granger et al., 2000). Because water quality can fluctuate with tidal cycles and seasonal and yearly weather changes, water quality trends are difficult to predict or to use as a reliable indicator of extended changes in the health of an estuarine system (D'Avanzo and Kremer, 1994; Cicchetti et al., 2006). Soil properties and characteristics develop in response to the environment, making subaqueous soils a potential long-term indicator of the degree that these ecosystems have become degraded (Germano and Rhoads, 1988; Valente et al., 1992). Such an indicator would allow estuary management teams to target particular estuaries for conservation, protection, and restoration of resources based on soil survey information. Understanding the degree and spatial distribution of the degradation is critical to managing coastal estuaries for any number of functions and values, especially aquaculture and restoration of commercially important shellfish populations.

Redox conditions in a soil have important impacts on chemical processes that occur in the soil such as denitrification, changes in forms of iron, manganese, or sulfur, and the solubility of heavy metals (Tomaszek, 1995; Teasdale et al., 1998). The decomposition of organic matter by microbes fuels the redox reactions in soil. Oxygen is the strongest oxidizing agent in aqueous systems and acts as an electron acceptor during microbial decomposition. In subaqueous systems, however, oxygen can quickly be depleted and other electron acceptors are used by the microbes. These other electron acceptors include nitrate, manganese, iron, sulfate, and carbon dioxide. Each species, respectively, is reduced at a lower range of redox potentials depending on the pH (Bohn et al., 1979). These processes produce a vertical profile of decreasing redox potential with depth as each oxidizing agent is reduced until all organic matter has been decomposed (Teasdale et al., 1998).

The first of these boundaries, where all oxygen has been depleted or reduced, is generally known as the redox boundary, the redoxocline, or the redox potential discontinuity (RPD) (Knox, 1986; Teasdale et al., 1998; Hinchey and Schaffner, 2005). The depth of the RPD can be influenced by the grain size, organic matter content, temperature of the soil, as well as the movement and dissolved oxygen content of water above the soil surface (Knox, 1986). A redox potential gradient found in subaqueous soils often includes the oxidized surface layer where oxygen is still present in the interstitial water, a zone of transition where other species are being reduced, and a sulfide zone that is totally anaerobic, H_2S is prevalent, and redox potentials are very low (Knox, 1986). This zonation plays an important role in determining layers in which chemical processes involving organic carbon, nitrogen, and sulfur occur and can serve as an indication of estuary health.

36.6.3 Submerged Aquatic Vegetation

One part of the definition of soil is the ability of soil to support rooted plants in a natural environment (Soil Survey Staff, 1999). Dense beds of SAV (or seagrass) are often found in subtidal estuaries. Unlike macroalgal species, which anchor themselves to a substrate, seagrasses are rooted vascular aquatic plants in which roots serve both structural and nutrient uptake purposes (Barko et al., 1991). One of the most important interpretations from an inventory of subaqueous soils may be seagrass restoration. SAV such as eelgrass provides nursery habitat for economically important fin and shellfish and is important for sediment and nutrient filtering, nutrient cycling, and buffering wave effects. In many estuaries, aerial coverages of seagrass beds have severely declined over recent years. Therefore, seagrass restoration has become a focus of many coastal managers. Seagrass revegetation and restoration projects cost on the order of $100,000 per acre, but few of these projects have been successful. It is highly likely that the projects fail because of site selection. Seagrass revegetation sites are commonly located where past seagrass meadows had been, not at sites where present soil conditions are optimum for success. However, loss of seagrass tends to result in erosion of the subaqueous soils. Thus, soils within areas that previously supported seagrass may be significantly different following the loss of vegetation and subsequent erosion. A detailed knowledge of the relationship between subaqueous soil properties and SAV is essential for improving the success of seagrass revegetation efforts. An understanding of the seagrass–subaqueous soil system will help resource managers identify the sites where revegetation efforts can be most successful (Bradley, 2001). Few studies have looked at these relationships. Bradley and Stolt (2006) examined subaqueous soil–eelgrass relationships in a northeastern U.S. coastal lagoon. Similar studies need to be made across regions with a focus on the predominant seagrass species and the breadth of tidal ranges and gradient of temperatures.

36.6.4 Carbon Storage and Sequestration

With the concern with global warming mounting as a result of increasing greenhouse gas emissions, there is a significant interest in carbon storage and sequestration in soil systems. These interests have led to numerous studies focused on soil carbon for various land types and covers. Although forested and emergent wetlands have been well studied in regard to carbon sequestration, carbon sequestration and storage studies have largely overlooked estuary soils as important carbon sinks (Chmura et al., 2003; Thom et al., 2003). Considering that the shallow subtidal component may occupy as much as 90% of the estuary, these areas likely represent a significant and unaccounted for sink for carbon. Little is known, however, regarding the contribution of the shallow subtidal portions of the estuaries to the regional carbon stocks.

Geologic studies focused on estuarine and oceanic substrates have included organic carbon as a parameter inventory; however, most of these studies focus on surficial soil samples with the goal of understanding the origin and formation of petroleum (Hedges and Oades, 1997). Utilizing a pedologic approach, it is possible to quantify the organic carbon content of the subaqueous soil with depth, where it is actually stored, not just within the soil surface. Once soil organic carbon is determined for specific sites within an estuary, it will be possible to scale up to a regional or global scale in order to better determine the estuarine soils importance as a global carbon storage unit. Jespersen and Osher (2007) and Payne (2007) investigated the carbon storage capabilities of subaqueous soils in the Taunton Bay estuary in Maine and three embayments in Rhode Island, respectively. In both studies, a soil survey of the estuary was completed as a component of the study to relate organic carbon storage to soil–landscape unit. In addition, carbon pools to a depth of a meter in subaqueous soils were compared to their adjacent subaerial upland and wetland soils. The estuarine soil organic carbon pools were found to be equal to, and in some cases greater than, subaerial soil organic carbon pools. Payne (2007) reported higher energy, sandier soil–landscape units, such as shoals and shorefaces, had lower carbon pools than the lower energy soil–landscape units such as bay bottoms. Similar relationships were observed by Jespersen and Osher (2007).

The studies made by Jespersen and Osher (2007) and Payne (2007) were focused on northeastern tidal embayments. Little is

known regarding the expansive coastal lagoons of the Atlantic coast or Gulf of Mexico estuarine subaqueous soils. Carbon pools and sequestration rates in freshwater subaqueous soils are also unknown. Future studies should be designed and implemented to investigate these subaqueous soil systems.

36.6.5 Moorings and Docks

With any body of water there are typically structures built or deployed to secure boats. The foundation for these docks or mooring (permanent anchor that boats are secured to in a harbor) are the subaqueous soils. Thus, how well the mooring or dock functions is dependent upon subaqueous soil type. The bearing capacity or n value of the surface and near surface soils is one of the most important characteristics. Surabian (2007) examined relationships between subaqueous soils and moorings and found that mushroom anchors work best in high n-value soils. These moorings sink into the low bearing capacity soils and are kept in place by surface area and suction forces. Deadweight anchors are best suited for low n-value soils or soils dominated by coarse fragments (Surabian, 2007).

36.6.6 Shellfish

Subaqueous soils are critical to the structure and function of many of the plants and animals in the estuarine ecosystem and are the foundation for commercial shellfish production and aquaculture. Worldwide the aquaculture industry continues to develop and expand. Although the economics are difficult to quantify worldwide, the value of aquaculture products per acre typically far exceed those of traditional agriculture. For example, in 2007 the average value of Rhode Island aquaculture products (oysters and clams) was $32,000 per hectare (Alves, 2007). Considering the cash value of these aquaculture products, developing an understanding of the relationships between the submerged landscape, the subaqueous soils, and the growth and productivity of aquaculture species such as clams, oysters, scallops, and mussels is essential. To date, very little information is available regarding the relationships between shellfish productivity and subaqueous soil type.

The few works that have studied aquaculture–soil type relationships have focused on clams (Pratt, 1953; Pratt and Campbell, 1956; Wells, 1957; Grizzle and Lutz, 1989; Grizzle and Morin, 1989). These studies investigated clam abundance and shell-size growth rates with environmental factors such as soil type. Soils in these studies were fairly crudely characterized (i.e., sand or mud); however, most of the studies found a relationship between growth and particle size existed. In general, increases in fines (muds, silt, and clay) were associated with retardation in growth of clams. Grizzle and Lutz (1989) concluded that substrate type is important in some instances but seston flux (the amount of suspended particulate matter including plankton and organic detritus that passes by over a given period in the water column) is more important. Since sandy substrates typically have higher energies, the seston flux is often higher relative to finer textured soils.

Thus, current views on the shellfish–soil relationship are that the increased growth associated with sandier substrates in the earlier studies has been reinterpreted to be a secondary result of sandier soil being associated with higher current velocities (Rice and Pechenik, 1992). This suggests that subaqueous soil type may not directly relate to shellfish growth, but may serve as a surrogate for identifying areas of favorable seston fluxes, and could thereby be used to predict areas of the subtidal estuary with the highest potentials for shellfish growth. A better classification of the soil that would come with a subaqueous soil survey (i.e., better than sands and muds) may prove a better predictor of shellfish growth and provide delineations for the best locations for aquaculture of clams and oysters.

36.7 Future Considerations for Subaqueous Soils

Although subaqueous soils have received occasional mention in the literature for more than 50 years, only in the past decade or so have these soils been investigated with any intensity or focus. The limited number of investigations to date suggests that additional mapping, characterization, and research is needed to better understand these soils. In the United States, nearly all of the subaqueous soils projects have been conducted on the eastern seaboard in coastal waters. The same resource, habitat, and ecosystem service issues that have begun to be addressed from a pedological perspective in eastern U.S. estuaries also need to be examined in other shallow subtidal habitats as well as freshwater systems. As we have shown, the application of subaqueous soils investigations to addressing environmental and ecosystem questions related to restoration, aquaculture, carbon accounting, water quality, etc. is dependent upon an inventory of the subaqueous soil resources. The soil survey landscape-level models developed for mapping soils of embayments and lagoons need to be tested further in other Atlantic shallow subtidal habitats and then in other areas of the country and of the world. Concerted efforts should be made to conduct widespread subaqueous soil survey projects that are founded on established standards and protocols such as those used in the National Cooperative Soil Survey. These subaqueous soil resource inventories should be conducted and published at a scale that will be useful to resource managers attempting to balance both use and conservation of aquatic ecosystems that are heavily taxed and impacted as increasing populations choose to inhabit areas near the water.

References

Alves, D. 2007. Aquaculture in Rhode Island 2007 yearly status report. Coastal Resources Management Council, Wakefield, RI.

August, P., and B. Costa-Pierce. 2007. Mapping submerged habitats: A new frontier. 41 Degrees North 4:3.

Barko, J.W., D. Gunnison, and S.R. Carpenter. 1991. Sediment interactions with submersed macrophyte growth and community dynamics. Aquat. Bot. 41:41–65.

Bockheim, J.G. 1990. Soil development rates in the Transantarctic Mountains. Geoderma 47:59–77.

Bockheim, J.G. 1997. Properties and classification of cold desert soils from Antarctica. Soil Sci. Soc. Am. J. 64:224–231.

Bohn, H., B. McNeal, and G. O'Connor. 1979. Soil Chemistry. John Wiley and Sons, New York, NY.

Bradley, M.P. 2001. Subaqueous soils and subtidal wetlands in Rhode Island. M.S. Thesis. Department of Natural Resources Science, University of Rhode Island, Kingston, RI.

Bradley, M.P., and M.H. Stolt. 2002. Evaluating methods to create a base map for a subaqueous soil inventory. Soil Sci. 167:222–228.

Bradley, M.P., and M.H. Stolt. 2003. Subaqueous soil-landscape relationships in a Rhode Island estuary. Soil Sci. Soc. Am. J. 67:1487–1495.

Bradley, M.P., and M.H. Stolt. 2006. Landscape-level seagrass-sediment relations in a coastal lagoon. Aquat. Bot. 84:121–128.

Burt, R. (ed.). 2004. Soil survey laboratory methods manual. Soil Survey Investigations Report No. 42. USDA-NRCS. National Soil Survey Center, Lincoln, NE.

Bush, G.W. 2004. U.S. ocean action plan. United States Government Publication, Washington, DC.

Campbell, I.B., and C.G.C. Claridge. 1987. Antarctica: Soils, weathering processes and environment. Elsevier, New York.

Chmura, G.L., S.C. Anisfeld, D.R. Cahoon, and J.C. Lynch. 2003. Global carbon sequestration in tidal, saline wetland soils. Global Biogeochem. Cycles 17:1111.

Cicchetti, G., J.S. Latimer, S.A. Rego, W.G. Nelson, B.J. Bergen, and L.L. Coiro. 2006. Relationships between near-bottom dissolved oxygen and sediment profile camera measures. J. Mar. Syst. 62:124–141.

D'Avanzo, C.D., and J.N. Kremer. 1994. Diel oxygen dynamics and anoxic events in an eutrophic estuary of Waquoit Bay, Massachusetts. Estuaries 17:131–139.

Demas, G.P. 1993. Submerged soils: A new frontier in soil survey. Soil Survey Horiz. 34:44–46.

Demas, G.P. 1998. Subaqueous soils of Sinepuxent Bay, MD. Ph.D. Dissertation. University of Maryland. College Park, MD.

Demas, S.Y., A.M. Hall, D.S. Fanning, M.C. Rabenhorst, and E.K. Dzantor. 2004. Acid sulfate soils in dredged materials from tidal Pocomoke Sound in Somerset County, MD, USA. Aust. J. Soil Res. 42:537–545.

Demas, G.P., and M.C. Rabenhorst. 1998. Subaqueous soils: A resource inventory protocol. *In* Proc. 16th World Congress Soil Sci. August 20–26, 1998, Symposium 17 on CD. International Society of Soil Science, Montpelier, France.

Demas, G.P., and M.C. Rabenhorst. 1999. Subaqueous soils: Pedogenesis in a submerged environment. Soil Sci. Soc. Am. J. 63:1250–1257.

Demas, G.P., and M.C. Rabenhorst. 2001. Factors of subaqueous soil formation: A system of quantitative pedology for submersed environments. Geoderma 102:189–204.

Demas, G.P., M.C. Rabenhorst, and J.C. Stevenson. 1996. Subaqueous soils: A pedological approach to the study of shallow-water habitats. Estuaries 19:229–237.

Ditzler, C., R.J. Ahrens, K. Hipple, M.C. Rabenhorst, and M.H. Stolt. 2008. Classification, mapping, and interpretation of subaqueous soils. Abstract: International conference and field workshops on soil classification. Santiago, Chile.

Fanning, D.S., and M.C.B. Fanning. 1989. Soil morphology, genesis, and classification. John Wiley & Sons, New York.

FAO. 2006. World reference base for soil resources 2006. World Soil Research Report No. 103. FAO, Rome, Italy.

Fegley, S.R. 2001. Demography and dynamics of hard clam populations, p. 383–422. *In* J.N. Kraeuter and M. Castagna (eds.) Biology of the hard clam. Elsevier Science, Amsterdam, the Netherlands.

Folger, D.W. 1972. Characteristics of estuarine sediments of the United States. Geol. Surv. Prof. Pap. 742. United States Department of the Interior, Washington, DC.

Fox, C.A. 1985. Micromorphological characterization of histosols, p. 85–104. *In* L.A. Douglas and M. Thompson (eds.) Soil micromorphology and soil classification. SSSA Special Publication No. 6. SSSA, Madison, WI.

Germano, J., and D. Rhoads.1988. Narragansett Bay sediment quality survey. Rhode Island State publication NBP-89-23. Rhode Island Department of Environmental Management, Providence, RI.

Glasgow, H.B., Jr., and J.M. Burkholder. 2000. Water quality trends and management implications from a five-year study of a eutrophic estuary. Ecol. Appl. 10:1024–1046.

Granger, S., M. Brush, B. Buckley, M. Traber, M. Richardson, and S. Nixon. 2000. Restoring water quality in Greenwich Bay: An assessment of eutrophication in Greenwich Bay. Graduate School of Oceanography. University of Rhode Island, Kingston, RI.

Grizzle, R.E., and R.A. Lutz. 1989. A statistical model relating horizontal seston fluxes and bottom sediment characteristics to growth of *Mercenaria mercenaria*. Mar. Biol. 102:95–105.

Grizzle, R.E., and P.J. Morin. 1989. Effects of tidal currents, seston, and bottom sediments on growth of *Mercenaria mercenaria*: Results of a field experiment. Mar. Biol. 102:85–93.

Harvell, C.D., K. Kim, J.M. Burkholder, R.R. Colwell, P.R. Epstein, D.J. Grimes, E.E. Hofmann et al. 1999. Emerging marine diseases—Climate links and anthropogenic factors. Science 285:1505–1510.

Hedges, J.I., and J.M. Oades. 1997. Comparative organic geochemistries of soils and marine sediments. Org. Geochem. 27:319–361.

Hinchey, E.K., and L.C. Schaffner. 2005. An evaluation of electrode insertion techniques for measurement of redox potential in estuarine sediments. Chemosphere 59:703–710.

Huddleston, J.H., and F.F. Riecken. 1973. Local soil map relationships in Iowa: Distributions of selected chemical and physical properties. Soil Sci. Soc. Am. Proc. 37:264–270.

Hudson, B.D. 1992. The soil survey as a paradigm-based science. Soil Sci. Soc. Am. J. 56:836–841.

Jenny, H. 1941. Factors of soil formation. McGraw-Hill, New York.

Jespersen, J.L., and L.J. Osher. 2007. Carbon storage in the soils of a mesotidal gulf of Maine estuary. Soil Sci. Soc. Am. J. 71:372–379.

Jongerius, A., and G.K. Rutherford. 1979. Glossary of soil micromorphology. Centre of Publication and Documentation, Wageningen, the Netherlands.

King, P. 2004. Subaqueous soil interpretations. Northeast Region National Cooperative Soil Survey Subaqueous Soils Working Group. *In* M.C. Rabenhorst, M.H. Stolt, P. King, and L. Osher (eds.) National workshop on subaqueous soils. Georgetown, DE.

Knox, G.A. 1986. Estuarine ecosystems: A systems approach. Vol. 1. CRC Press, Inc., Boca Raton, FL.

Kubiëna, W. 1953. The soils of Europe. Thomas Murby, London, U.K.

Lanesky, D.E., B.W. Logan, R.G. Brown, and A.C. Hine. 1979. A new approach to portable vibracoring underwater and on land. J. Sediment. Petrol. 49:654–657.

Madden, C., K. Goodin, B. Allee, M. Finkbeiner, and D. Bamford. 2008. Coastal and marine ecological classification standard. NOAA and NatureServe. http://webqa.csc.noaa.gov/benthic/cmecs/Version_III_Official_Review_Draft.doc; accessed December 17, 2009.

Mapping Partnership for Coastal Soils and Sediments. 2009. Soil survey data for Ninigret Pond. http://www.ci.uri.edu/projects/mapcoast/data/default.htm; accessed December 17, 2009.

Muckenhausen, E. 1965. The soil classification system of the Federal Republic of Germany. Pedologie (Belgium) Special Issue 3:57–74.

NCSS. National Cooperative Soil Survey. 2005. Glossary of terms for subaqueous soils, landscapes, landforms, and parent materials of estuaries and lagoons. Subaqueous Soils Committee Report, National Cooperative Soil Survey. http://nesoil.com/sas/Glossary-Subaqueous%20Soils.pdf; accessed July 30, 2009.

NCSS. National Cooperative Soil Survey. 2009. Amendments to soil taxonomy to accommodate subaqueous soils. Subaqueous Soils Committee Report, National Cooperative Soil Survey. http://nesoil.com/sas/Subaqueoustaxonomy.htm; accessed July 30, 2009.

NGDC. National Geophysical Data Center. 1996. National ocean service hydrographic survey data CD-ROM set. NOAA, Boulder, CO.

Osher, L.J., and C.T. Flannagan. 2007. Soil/landscape relationships in a mesotidal Maine estuary. Soil Sci. Soc. Am. J. 71:1323–1334.

Payne, M.K. 2007. Subaqueous soils and water quality. M.S. Thesis. Department of Natural Resources Science, University of Rhode Island, Kingston, RI.

Payne, M.K., and J. Turenne. 2009. Mapping the "new frontier" of soil survey: Rhode Island's MapCoast partnership. Soil Survey Horiz. 50:86–89.

Ponnamperuma, F.N. 1972. The chemistry of submerged soils. Adv. Agron. 24:29–96.

Pratt, D.M. 1953. Abundance and growth of *Venus mercenaria* and *Callocardia morrhuana* in relation to the character of the bottom sediments. J. Mar. Res. 7:60–74.

Pratt, D.M., and D.A. Campbell. 1956. Environmental factors affecting growth in *Venus mercenaria*. Limnol. Oceanogr. 1:2–17.

RIDEM. Rhode Island Department of Environmental Management. 1998. State of the State's Waters Rhode Island, A report to congress (PL 94–500, Section 305(b) Report). Supporting assessment data. Division of Water Resources, Providence, RI.

RIDEM. Rhode Island Department of Environmental Management. 2003. The Greenwich Bay fish kill—August 2003: Causes, impacts, and responses. Division of Water Resources, Providence, RI.

Rice, M.A., and J.A. Pechenik. 1992. A review of the factors influencing the growth of the northern quahog, *Mercenaria mercenaria* (Linnaeus, 1758). J. Shellfish Res. 11:279–287.

Ruhe, R.V. 1960. Elements of the soil landscape. Proc. 7th Int. Congr. Soil Sci. 23:165–168.

Saarse, L. 1990. Classification of lake basins and lacustrine deposits of Estonia. J. Paleolimnol. 3:1–12.

Schoeneberger, P.J., D.A. Wysocki, E.C. Benham, and W.D. Broderson (eds.) 2002. Field book for describing and sampling soils, version 2.0. USDA-NRCS, National Soil Survey Center, Lincoln, NE.

Scull, P., J. Franklin, and O.A. Chadwick. 2005. The application of classification tree analysis to soil type prediction in a desert landscape. Ecol. Appl. 181:1–15.

Simonson, R.W. 1959. Outline of a generalized theory of soil genesis. Soil Sci. Soc. Am. Proc. 23:152–156.

Soil Survey Division Staff. 1993. Soil survey manual. USDA-SCS. Agriculture Handbook 18. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1975. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. USDA-SCS Agriculture Handbook 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd Ed. USDA-SCS Agriculture Handbook 436. U.S. Government Printing Office, Washington, DC.

Soil Survey Staff. 2010. Keys to soil taxonomy. 11th Ed. USDA-NRCS. U.S. Government Printing Office, Washington, DC.

Stevenson, J.C., L.W. Staver, and K.W. Staver. 1993. Water quality associated with survival of submersed aquatic vegetation along an estuarine gradient. Estuaries 16:234–361.

Stolt, M.H., J.C. Baker, and T.W. Simpson. 1993. Soil–landscape relationships in Virginia: I. Soil variability and parent material uniformity. Soil Sci. Soc. Am. J. 137:172–176.

Surabian, D.A. 2007. Moorings: An interpretation from the coastal zone soil survey of Little Narragansett Bay, Connecticut and Rhode Island. Soil Survey Horiz. 48:90–92.

Teasdale, P.R., A.I. Minett, K. Dixon, T.W. Lewis, and G.E. Batley. 1998. Practical improvements for redox potential (E_H) measurements and the application of a multiple-electrode redox probe (MERP) for characterizing sediment in situ. Anal. Chim. Acta 367:201–213.

Thom, R.M., A.B. Borde, G.D. Williams, D.L. Woodruff, and J.A. Southard. 2003. Climate change and seagrasses: Climate-linked dynamics, carbon limitation and carbon sequestration. Gulf Mexico Sci. 21:134.

Tomaszek, J.A. 1995. Relationship between denitrification and redox potential in two sediment-water systems. Mar. Freshw. Res. 46:27–32.

Trenhaile, A.S. 1997. Coastal dynamics and landforms. Clarendon Press, Oxford, U.K.

Valente, R., D.C. Rhoads, J.D. Germano, and V. Cabelli. 1992. Mapping of benthic patterns in Narragansett Bay, RI. Estuaries 15:1–17.

Wells, H.W. 1957. Abundance of the hard clam *Mercenaria mercenaria* in relation to environmental factors. Ecology 38:123–128.

Digital Soil Mapping

Alex B. McBratney
The University of Sydney

Budiman Minasny
The University of Sydney

Robert A. MacMillan
International Soil Reference and Information Centre

Florence Carré
European Commission-Joint Research Centre

37.1 Introduction

The current explosion in computation and information technology comes with vast amounts of data and tools in all fields of endeavor. This has motivated numerous initiatives around the world to build spatial data infrastructures aiming to facilitate the collection, maintenance, dissemination, and use of spatial information. Soil science potentially contributes to the development of such generic spatial data infrastructure through the ongoing creation of regional, continental, and worldwide soil databases. The principal manifestation is soil resource assessment using geographic information systems (GISs), that is, the production of digital soil property and class maps with the constraint of limited relatively expensive fieldwork and subsequent laboratory analysis.

The production of digital soil maps ab initio, as opposed to digitized (existing) soil maps, is moving inexorably from the research phase (Skidmore et al., 1991; Favrot and Lagacherie, 1993; Moore et al., 1993) to production of maps for regions and catchments and whole countries. The map of the Murray–Darling basin of Australia (Bui and Moran, 2001) comprising some 19 million $250\,\text{m} \times 250\,\text{m}$ pixels or cells and the digital Soil Map of Hungary (Dobos et al., 2000) are the most notable examples. The progress and development of digital soil mapping is marked by adoption of new mapping tools and techniques, data management systems, innovative delivery of soil data, and methods to analyze, integrate, and visualize soil and environmental data sets (Grunwald, 2009).

McBratney et al. (2000) reviewed pedometric methods for soil survey and suggested three resolutions of interest, namely, $>2\,\text{km}$, $20\,\text{m}$–$2\,\text{km}$, and $<20\,\text{m}$, corresponding to national to global, catchment to landscape, and local extents. Table 37.1 provides an overview with five resolutions of interest. The third one (D3) that deals with subcatchments, catchments, and regions is the one that attracts the most attention. In the language of digital soil maps (Bishop et al., 2001), which differs from that of conventional cartography, scale is a difficult concept and is better replaced by resolution and spacing. D3 surveys, which in conventional terms have a scale of 1:20,000 down to 1:200,000, have a block or cell size from 20 to 200 m, a spacing also of 20–200 m, and a nominal spatial resolution of 40–400 m (see Table 37.1).

Unfortunately, the existing soil databases are neither exhaustive nor precise enough for promoting extensive and credible use of the soil information within the spatial data infrastructure that is being developed worldwide. The main reason is that their present capacities only allow the storage of data from conventional soil surveys, which are scarce and sporadically available.

The Netherlands has complete coverage at a nominal spatial resolution of 100 m. In France, on the other hand, a highly developed western economy, but with a large land area, only 26% of the country is covered at a nominal spatial resolution of 500 m and 13% at a nominal spatial resolution of 200 m (King et al., 1999). One third of Germany is covered with soil maps at a nominal spatial resolution of 10 m (1:10,000) but most of these are not yet digital (Behrens and Scholten, 2006). On the other hand, complete coverage of Germany at coarser resolutions (nominally 100 and 400 m) is available. It was projected that in 2010, 65% of Germany's soil maps will be digitized.

TABLE 37.1 Suggested Resolutions and Extents of Digital Soil Maps

Name	Approx. USDA Survey Order[a]	Nominal Resolution and Spacing (Pixel Size)	Resolution "loi du quart"[b]	Extent[c]	Cartographic Scale[d]
D1	0[e]	<(5 m × 5 m)	<(25 m × 25 m)	<(50 km × 50 km)	>1:5,000
D2	1, 2	(5 m × 5 m)–(20 m × 20 m)	(25 m × 25 m)–(100 m × 100 m)	(500 m × 500 m)–(200 km × 200 km)	1:5,000–1:20,000
D3	3, 4	(20 m × 20 m)–(200 m × 200 m)	(100 m × 100 m)–(1 km × 1 km)	(2 km × 2 km)–(2,000 km × 2,000 km)	1:20,000–1:20,0000
D4	5	(200 m × 200 m)–(2 km × 2 km)	(1 km × 1 km)–(10 km × 10 km)	(20 km × 20 km)–(20,000 km × 20,000 km)	1:200,000–1:2,000,000
D5	5	>(2 km × 2 km)	>(10 km × 10 km)	>(200 km × 200 km)	<1:2,000,000

[a] Soil Survey Staff (1993), Table 2.1.

[b] According to Boulaine (1980), the smallest area discernible on a map is 0.5 cm × 0.5 cm or one quarter of a square centimeter, hence the term "loi du quart." The USDA *Soil Survey Field Handbook* (Table 2.2) (Soil Survey Staff, 1993) quotes 0.6 cm × 0.6 cm. Both of these really refer to conventional map delineations, and resolution estimates based on these minimum areas should be regarded as very conservative.

[c] Calculated as minimum resolution times 100 (pixels) up to maximum resolution times 10,000 pixels.

[d] Digital soil maps are defined by their resolution and spacing—which here we equate with pixel size—the cartographic scale is calculated as 1 m/(side length of 1,000 pixels), assumes that the smallest area discernible is 1 mm × 1 mm. Conversely, the resolution (ρ) of a 1:100,000 conventional map can be calculated as, $\rho = 1/\chi \times \lambda = 100,000 \times 0.001 = 100$ m if we consider the smallest area resolvable on a map (λ), with representative fraction χ, to be 1 mm × 1 mm. It could be argued that the minimum resolution is the size of 2 × 2 pixels.

[e] This resolution was suggested by Dr. Pierre C. Robert, University of Minnesota, for applications in precision agriculture.

The situation in larger countries such as Australia and Brazil is much worse. In Australia, for example, prior to Moran and Bui's (2002) work, the Murray–Darling Basin, Australia's most important agricultural region comprising some 14% of the land area, had 50% coverage at 500 m and 3% at 200 m. In Brazil, the country is uniformly covered by the Soil Map of Brazil and the Agricultural Suitability Map of Brazil at a nominal spatial resolution of 10 km, exploratory soil maps by the RADAM/EMBRAPA Solos project (1:1,000,000 or nominally 2 km) and Agroecological Zoning (diagnosis of environmental and agro-socioeconomic features, nominally 4 km or 1:2,000,000).

The main reason for this lack of soil spatial data is simply that conventional soil survey methods are relatively slow and expensive. Furthermore, there is a worldwide crisis in collecting new field data in general that leads some to be very pessimistic about future developments in conventional soil surveying. To address this situation, current spatial soil information systems have to extend their functionalities from the storage and the use of digitized (existing) soil maps to the production of soil maps ab initio. This is the aim of digital soil mapping, which can be defined as the creation, and population of spatial soil information systems by the use of field and laboratory observational methods coupled with spatial and nonspatial soil inference systems (Lagacherie and McBratney, 2006).

Thus, digital soil mapping refers to the production of spatial soil databases, based on soil observations combined with environmental data through quantitative statistical relationships. In that sense, digital soil mapping is not simply digitizing existing soil maps and is more than just producing paper maps. There are other terminologies such as "predictive soil mapping" (Scull et al., 2003) that refers to the production of digital soil maps, and "environmental correlation" (McKenzie and Ryan, 1999) that is an aspect of the spatial soil prediction function.

Digital soil mapping uses a range of technologies allowing for more accurate and efficient prediction of soil properties through optimal sampling strategies, rapid analysis of soil properties, and rapid acquisition of environmental variables over a large extent. Combined with pedometric techniques, it can provide the best prediction of soil properties at the required resolution with associated uncertainties. Digital soil mapping can be thought as a means for modernizing and systematizing traditional soil survey.

In this chapter, we review various approaches with numerous examples from the literature, which are largely seen as special cases of the approach suggested here. First, we trace the development of the quantitative ideas and methods over the last 60 or so years.

37.2 Brief Review of Approaches to Soil Spatial Prediction

Hudson (1992) contended that soil survey is a scientific strategy based on the concepts of factors of soil formation coupled with soil–landscape relationships. Hewitt (1993) pleaded for the need for explicitly stated, but not necessarily quantitative, models for soil survey. Such models may be knowledge based (Bui, 2004). In this view, soil maps are representations of soil surveyors' knowledge about soil objects. In this chapter, we argue in favor of quantitative predictive models.

37.2.1 Jenny

Recalling Jenny's famous equation (Jenny, 1941), which he intended as a mechanistic model for soil development,

$$S = f(c, o, r, p, t, \ldots) \qquad (37.1)$$

implicitly

S stands for soil

c (sometimes *cl*) represents climate

o represents organisms including humans

r represents relief

p represents parent material

t represents time

some have asserted that it cannot be solved; nonetheless, since Jenny published his formulation, it has been used by innumerable surveyors all over the world as a qualitative list for understanding the factors that may be important for producing the soil pattern within a region. Numerous researchers have taken the quantitative path and have tried to formalize this equation largely through studies of cases where one factor varies and the rest are held constant. So quantitative climofunctions, topofunctions, etc. have been developed. Much of this work was done before sophisticated numerically intensive statistical methods became available. Here are some brief examples.

c (sometimes *cl*): Climofunctions were the ones most developed by Jenny in his 1941 book. Jones (1973) found relationships between carbon, nitrogen, and clay and annual rainfall and altitude in West African savanna using linear and multiple linear regression. Simonett (1960) found a power–function relationship between mineral composition of soil developed on basalt in Queensland and annual rainfall.

o: There seems less development of organofunctions, many believing that the principal organofunction or biofunction, that of vegetation, is dependent on soil rather than the converse. Noy-Meir (1974) found relationships between vegetation and soil type in S.E. Australia. The other principal organofunction, the anthropofunction, has only been recently quantified—much of the work on soil degradation and soil quality is evidence of the effect of humanity on soil. The classic work of Nye and Greenland (1960) is an early quantitative example.

r: The relationship between soil and topographic factors has been evident at least since Milne's paper (Milne, 1935). Quantitative topofunctions are manifold. For example, Furley (1968) and Anderson and Furley (1975) found a piecewise linear relationship between organic carbon, nitrogen, and pH of surface horizons and slope angle for profiles developed on calcareous parent materials around Oxford in England.

p: Quantitative lithofunctions have not been developed often, perhaps due to a difficulty in recognizing and quantifying the dependent and independent variables. Barshad (1958) quantified mean clay content as a function of rock type.

t: Some consider this the only truly independent variable (but if that is the case why is space not also included?). Chronofunctions are often theoretical or hypothetical rather than observed. Hay (1960), however, found an exponential relationship between clay formation and time for soil developed in volcanic ash on the island of St. Vincent, as would be expected from first-order kinetics.

A lot of this early quantitative work was very well summarized by Yaalon (1975). Many of the relationships found are not linear. It should be remembered that the aim of these investigations was to understand soil formation and not necessarily to predict soil from the other factors.

Recognition of interactions between the soil-forming factors is potentially important because it is one possible source of detailed soil pattern. It is difficult to find work that considers such interactions explicitly. Webster (1977) perhaps came closest with his canonical correlation studies of sets of soil properties and environmental factors. From this work, he suggested, for example, that soil will reflect a strong interaction between topography and lithology particularly on upper slope positions, but this will be time dependent. Odeh et al. (1991), using closely related methods, made similar findings.

37.2.2 Geographic and Neighborhood (or Purely Spatial) Approaches

Since the late 1960s, there has been an emphasis on what might be called geographic or purely spatial approaches, that is, soil attributes* can be predicted from spatial position largely by interpolating between locations of soil observation. Another way of thinking about this is as a *neighborhood law* expounded first perhaps by Lagacherie (1992), but is the basis underlying the soil combinations of Fridland (1972) and also of soil geostatistics (Giltrap, 1977), etc. Generally, we can consider the soil at some location (x, y) to depend on the geographic coordinates x, y and on the soil at neighboring locations $(x + u, y + v)$, that is,

$$s(x, y) = f((x, y), s(x + u, y + v)), \qquad (37.2)$$

the dependence usually being some decreasing function of the magnitude of u and/or v.

This approach arose originally out of the need for spatial prediction to make soil maps, and because of a failure to obtain prediction from the soil-forming factors largely because the quantitative variables describing these factors were not readily available to do such predictions. These purely spatial approaches are almost entirely based on geostatistics and its precursor trend-surface analysis, although thin-plate smoothing splines have been suggested and used occasionally (Laslett et al., 1987; Hutchinson and Gessler, 1994). Exact-fitting splines do not perform well (Laslett et al., 1987; Voltz and Webster, 1990).

37.2.2.1 Geostatistics and Related Methods Trend Surfaces: $s(x, y) = f(x, y)$

Trend surfaces are low-order polynomials of spatial coordinates. Several applications have been reported in the literature. Davies and Gamm (1969) applied this technique to soil pH values from the county of Kent in England. Edmonds and Campbell (1984) described the average annual soil temperatures at locations within a network of stations from Virginia and neighboring states with a third-degree polynomial that explained 71% of the observed variation. On the other hand, Kiss et al. (1988) found

---

* Soil attributes is a general term to mean that which can be attributed to the soil by measurement or inference, for example, soil properties like pH, or classes like a soil profile class, or the presence or the absence of a soil horizon class.

that the spatial pattern of ^{137}Cs activity in well-drained, native noneroded soil in the agricultural portion of Saskatchewan was complex and could not be adequately described by a second-order trend surface. There appears to be no literature on trend surfaces for soil classes; nevertheless, Wrigley (1978) has made an attempt to map the probability of the occurrence of soil classes. Spatially, trend surfaces are rather simplified "unnatural" representations, and more complex spatial patterns often need to be modeled.

$$\text{Kriging—}s(x, y) = f(s(x + u, y + v)). \tag{37.3}$$

It was recognized that more complex spatial patterns could be accommodated by treating soil variables as regionalized variables using the methods of geostatistics, particularly various forms of kriging. The papers by Burgess and Webster (1980a, 1980b; Webster and Burgess, 1980) are probably regarded as the most seminal. These kriging methods, reviewed by Burrough (1993), Goovaerts (1999), and Heuvelink and Webster (2001) could deal with continuous soil properties and classes, could give estimates for blocks or pixels of varying size, and moreover could estimate uncertainty.

$$\text{Co-kriging—}s(x, y) = f(s(x + u, y + v), \{c, o, r, p, t\}(x, y)). \tag{37.4}$$

It was recognized early in the development of soil geostatistics that soil could be better predicted if denser sets of secondary variables (spatially cross-) correlated with the primary variable were available; this technique is called co-kriging. In the early co-kriging studies (McBratney and Webster, 1983; Vauclin et al., 1983; Goulard and Voltz, 1992), these ancillary variables were other soil variables, indicating that other soil variables are themselves useful predictors of soil. Later, with the advent of GIS and improved technology, co-kriging was performed with detailed secondary data sets of environmental variables derived from digital elevation models and satellite images (Odeh et al., 1994, 1995).

37.2.2.2 Jenny and Geography—Corpt or Clorpt: $s(x, y) = f(\{c, o, r, p, t\}(x, y))$

An alternative spatial prediction strategy to the purely geographic approaches was developed in the early 1990s, although there were precursors. In these studies, the state-factor equation was put explicitly into a spatial framework, and the factors were also observed in the same spatial domain. This approach probably resulted from the advent of the first GISs, and also possibly as a pedological response to geostatistics. It seems to be based on a much earlier 1D example of using environmental (terrain, representing *r*) attributes for soil prediction, namely, that of Troeh (1964) and Walker et al. (1968). Probably the first of its kind, Aandahl (1948) quantitatively relates landscape attributes to soil properties. He derived the distribution of N based on slope length. Troeh (1964) analyzed the elevation data from two catenas and derived slope and profile curvature. He then plotted the slope gradient and profile curvature and found that the soil drainage classes could be distinguished by paraboloid of

revolution equations. Walker et al. (1968) used slope, curvature, aspect, and distance from the local summit, in combination with multiple linear regression to predict soil morphological properties such as A horizon depth, depth to mottling, and carbonates along a transect. An early, perhaps the first, 2D example is Legros and Bonneric (1979), based on earlier work by Legros (1975). They described a soil–environment relationship using various factors (altitude, slope, exposure, and parent material) that were observed on a 500 m grid-cell basis to predict the degree of podsolization in Massif du Pilat of France, and mapped it digitally at a resolution of 500 m. The prediction was achieved by a kind of taxonomic distance relative to reference sites. This was done well before the advent of formal GIS.

The GIS-based studies started at the beginning of the 1990s. Terrain analysis had improved and secondary rasterized layers, providing a kind of complete enumeration of the area could be put in GIS. The soil observation points were intersected with the layers of secondary data, a model fitted by various means, and then the model was used to predict all other locations on the raster. Moore et al. (1993) gave the first 2D example using a set of terrain attributes derived from a digital elevation model on a 15 m grid to predict continuous soil properties such as A horizon thickness and pH for a small catchment in Colorado. Odeh et al. (1994) did a similar study in South Australia. Skidmore et al. (1991) predicted forest soil classes in New South Wales from layers of natural vegetation data (representing *o*) and terrain attributes on a 30 m grid. Bell et al. (1992, 1994) predicted soil drainage class from terrain data, and Lagacherie and Holmes (1997) predicted soil classes in the Languedoc using layers of lithological and terrain data. Favrot and Lagacherie (1993) foreshadowed this as a general approach for making soil class maps.

For quantitative prediction purposes, this has been called the "*clorpt*" or "*corpt*" equation (McBratney et al., 2000). Some people have termed the approach "environmental correlation" (McKenzie and Austin, 1993). McKenzie and Ryan (1999) used environmental correlation associated with stratigraphy, digital terrain models and gamma-radiometric survey, respectively, to predict soil properties in Australia. Ryan et al. (2000) reviewed the concepts and applications of spatial modeling using the "environmental correlation" approach and used it to predict forest soil properties at the landscape level.

For predicting soil classes, S_c, or soil properties, S_a, often only a subset of the five soil-forming factors has been used, for example, when information from a digital elevation model is available,

$$\begin{aligned} S_c &= f(r), \quad \text{e.g., Bell et al. (1992)} \quad \text{or} \\ S_a &= f(r), \quad \text{e.g., Moore et al. (1993),} \end{aligned} \tag{37.5}$$

or relief and a lithology map,

$$S_c = f(r, p), \quad \text{e.g., Lagacherie and Holmes (1997),} \tag{37.6}$$

or relief and vegetation,

$$S_c = f(r, o), \quad \text{e.g., Skidmore et al. (1991).} \tag{37.7}$$

37.2.2.3 Combinations: Clorpt (or Corpt) and Kriging

Alert readers will have noted that there has been a certain similarity and convergence between the co-kriging and the environmental correlation approach. Some workers recognized this in the mid-1990s and combined the two in what is generically known as regression-kriging (Knotters et al., 1995; Odeh et al., 1995). In this approach, "*clorpt*" is used to predict the soil property of interest from environmental variables, and kriging is used on the residuals. Bourennane et al. (1996) used kriging with external drift, which is related to regression-kriging but only allows for linear relationships between the variable of interest and the environmental variables (the external drifts).

37.2.3 Predicting Soil Attributes from Other Soil Attributes: $s_1 = f(s_2)$

As noted above, some of the co-kriging studies (McBratney and Webster, 1983; Vauclin et al., 1983) showed that soil could be predicted from other soil attributes. This observation in itself is not very useful unless there are much denser secondary variables available. Remote (e.g., gamma radiometrics) and proximal sensing (e.g., electromagnetic induction) offer this possibility. This becomes increasingly important because Phillips (2001) gives several examples where "*clorpt*" apparently does not work, particularly at fine resolutions. This suggests that for predictive purposes, s (for soil) should be added to the "*corpt*" list.

37.2.4 Some Brief Conclusions

From this brief review, we see that

1. Quantitative relationships have generally been most easily found between soil and topography but there is evidence of quantitative relationships with the other four soil-forming or soil-altering factors.
2. In general, the relationships cannot be assumed to be linear.
3. Little work has been done on interactions between factors.
4. Soil can be spatially predicted from geographic position using a variety of techniques.
5. Soil can be predicted from other soil attributes at the same location.
6. Soil can be predicted from itself, other soil attributes, and environmental attributes at neighboring locations.

37.3 The Scorpan Model

McBratney et al. (2003) generalized and formalized the digital soil mapping approach using a Jenny-like formulation not for explanation but for empirical quantitative descriptions of relationships between soil and other spatially referenced factors with a view to using these as soil spatial prediction functions. This is called the "*scorpan*" model, which can be written as follows:

$$S_c = f(s, c, o, r, p, a, n) \quad \text{or} \quad S_a = f(s, c, o, r, p, a, n), \quad (37.8)$$

where
S_c is soil classes
S_a is soil attributes

There are seven factors:

1. s—soil, other properties of the soil at a point
2. c—climate, climatic properties of the environment at a point
3. o—organisms, vegetation or fauna or human activity
4. r—topography, landscape attributes
5. p—parent material, lithology
6. a—age, the time factor
7. n—space, spatial position

Soil is included as a factor because soil can be predicted from its properties, or soil properties from its class or other properties. The s refers to soil information either from a prior map, or from remote or proximal sensing or expert knowledge. Implicit in this are the spatial coordinates x, y (and probably not z) and an approximate or vague time coordinate $\sim t$. This time coordinate can be expressed as "at about some time t." So explicitly, for example,

$$S_c[x, y, \sim t] = f(s[x, y, \sim t], c[x, y, \sim t], o[x, y, \sim t],$$
$$r[x, y, \sim t], p[x, y, \sim t], a[x, y], [x, y]). \quad (37.9)$$

Each factor will be represented by a set of one or more continuous or categorical variables, for example, c by average annual rainfall and average annual temperature or climate class.

We shall not consider the direction of causality. For example, many reckon vegetation to be dependent on soil and we could write $o = g(S)$, where o is set of vegetation classes or percentage cover of a species, g is some arbitrary function, and S is a set of soil classes or attributes. For our purpose, we could write $S = g^{-1}(o)$, where g^{-1} is the inverse function of g, $S = g^{-1}(o) = f(o)$. We stress that the approach in general is not theoretical, and it is largely empirical—where there is evidence of a relationship we use it. Clearly, although we do not require causality, we should be mindful of potential problems of nonuniqueness if g is not a monotonic function, as well it might be if S is say topsoil pH and o is the number of plants of a particular species per unit area.

A general soil prediction model would be

$$S(x, y, z, t) = f(Q), \quad (37.10)$$

where Q is predictor variable(s). Here we will consider some restrictions, in cases where S stands for $S(x, y, (z), t)$, that is, the soil class or attribute at some spatial location x, y, (z) and at some time t.

37.3.1 What Is *S*? Soil Classes *S*$_c$ or Individual Soil Attributes *S*$_a$

The model must be able to predict the probability of a set of classes, for example, for the case of five classes, say, A, …, E, the model would predict the probability vector (p[A], p[B], p[C], p[D], p[E]), for example, $S_c[x, y]$ = (0.01, 0.72, 0.01, 0.02, 0.25) along with some measure of uncertainty. The problem will generally consist of a preexisting soil class label (from some soil classification system) at each soil observation location and a set of colocated environmental variables. These are called the training data. This represents a supervised classification or supervised learning problem. (More rarely, unsupervised learning also known as numerical classification may be used on observed soil attributes to first generate the class labels.) The supervised learning rules are fitted using the training data and then used at other locations where only environmental variables are observed. Most prediction methods treat soil classes as "labels," and their prediction only considers the minimization of the misclassification error. Soil classes at any taxonomic level have taxonomic relationships between each other, and in some instances, the errors in prediction of certain classes are more serious than the others. Minasny and McBratney (2007a, 2007b) proposed the incorporation of taxonomic distance between the soil classes in the prediction.

The model should also be able to predict individual soil attributes S_a along with a measure of uncertainty. The S_a might be the value of a given soil attribute at a certain depth, for example, the clay content at 60 cm, that is, $S_a[x, y]$ = 310 g/kg, along with an uncertainty measure. Similarly to the class problem, this will generally consist of a measured soil attribute at each soil observation location and a set of colocated environmental variables. These are the training or calibration data. This represents a generic (multiple) regression problem. The generic regression equations or rules are fitted using the calibration data and are then used at other locations where only environmental variables are observed.

Heuvelink and Webster (2001) have discussed the merger of discrete and continuous models of spatial variation. Heuvelink (1996) suggested the mixed model of spatial variation, in which the soil property is treated as the sum of a global mean, a class-dependent deviation from the mean, and a spatially correlated residual. Prediction with this model boils down to kriging with an external drift (Delhomme, 1978), which in this case is a classification. Its main advantage is that it performs well over the whole range of spatial variation, from exclusively discrete realities. A more general interpretation of this kind of idea, and the one we use here, is that the external drift represents *f*() and can be any kind of function. The discreteness or continuity of *S* will depend on the magnitude and form of *f*(). In the Heuvelink (1996) case, the *f*() is a one-way analysis of variance model, a special case of a generalized linear model (GLM) (McCullagh and Nelder, 1983; Lane, 2002).

37.3.2 The General Approach

If we write the equation as *S* = *f*(*Q*) + *e*, then the general approach we shall use is to take some observations of *S* in the field at known locations [*x*, *y*] and fit some kind of function to a set of pedologically meaningful predictor variables *Q*, which will generally be raster data layers of size *M* in a GIS. Once the model is fitted at the *m* observation points, the prediction can be extended to the *M* points or cells in the raster thereby giving a digital map. The efficiency of the method relies on the fact that hopefully *m* ≪ *M*, and because *S* is much more difficult and expensive to measure than the *Q*. The success will depend on the following:

1. Having sufficient predictor variables observed everywhere or at least with a relatively high data density
2. Having enough soil observations (data points) to fit a relationship
3. Having functions *f*() flexible enough to fit a nonlinear relationship
4. Having a good relationship between the soil and its environment

Followed by a discussion of quantitative procedures for fitting *f*() in Section 37.3.3, we present some considerations concerning *e* in Section 37.3.4, and a review of previous studies in Section 37.3.5.

37.3.3 Form of *f*()

We will now discuss some forms of *f*(), most, but by no means all, of which have been or can be used for this kind of problem. For the sake of brevity, we shall not delve deeply into the mathematics of the methods. The advance in statistical learning techniques, enhanced by ongoing developments in data mining, has aided the use of different forms of *f*() in soil science. Recent developments and technical details of the statistical modeling have been recently and extensively reviewed by Hastie et al. (2009). When predicting soil classes, some kind of *supervised classification* will be used, and for soil attributes, some kind of *generic regression* will be used. These are now discussed.

37.3.3.1 Linear Models

Linear models include regression for predicting soil attributes and classification for predicting soil classes. Linear regression included in this section is linear models using ordinary or generalized least squares (GLS). Linear methods for classification include discriminant analysis.

37.3.3.1.1 Ordinary Least Squares

For multiple linear regression, the model is written as follows:

$$\mathbf{s} = \mathbf{Qb} + \mathbf{e}, \tag{37.11}$$

where
  s is the vector of response (predicted soil attribute)
  Q is the matrix of predictor variables
  b is parameter vector of the linear function
  The error component, **e**, represents the deviations of the model to the observed value

The parameter is usually solved using ordinary least squares (OLS), with assumptions that **e**

1. Are independently and identically distributed (independence assumption)
2. Have zero mean and finite variance (homoscedasticity assumption)
3. Are normally distributed (normality assumption)

OLS has been used widely in prediction of soil attributes, because of the ease of use and wide availability. The predictors are usually continuous variables; however, qualitative factors or discrete variables can also be integrated.

37.3.3.1.2 Principal Component Regression and Partial Least Squares

When a large number of correlated predictor variables is present (such as electromagnetic spectra), usually principal component analysis is used to produce linear combinations of the original inputs. Selected principal components are then used in place of the original predictors. Alternatively, partial least squares (PLS) (Martens and Naes, 1989) are developed, which constructs a new set of components as regressor variables that are linear combination of the original variables. Unlike principal component regression, which only uses a combination of the predictors, the components in PLS are determined by both the response variable(*s*) and the predictor variables.

Principal component regression and PLS have been used quite extensively in predicting soil attributes from the electromagnetic spectrum, especially in the near and mid-infrared ranges (e.g., Chang et al., 2001). This method may be necessary if the environmental covariates consist of hyperspectral imagery.

37.3.3.1.3 Linear Discriminant Analysis

Discriminant analysis (Fisher, 1936) is the seminal supervised learning technique. It has been applied in soil science for more than 70 years. Webster and Burrough (1974) used the method to allocate soil observations into existing classes. Henderson and Ragg (1980) employed a multivariate logistic method to assess the usefulness of soil properties for distinguishing between taxonomic units. The method was perhaps first used for digital soil mapping by Bell et al. (1992, 1994) who related soil drainage classes to landscape parameters and used the resulting discriminant functions for spatial predictions.

The theory is readily accessible in Webster and Oliver (1990) and Hastie et al. (2009). Triantafilis et al. (2003) generalized the theory to a fuzzy linear discriminant analysis. This considers the a priori membership of each individual to each of the classes.

37.3.3.2 Generalized Linear Models

GLMs extend the linear regression models to accommodate nonnormal response distributions (Hastie and Pregibon, 1992). The theory and applications in soil science have been reviewed by Lane (2002). Usually, to accommodate for nonlinearity, transformation of variable is introduced, GLM attempt to modify the model rather than transforming the data (Lane, 2002). GLMs have the assumption of independence between the response and predictor variables.

37.3.3.2.1 Prediction of Continuous Soil Attributes S_a

McKenzie and Austin (1993) used GLMs to predict soil attributes (clay content, CEC, EC, pH, bulk density, and COLE) using environmental variables (geomorphic unit, local relief, etc.) as predictors. Other examples include Odeh et al. (1995, 1997) and McKenzie and Ryan (1999) who used GLMs to predict nonnormally distributed continuous variables. In both of these cases, the GLM was used in preference to standard linear regression because of the nonnormal distribution of the response variable. Park and Vlek (2002) found that GLMs performed better than neural networks and regression trees in predicting soil attributes from environmental variables.

37.3.3.2.2 Prediction of Soil Classes S_c

Gessler et al. (1995) used GLMs to predict the presence or absence of a bleached A2 horizon using digital terrain information. In this case, a logit link function was used due to the binomial distribution. Campling et al. (2002) used logistic regression to model soil drainage classes from terrain attributes and vegetation indices as calculated from a Landsat TM image. Kempen et al. (2009) developed a logistic regression model for predicting soil groups in the Netherlands from DEM, groundwater maps, land cover, paleogeography, geomorphology, and soil maps. The model-building process was guided by pedological expert knowledge to ensure that the final regression model is not only statistically sound but also pedologically plausible.

37.3.3.3 Generalized Additive Models

Generalized additive models (GAMs) attempt to characterize the nonlinear effect, which is not considered in GLMs. GAMs have the form:

$$S = \alpha + \sum_{j=1}^{p} f_j(q_j) + e, \qquad (37.12)$$

where

α is the intercept
f_j is a nonparametric "smoothing" functions for predictor q_j
p is the number of covariates
e is the error

The smoothing functions can be splines, loess, kernel, and other, smoothers (Venables and Ripley, 1994). Smoothers fit the data locally, and crucial to the fit is the size of the local neighborhood, which is generally controlled by a smoothing parameter. The smoothing parameter controls the variance–bias trade-off. A large neighborhood produces estimates with low variance and potentially high bias, a small neighborhood produces the reverse effect (Hastie and Tibshirani, 1990).

The reported use of GAMs in the soil science literature has been minimal. One of the few studies has been Odeh et al. (1997) who compared GAMs with GLMs and linear regression for the prediction of organic carbon with digital terrain information as secondary information. GAMs were found to be superior. Bishop and McBratney (2001) used GAMs for mapping

of soil cation exchange capacity from environmental factors (terrain attributes, bare soil color aerial photograph, bare soil LANDSAT imagery, crop yield data, and soil apparent electrical conductivity).

37.3.3.4 Tree Models: Classification and Regression

Various implementations of decision trees (DT) or classification trees (CART) have been used to extract and apply knowledge about soil spatial patterns. Tree-based models are fitted by successively splitting a data set into increasingly homogenous subsets. The response variable can be either a factor (classification trees) or a continuous variable (regression trees), and the explanatory variables can be of either type (McKenzie and Ryan, 1999). Once the partitioning has ceased, the subsets are called terminal nodes. Each terminal node is typically assigned the label of the majority class. Splits, or rules defining how to partition the data, are selected based on information statistics that measure how well the split decreases impurity (heterogeneity or variance) within the resulting subsets (Scull et al., 2005). All possible organizations of explanatory variables into two groups are examined recursively to evaluate the effectiveness of each possible split.

Decision trees have found favor in digital soil mapping (DSM) because they can handle missing values, can use continuous and categorical predictors, are robust to predictor specification, and make very limited assumptions about the form of the regression model (Henderson et al., 2005).

Lark et al. (2007) expressed some reservations about overfitting with data mining methods such as regression trees unless the methods were evaluated against truly independent test data.

Henderson et al. (2005) concluded that regression trees had been widely adopted because they were robust, could handle continuous and categorical data, and were reasonably easy to interpret. The final trees are essentially binary keys with yes/no decisions based on a single class or parameter value at each node. This makes it easy to express the splitting rules in semantic terms that are easily understood and appreciated by human interpreters.

37.3.3.4.1 Prediction of Continuous Soil Attributes S_a

This is called a regression tree. Pachepsky et al. (2001) used regression tree to predict sand and silt contents and water retention from terrain attributes (slope, curvature). One of the limitations in regression trees is the discrete predictions from each terminal node, which result in a lack of smoothness of the prediction surface. This can result in unrealistic representations of soil variability if the tree has a small number of terminal nodes (McKenzie and Ryan, 1999). An improvement to regression trees involves building multivariate linear models in each node (leaf). This type of model, which is analogous to using piecewise linear functions, has been implemented in the program Cubist (RuleQuest Research, 2000). This is used by Henderson et al. (2005) for mapping soil properties throughout Australia.

37.3.3.4.2 Prediction of Soil Classes S_c

This is called a decision tree or a classification tree. The binary decision tree algorithm uses a binary split that has exactly two branches at each internal node. There are different decision tree methods. The most commonly used is CART (Breiman et al., 1984). Lagacherie and Holmes (1997) discussed the application of CART for soil classification and its sensitivity to error. Another popular algorithm is C4.5 (Quinlan, 1992) and its later version See5 (RuleQuest Research, 2000). Bui and Moran (2003) utilized this program for mapping soil classes across the Murray–Darling Basin in eastern Australia. Moran and Bui (2002) refined the analysis of Bui et al. (1999) by using a "boosted" tree to reduce the classification error.

37.3.3.5 Neural Networks

Neural networks attempt to build a mathematical model that supposedly works in an analogous way to the human brain. Neural networks have a system of many elements or "neurons" interconnected by communication channels or "connectors," which usually carry numeric data, encoded by a variety of means, and organized into layers. Neural networks can perform a particular function when certain values are assigned to the connections or "weights" between elements. To describe a system, there is no assumed structure of the model, instead the networks are adjusted or "trained" so that a particular input leads to a specific target output. The mathematical model of a neural network comprises of a set of simple functions linked together by weights. The network consists of a set of input units, output units, and hidden units, which link the inputs to outputs. The hidden units extract useful information from inputs and use them to predict the outputs (Hastie et al., 2009).

Neural networks are now widely described in the soil science literature, mainly for predicting soil attributes. The application of neural networks as pedotransfer functions for predicting soil hydraulic properties is the most common.

37.3.3.5.1 Prediction of Continuous Soil Attributes S_a

The practice of predicting soil hydraulic properties in the form of pedotransfer functions can be found in many studies such as Minasny and McBratney (2002). Chang and Islam (2000) predict soil texture from multitemporal remotely sensed brightness temperature and soil moisture maps. Minasny et al. (2006) used a modified form of neural networks for mapping the parameters of a negative exponential depth function.

37.3.3.5.2 Prediction of Soil Classes S_c

Neural networks can be used to predict the probability of classes using multilogit transformation of the output. Another type of network is called self-organising maps (SOM, in this case, *not* soil organic matter) (Kohonen, 1982). Kohonen's network is an unsupervised classification splitting input space into patches with corresponding classes. It has the additional feature that the centers are arranged in a low-dimensional structure (usually a string, or a square grid), such that nearby points in the topological structure (the string or grid) map to nearby points in the attribute space.

Zhu (2000) used neural networks to predict the probability of soil classes from soil environmental factors. Fidêncio et al. (2001) applied two types of neural networks (radial basis function networks and

SOMs) to classify soil samples from different geographical regions in Sao Paulo, Brazil by means of their near-infrared (diffuse reflectance) spectra. Behrens et al. (2005) used neural networks to predict soil classes based on existing soil maps. The predictors are terrain attributes, geologic–petrographic units, and land use.

37.3.3.6 Fuzzy Systems

Zadeh (1965) defined a fuzzy set as "a class of objects with a continuum of grades of membership." Such a set is characterized by a membership (characteristic) function that assigns to each object a grade of membership ranging between 0 and 1. In classical Boolean logic, each individual, or site, either belongs to a class (membership = 1) or does not belong to a class (membership = 0). In Fuzzy logic, each individual, or location, is assigned a value between 0 and 1 that expresses the relative degree to which it belongs to a class. The degree of belonging to a given class (range 0–1) is assessed by comparing the properties that define the central concept for a class to the properties exhibited by any individual, or location, that one wishes to classify.

In the context of digital soil mapping, fuzzy systems represent one approach to assigning an estimate of the likelihood of a particular soil class or soil property value occurring at any given location under a given set of environmental conditions. Typically, the environmental conditions are represented by a series of covariates or predictor variables that exhaustively cover the spatial extent of an area for which predictions are required. The fuzzy likelihood that a given soil class or soil property value will occur at any point, and under any particular set of conditions, has been estimated using several quite different approaches. These approaches can be broadly characterized as unsupervised, supervised, and knowledge based (Hengl and MacMillan, 2009).

Unsupervised approaches typically do not have an a priori set of target classes defined and do not use spatially located reference locations to establish relationships between observed classes and values of the covariate predictor data sets. For example, Irvin et al. (1997) used a fuzzy *k*-means unsupervised approach to derive groupings of fuzzy soil–landform classes based solely on consideration of the input data values without reference to any existing classification. This approach used iterative processes to determine class means by minimizing distances in multidimensional attribute space. The resulting continuous classification created "partial" class memberships for each data point with membership values ranging from 0 to 1.

In situations in which the optimum number of classes to be identified, and the conditions or criteria that define the central concepts of each class, are not known a priori, statistical ordination techniques can be applied to extract class numbers and definitions. Fuzzy *k*-means or fuzzy *c*-means has been used extensively in soil science and DSM to identify optimum class numbers and central concepts and to then determine the fuzzy degree of membership of unclassified individuals or locations to the central concepts of each of the statistically defined classes (McBratney and De Gruijter, 1992; Odeh et al., 1992; De Gruijter et al., 1997; McBratney and Odeh, 1997; Burrough et al., 2000; Carré and Girard, 2002; Zhu, 2006). Detailed descriptions of the equations used in Fuzzy *k*-means, and the logic behind them, can be found in Odeh et al. (1992), McBratney and Odeh (1997), and Zhu (2006).

In digital soil mapping, most applications of fuzzy *k*-means classification have adopted an approach in which environmental covariates (e.g., slope, aspect, curvatures, climate variables, or vegetation cover) were used to define the central concepts of classes of soil of interest.

Supervised approaches all start out with a defined list of a priori classes that a user wants to predict and they all make use of spatially located reference data to establish the fuzzy membership functions that relate values of predictor variables to the corresponding likelihood that a particular soil class or property value will occur. These approaches all represent a form of data mining in which instances of known classes at known locations are used to identify values of predictor variables that are associated with the presence of a given class or property value. The training data can be drawn from a variety of different sources, including preexisting soil maps, preexisting (legacy) point observations or sample sites, newly collected, geolocated field observations or sample data or even points or areas identified interactively, on-screen, by a knowledgeable expert as being likely to belong to a particular class or to have a particular property value. A key requirement for successful use of supervised approaches is that there must be instances in the training data of all classes that exist in an area and that a user wants to predict. Additionally, the training data should cover, or represent, the full range of values exhibited by each and everyone of the covariates used to predict the fuzzy class memberships. A Latin hypercube sampling (LHS) procedure (Minasny and McBratney, 2006) can be used to analyze the feature space of the covariates to establish whether the instances in the training data represent the full range of covariate values in the feature space. If the training data do not fully cover the feature space, then there will be confusion as to which classes or property values are most likely to occur at locations in the feature space where there are no instances of training data for a particular range of covariate values.

Several different approaches have been described for collecting and analyzing training data to establish fuzzy membership functions for soil classes or properties. Some involve identifying multiple instances of each class of interest within an area of interest. Each instance, or location, is treated as a representative profile for a soil class of interest, and each unclassified location is compared to each instance of each class to identify the instance to which it is most similar (Zhu et al., 1996; Zhu, 2000). In a variation of this approach, Shi et al. (2004, 2005, 2008) treated individual locations are cases and assigned different weights to cases such that cases closer to an unclassified location carried a greater weight than cases farther away. This case-based reasoning (CBR) permitted instances of a class to influence the classification of unclassified locations that were located near to them while not exerting much influence on classification of locations that were further away.

37.3.3.7 Other Methods

Ballabio (2009) used support vector machine (SVM) algorithm for mapping soil organic carbon in mountainous regions in Italy.

SVM is a set of supervised learning methods used for classification and regression, and its algorithm aims to match model complexity to data complexity. SVM performs classification by constructing a nonlinear *n*-dimensional hyperplane that optimally separates the data into two categories. This is done by constructing a linear boundary in a large, transformed space of the feature space. Support vector regression is a form of SVM that is applied to regression problems.

Genetic algorithms (GAs) (Goldberg, 1989) are randomized search and optimization techniques guided by the principles of biological evolution and natural genetics. They have been used mainly in optimization of large multidimensional problems. Pal et al. (1998) developed a GA-classifier and applied it to satellite imagery (Pal et al., 2001). It attempts to approximate the class boundaries of a data set with a fixed number of hyperplanes in such a manner that the associated misclassification of data points is minimized. Nelson and Odeh (2009) explored the Genetic Algorithm for Rule-set Production (GARP) (Stockwell and Noble (1992). GARP was developed to model the habitat distribution of plant and animal species using locations of known species presence and environmental variables. The algorithm has stochastic elements that produce a population of rules prior to iterative use of the best rules in a given generation to develop the next generation until some convergence criteria are met and a solution given. GARP is noteworthy in that it was developed specifically to utilize legacy data. However, Nelson and Odeh's results showed that GARP did not perform as well as decision tree algorithm, implying the need to improve the algorithm for its full potential to be realized for digitally mapping soil classes.

37.3.3.8 Strengthening Models: Bagging, Boosting

There has recently been empirical evidence that the accuracy of *f*() prediction can be enhanced by generating multiple models and aggregating them to produce an estimate. There are two widely used approaches for producing and using several models that are applicable to a wide variety of statistical learning methods. Bootstrap aggregating or bagging (Breiman, 1996) and boosting (Freund and Schapire, 1997) manipulate the training data in order to generate different models. These methods arise more naturally in the supervised classification problem, but they can be extended to generic regression.

Bootstrap methods (Efron and Tibshirani, 1993) assess the accuracy of a prediction by sampling the training data with replacement. Suppose the training data are composed of predictors *Q* and response *S* of size *N*, we draw *B* data sets each of size *N* of the training data by sampling with replacement. For each of the bootstrap data set Z^b, $b = 1, 2, …, B$, we fit model $\hat{f}^b(q)$. The bagging estimate is calculated as

$$\hat{f}(q) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^b(q). \tag{37.13}$$

The bagging method is the basis of the random forests algorithm.

Boosting combines the outputs of many "weak" models to produce a powerful "committee." Boosting uses all the data at each repetition, but maintains a weight for each instance in the training set that reflects its importance. Adjusting the weights causes the model to focus on different data and, hence, leads to different models. The multiple models are then aggregated by voting to form a composite model. In bagging, each component model has the same vote, while boosting assigns different voting strengths to component classifiers on the basis of their accuracy. Moran and Bui (2002) used boosting to improve their digital soil map of the Murray–Darling basin. Henderson et al. (2005) described how fitting multiple trees through boosting and bagging permitted tree averaging, which could lead to potential improvements in the predictive power of decision trees.

Grimm et al. (2008) described random forest (RF) as an implementation of randomized classification and regression trees in which numerous trees are generated within the algorithm and finally aggregated to give one single prediction (Hastie et al., 2009). Random forests combine the tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. It has been used successfully for mapping soil carbon (Grimm et al., 2008). The model can handle lots of inputs, has a high accuracy, and apparently does not overfit the data. This is achieved by using different subsets of the training data (with bootstrap) and using different subsets of the predictors for training the tree (determined randomly). Thus, only patterns that are present in the data would be detected consistently by a majority of the trees. However, this claim is not always true. Hastie et al. (2009) showed that when the number of variables is large but the fraction of relevant variables for prediction is small, random forests are likely to perform poorly and random forests can certainly overfit the data; the average of fully grown trees can result in a model that is too rich and incur unnecessary variance.

37.3.3.9 Expert (Knowledge-Based) Systems

Expert systems (Dale et al., 1989) are ways of harvesting and engineering knowledge. The terms expert system and knowledge-based system have been used both interchangeably and for quite fundamentally different approaches. Zhu (2008) observed that knowledge-based approaches used in digital soil mapping have extracted and codified knowledge in three main ways, these being as follows:

- Extraction of knowledge from existing documents, legends, and keys
- Extraction of tacit empirical knowledge from human domain experts
- Extraction of knowledge from spatial data sets via data mining

Knowledge already formalized and codified in documents, legends, and keys may simply require implementation as either Boolean or Fuzzy rules and not require any specific extraction process. Empirical knowledge about likely relationships between classes of interest and formative environmental conditions can be extracted by asking local experts to describe and quantify their heuristic, and mostly tacit, domain knowledge through

structured interview and evaluation sessions (Zhu et al., 1996; Zhu, 2008). Extraction of knowledge via application of statistically based data mining techniques works by comparing the spatial co-occurrence of known examples of a particular class or entity with the spatial pattern exhibited by sets of covariates that are considered likely to possess some predictive relationship with the class or entity of interest. Interestingly, the widely used SoLIM (Zhu, 1997, 2008) and SIE (Shi et al., 2004, 2008) fuzzy inference systems have both utilized all three approaches to extract expert knowledge.

Let us first consider those expert systems that make use of some form of data mining technique. Data mining methods require two main inputs. First, they require evidence in the form of known observations of the classes or values of interest at known locations. These provide the training data. Training data may exist in either point or map form or both. Second, they require data sets of explanatory variables that are spatially referenced and cover entire areas of interest exhaustively. These explanatory data sets provide the means of developing and then extrapolating predictive relationships from point (or localized map) observations to entire areas of interest. Data mining can be considered as a form of supervised classification as described elsewhere in this chapter.

Decision trees, or classification trees (see Section 37.3.3.4), have emerged as a favored data mining approach in digital soil mapping. Another data mining method that has found some use in digital soil mapping is Bayesian Analysis of Evidence (Skidmore et al., 1991). Cook et al. (1996a) and Corner et al. (1997) developed and applied the custom software program *Expector* to capture and apply human heuristic beliefs in a structured belief system. Mayr et al. (2008) used commercial Bayesian analysis of evidence software, *Netica* by Norsys, to extract rules for classifying soil types by extracting knowledge from analysis of spatially referenced auger bore data and existing soil maps.

Bui (2004) observed that knowledge engineering is the study of expert systems and how experts build mental models. She argued that soil maps and their legends are representations of structured knowledge, namely, the soil surveyor's mental soil–landscape model. Soil map unit descriptions describe, in words, the soil–landscape relationships. For example, in a narrative legend, a soil toposequence can be described as a function of relative elevation/slope position, slope length and slope gradient, and curvature. Sometimes, the soil–landscape model for each map unit is represented as a 2D cross section or 3D block diagram. Soil map unit descriptions and block diagrams are usually the only way the soil surveyors' mental models are transmitted to others. This soil–landscape model has been described by Hudson (1992) as the operative paradigm for soil survey. Bui (2004) and Wielemaker et al. (2001) suggested methodological frameworks to formalize the landscape knowledge of the soil surveyor by structuring terrain objects in a nested hierarchy followed by inference and formalization of knowledge rules.

Knowledge-based expert systems have been characterized as consisting of two main components, these being a "knowledge base" and an "inference engine" (Skidmore, 1989). Zhu et al. (1996) described knowledge-based systems as being organized on three levels of data, knowledge base, and inference engine. Data refer to information on environmental conditions of an area such as elevation, slope, and aspect. The knowledge base contains the declarative knowledge about a particular problem being solved. In the case of predictive soil mapping, the knowledge base contains information about the relationships between the spatial distribution of soils in an area and the environmental factors that are believed, or known, to influence this distribution. The inference engine applies the knowledge to the data to produce predictions or results. It controls when and how the specific problem solution knowledge is used. A particular concern of an inference engine is often the order, or sequence, in which a particular piece of information (or a rule) is processed (Skidmore, 1989).

Shi et al. (2008) recognized several useful distinctions regarding knowledge used in devising and applying fuzzy logic rules. First, they distinguished between *global knowledge* and *local knowledge*. Global knowledge applies with equal validity across an entire area of interest whereas local knowledge is applicable only within certain restricted extents. Second, they distinguished between knowledge based on understanding of relationships between classes of interest and values of environmental variables in *parameter space* and knowledge based on the existence of known classes or conditions in particular locations in *geographical space*. They differentiated global rule-based reasoning (RBR), capable of handling global knowledge in parameter space, from global and local CBR, capable of handling global versus local knowledge in geographical space. CBR is used when local experts can identify locations at which specific classes of interest are known to occur and can state whether these classes always occur in these same settings throughout an area or only occur in these settings within a restricted geographical extent. RBR is used when local experts can identify the environmental conditions, as defined by available covariates, under which conceptual classes of interest are known to occur in parameter space.

Shi et al. (2008) provide an excellent description of eight main steps that are involved in devising, applying, evaluating, refining, and then finalizing knowledge-based rules to predict the spatial pattern of soil classes using the soil inference engine (SIE). They noted that MacMillan et al. (2007) independently, but not surprisingly, employed a procedure that was almost identical to theirs. The eight steps described by Shi et al. (2008) are as follows:

1. The soil scientist provides the global knowledge, including names of the soils he/she expects to see in the mapping area and descriptions of the environmental conditions of these soils. The environmental conditions depicted by environmental values are formalized into Fuzzy rules and saved into rule bases; those represented by geographical locations are formalized into cases and saved into global case bases.
2. The soil scientist or a GIS specialist prepares data layers for characterizing the environmental conditions. The data layers may cover terrain attributes, geology, vegetation, climate, and other features. These data layers are stored in a GIS database.

3. SIE performs RBR or global CBR, using the global knowledge and the GIS database, to generate maps of the general pattern of soil distribution in the mapping area.
4. The soil scientist verifies the draft maps from (3). If he/she is satisfied with the maps, the mapping is done. Otherwise, he/she may go back to (1) to adjust the rules or global cases, or go to (5) to fine-tune the draft maps.
5. The soil scientist provides the local knowledge, in the form of cases, to address local exceptions. The cases are saved in local case bases.
6. SIE performs local CBR using the local knowledge and the GIS database.
7. The soil scientist verifies the maps from (6). He/she can adjust the cases and run local CBR again. He/she repeats this process until he/she is satisfied with the result.
8. The soil scientist uses SIE and other GIS tools to integrate the results from (3) and (6) to generate the final maps.

The development of the SoLIM inference model (Zhu and Band, 1994; Zhu, 2006, 2008) through time provides a revealing example of trends in how fuzzy knowledge of soil–landscape relationships has been captured and applied. Early implementations of SoLIM were mostly based on using interviews to capture the tacit knowledge of local experts. However, they often did not take into account the consideration of the cognitive aspects of knowledge formulation and required users to provide either exact forms for membership functions for each class to be defined or a very large set of typical cases or instances that could be used as exemplars in CBR (Qi et al., 2006). In a second stage of development, many SoLIM applications relied on acquisition of large numbers of user-selected spatially referenced instances or cases to supply the information required to define central concepts, and local exceptions, for representative or typical soil classes (Zhu, 2000; Shi et al., 2004, 2005). As SoLIM subsequently evolved and was increasingly applied for operational mapping of larger areas, the use of large volumes of geolocated field observations to serve as training data sets became increasingly more costly and less viable. More recent applications of SoLIM have notably adopted approaches that were based on rapidly identifying prototypes (Qi et al., 2006) or semantically defined central concepts (Liu and Zhu, 2009) by using expert tacit knowledge to define conceptual or mental models of central concepts of soil classes to be predicted. These semantically defined representative concepts not only require far less time and effort to produce than geolocated reference sites but they have also been shown (Qi et al., 2006; Liu and Zhu, 2009) to be more effective and to produce more accurate predictions of the spatial distribution of soil classes than earlier case-based approaches.

The equation for computing an estimate of any given soil property of interest is given by Zhu et al. (1997) as follows:

$$V_{i,j} = \frac{\sum_{k=1}^{n} \mathbf{S}_{i,j}^k V^k}{\sum_{k=1}^{n} \mathbf{S}_{i,j}^k}, \tag{37.14}$$

where
$V_{i,j}$ is the continuous soil property to be predicted
$S_{i,j}$ is the soil similarity value for the soil k at location i,j
V^k is the mean or representative value for the soil property of interest for soil k

While this approach produces estimates that are not quite as accurate or reliable as those produced by other, more data intensive methods such as regression-kriging or regression trees, it is an attractive option for areas that lack extensive databases of geolocated point information on soil properties. In these circumstances, all that is required is a single representative value for each soil property of interest, for each depth or horizon of each soil class of interest and an ability to develop fuzzy knowledge-based rules about which environmental factors influence the development and occurrence of the soil classes of interest.

37.3.3.10 Unsupervised Classification

In the previous sections, when we were discussing classes we were considering "supervised classification." This is also known as allocation or identification. This is where we wish to produce prediction equations for placing soil existing soil classes, such as a particular categorical level in a national or international classification system. However, we may first wish to make new classes from the observed soil properties. This is known as unsupervised classification. Much of the early work on pedometrics, in the 1960s, focussed on this topic. The numerical classification methods that have been used quite extensively in soil science more recently are k-means and fuzzy k-means (Odeh et al., 1992; De Bruin and Stein, 1998; Triantafilis et al., 2001). See the section on fuzzy systems on the application of fuzzy k-means. There is also a semi-supervised classification considering classification in the presence of some labeled data (Pedrycz and Waletzky, 1997).

Unsupervised classification is an option for making digital soil class maps particularly where the national or international scheme does not project well onto the soil–landscape. However, once the new classes have been established at the soil observation locations, then one of the previous methods, inter alia discriminant analysis, multiple logistic regression, regression trees, needs to be applied to fit equations and then make predictions from environmental covariates at the other locations where no soil properties have been observed.

Carré and Girard (2002) used a continuous method for horizon and profile classification called OSACA. This was based mainly on field soil morphological attributes. Carré and Jacobson (2009) further developed a program for quantitative grouping of soil layer descriptions into profile classes. The program calculates the taxonomic distances between observed profiles based on layer (horizon) characteristics. Characteristics can be either observed soil properties or layer class memberships. OSACA can allocate observed soil profiles to existing classes or create a new classification of the profiles. Their method is unique in that it models the taxonomic distance to each of the class centroids at

each observation site. Because these distances are continuous variables, multiple linear regression on environmental variables was used as the "supervised classification" step. One regression equation was developed for each class and the distances predicted at each site on their prediction raster.

37.3.4 Spatial Considerations

The older *corpt* approach has no intrinsic or formal spatial component other than the functions are predicted in a spatial context, that is, spatial position is not taken into consideration. This seems unwise for a mapping application. Spatialization can be introduced by considering spatial components of the environmental and soil variables (Section 37.3.4.1) and by perpending the spatial correlation structure of the residuals (Section 37.3.4.1), as was briefly discussed in Section 37.3.3.1.

37.3.4.1 Decomposition of *Q* Factors into Spatial Components

DSM studies have increasingly recognized the need to identify which scales of variation, which are operative and discernable in the predictor data layers (*Q*), are most strongly related to observed variation in a property or class of interest (Bui et al., 1999, 2006; Fisher et al., 2004, 2005; Smith et al., 2006; Arrell et al., 2007; Deng, 2007; Wu et al., 2008; Zhu, 2008). Several approaches have been used to investigate the spatial structure that may be operative in predictor data sets and to identify the ranges of distance over which this structure operates.

Bui et al. (2006) observed that a spatial hierarchy emerged from their decision tree analysis of predictive relationships between environmental covariates and soil properties over the entire extent of the continent of Australia. They undertook a very logical, but frequently neglected, interpretation of their results in terms of physical and biological processes that influence pedogenesis. They sought to interpret their results in terms of the degree to which different environmental covariates, operative at different scales, could be related to soil-forming processes that influenced the values of the soil property variables they were modeling. Their results indicated that the state factors of soil formation formed a hierarchy of interacting variables, with climate being the most important factor at a continental scale, but with different climatic variables dominate in different regions. Their results also showed that lithology was almost equally important in defining broadscale spatial patterns of soil properties and that shorter-range variability in soil properties appeared driven more by terrain variables. The message to be taken from this thoughtful and rigorous analysis is that different environmental covariates are likely to drive the significant processes of soil formation at different scales and over different distances. Individuals seeking to model the spatial distribution of soil properties ought to first systematically examine the variation in those properties at different scales and seek to identify the environmental covariates (and their scales) that are most closely associated with causing variation in soil property values at specific scales.

In this same context, there has been an increasing number of studies that investigated the effects of different DEM resolutions (Zhang and Montgomery, 1994; Thompson et al., 1997, 2001; Fisher et al., 2004, 2005; Kienzle, 2004; Arrell et al., 2007; Erskine et al., 2007; Seibert et al., 2007; Wu et al., 2008) and neighborhood (window) sizes (Schmidt et al., 2005; Smith et al., 2006; Behrens et al., 2007a, 2007b; Zhu, 2008) on the performance of models designed to predict soil properties or soil classes. Typically, such studies have been applied to smaller areas over which the regional climate and lithology could be assumed to hold constant, or to vary only slightly, relative to more significant variation associated with topography. When predicting individual soil properties, the approach has often been to fit regression models in which the soil property of interest is predicted using environmental covariates computed at different neighborhood sizes and/or grid resolutions. For example, Schmidt and Hewitt (2004) showed that the predictive value of profile curvature for estimating soil properties in a regression model varied with the scale (window size) for which profile curvature was calculated. Similarly, Smith et al. (2006) showed that there was a range of window sizes, which varied from landscape to landscape, which produced the most accurate maps of soil classes, and that finer resolution DEMs did not necessarily create more accurate maps. They observed that there was no apparent physical process-based significance behind using a fixed 3×3 neighborhood for computing terrain attributes. Instead, they concluded that it is much more important to match the terrain characteristics computed from the DEM using a specified neighborhood size, with the characteristics of the real world landscape. In a subsequent publication, Zhu (2008) observed that soil-forming processes operated at specific scales, and that spatial analysis should employ a neighborhood size comparable to the spatial scale of the process under consideration. He argued for use of joint distributions of processes and parameters with recognition of effective neighborhoods, defined in terms of the extent over which a process of interest operates. This spatial extent is comparable to the concept of grain size and, for many soil–landscape studies, is equivalent to the extent defined by a hillslope from crest to channel.

Other approaches that have been used to investigate the spatial structure of environmental covariates include factorial kriging analysis (Bourgault, 1994; Wen and Sinding-Larsen, 1997; Oliver et al., 2000) and wavelets (Zhu and Yang, 1998; Carvalho et al., 2001; Mendonça-Santos et al., 2006). Both of these methods decompose the separate variables into separate hierarchical spatial components of decreasing spatial resolution. The factorial kriging method assumes stationarity while the wavelet method does not. The factorial kriging method finds the scale of the components from the observations, whereas in wavelets the various scales are dictated by the size of the image, that is, the scales are increasing powers of 2 pixels.

The various components of *Q* could conceivably all be derived at different spatial scales and used as separate layers in the fitting of *s*. It is more than likely that the short spatial range components (e.g., the nugget component) might not relate to soil and

can be removed (see Smith et al., 2006). In a quite a different application Oliver et al. (2000) found that land use was related to a long-range spatial component in SPOT imagery and not to two shorter range components.

37.3.4.2 Structure in *e*: Generalized Least Squares and Geostatistics

It would be naïve to imagine that there is no spatial structure in *e*. If *e* has a spatial structure, then GLS or geostatistics can be applied. Why would *e* have a spatial structure? The answers could be as follows:

- *Scorpan* is incorrect.
- Attributes used to describe *scorpan* are inadequate.
- Interactions are misspecified.
- Form of *f*() is misspecified.
- Something intrinsic—such as spatial diffusion, interaction, or inhibition processes.

Variograms of the fitted parts of the soil spatial prediction functions for the various factors will be instructive in elaborating these possibilities.

37.3.4.2.1 Generalized Least Squares

In GLS (Cressie, 1993): $\mathbf{s} = \mathbf{Qb} + \mathbf{e}$; errors $\mathbf{e}$ belong to multivariate normal distribution with mean 0 and covariance matrix $\mathbf{V}$: $N(0, \mathbf{V})$. For spatial data, it can be further simplified assuming the error is homogenous with variance σ^2; thus, $\mathbf{V}$ can be replaced by $\sigma^2\mathbf{C}$, where $\mathbf{C}$ is the correlation matrix of the errors (Lark, 2000).

GLS has been used in soil science literature, for example, Samra et al. (1991) predicted tree growth from soil sodicity parameters with spatially correlated errors. Other examples include Aiken et al. (1991), Opsomer et al. (1999), and Vold et al. (1999). Lark (2000) provided the theory and example of using GLS for mapping soil organic matter content. Hengl et al. (2003) utilized GLS as a regression-kriging procedure for spatial prediction of soil properties in Croatia.

The drawback with this method is the heavy computation time when a large volume of data is involved, as the calculation time of semivariance and inverse of the correlation matrix $\mathbf{C}$ will increase (approximately cubically with sample size); nevertheless, Opsomer et al. (1999) showed some mathematical manipulation to avoid the inversion of the whole matrix. Pace and Barry (1997) developed a spatial autoregressive model, which utilized the sparse matrix technique to allow for quick computation for large spatial data. Another approach has been to choose classes of covariance functions for which kriging can be done exactly, even though the data set is massive (Cressie and Johannesson, 2008). In this approach, a multiresolution spatial process is constructed explicitly so that (simple) kriging can be computed extremely rapidly, with computational complexity linear in the size of the data.

37.3.4.2.2 Geostatistics: Scorpan Kriging

Here, we recognize that the spatial "trend" can be described by $f(s, c, o, r, p, a, n)$, and the residuals e modeled separately by variograms and kriging. The final prediction is the sum of $f()$ and e.

37.3.4.2.2.1 Scorpan: Universal Kriging Universal kriging allows incorporation of both deterministic and stochastic components in kriging:

$$S(\mathbf{x}) = \sum_{j=0}^{p} b_j q_j(\mathbf{x}) + e(\mathbf{x}) + \varepsilon, \tag{37.15}$$

where

the first term represents the nonstationary trend, which is modeled as a set of linear functions of the environmental variables $\mathbf{Q}$ with parameter vector $\mathbf{b}$

the second term is the stochastic component modeled by variogram

the third component ε is the error term that occurs below the level of spatial resolution or support of the covariates

Universal kriging can be solved by modifying the kriging system. However, the trend function is only limited to linear functions, and when the number of variables p is large, the matrix inversion to solve the system can consume heavy computation time.

Alternatively, the trend function can be modeled separately, where kriging is combined with regression (Ahmed and DeMarsily, 1987; Knotters et al., 1995). This method involves regression of the soil attributes as a function of predictor variables. This is followed by kriging of the regressed values, where the variance of the predicted (from the regression model) is used as the uncertainty of the modified kriging system. This is also known as kriging with uncertain data (Ahmed and DeMarsily, 1987). Odeh et al. (1994, 1995) defined regression-kriging where model $f()$ is used to describe the relationship between predictors and soil attributes:

$$S(\mathbf{x}) = f(\mathbf{Q}, \mathbf{x}) + e'(\mathbf{x}) + \varepsilon, \tag{37.16}$$

where

$f(\mathbf{Q}, \mathbf{x})$ is a function describing the structural component of S as a function of $\mathbf{Q}$ at $\mathbf{x}$

$e'(\mathbf{x})$ is the locally varying, spatially dependent residuals from $f(\mathbf{Q}, \mathbf{x})$

In regression-kriging, the soil property S at an unvisited site is first predicted by $f()$ and then followed by kriging of the residuals of the model.

As discussed at the end of Section 37.3.3.10, Carré and Girard (2002) used a continuous method for horizon and profile classification followed by multiple linear regression on environmental variables. One regression equation was developed for the taxonomic distance to each class centroid, and the taxonomic distances were predicted at each site of their prediction raster. The residual taxonomic distances to each class centroid were then spatially predicted onto the raster using ordinary kriging and added to the taxonomic distances from regression analysis. The summed taxonomic distances were then displayed and manipulated to make class maps. In this way, regression-kriging can be

used to make continuous or discrete soil class maps taking into account the spatial correlation structure of the residuals from the fitted classes at each data point.

37.3.4.2.2.2 Scorpan: Simple Kriging Because $f()$ is modeled under the assumption that e has zero mean, simple kriging can be applied to the residuals of the model. Simple kriging allows prediction of the spatially correlated residuals with known mean where the weights of the kriging equation do not need to sum to unity (Webster and Oliver, 1990).

37.3.4.2.2.3 Scorpan: Compositional Kriging So far, kriging has been used mainly to predict soil attributes, for prediction of soil classes incorporating predictor variables **Q**, a form of compositional kriging with external trend is proposed:

$$\Pr[S_c(\mathbf{x})] = f(\mathbf{Q}, \mathbf{x}) + e_c'(\mathbf{x}) + \varepsilon, \qquad (37.17)$$

where $\Pr[S_c(\mathbf{x})]$ is the probability of the soil at **x** that belongs to soil class c. The probability of the soil classes $c = 1, \ldots, K$ must sum to 1 and the residuals of the probability must sum to 0:

$$\sum_{c=1}^{K} \Pr[S_c] = 1, \qquad (37.18)$$

$$\sum_{c=1}^{K} e_c' = 0. \qquad (37.19)$$

Solution of this method will involve prediction of probability of occurrence of soil classes using a form of $f()$, such as logistic regression, and compositional kriging of their residuals (Walvoort and de Gruijter, 2001).

37.3.4.2.3 Geostatistics: Co-Kriging and Coregionalization Analysis

Another method is co-kriging. Any of the q layers can be a covariate in co-kriging. Indeed, many people have used this. The major problems with co-kriging have been twofold. First, the parameters of all the $[(q + 1)q]/2$ variograms and cross-variograms have to be estimated and the parameters have to obey a strict inequality (Wackernagel, 1987). Secondly, and more importantly, the co-kriging model really assumes linear relationships between the predictor and predicted variables. The use of categorical predictors and predicted variables is also difficult. Odeh et al.'s experience (Odeh et al., 1995) was that co-kriging did not perform as well as $S = f(r) + e$ (regression-kriging) and was more cumbersome to use. We prefer the *scorpan* model, but co-kriging should not be dismissed and the difficulties and restrictions will be overcome. It can be argued that co-kriging is a kind of GLM. Co-kriging can be used for soil attributes, and compositional (co-)kriging (De Gruijter et al., 1997; Walvoort and De Gruijter, 2001) can be used for indicators or probabilities of discrete soil classes or memberships of continuous ones.

37.3.4.2.3.1 Coregionalization Analysis Even if co-kriging is not done, coregionalization analysis (e.g., Lark and Papritz, 2003) is a very instructive way of studying the linear spatial relationships between soil and the predictor variables Q. This will indicate the spatial scales over which we might expect linear relationships to hold.

37.3.4.3 Other Spatial Methods

Bayesian maximum entropy (BME) was introduced by Christakos (1990, 2000). This approach allows the incorporation of a wide variety of hard and soft data in a spatial estimation context. The data sources may come in various forms, such as intervals of values, probability density functions (pdf), or physical laws (Christakos, 2000). Bogaert and D'Or (2002) used BME algorithm and a Monte Carlo procedure (BME/MC) to generate a map of particle-size distributions from a limited number of accurate measurements and a spatially exhaustive soil map. Compared with ordinary kriging (OK), this approach has the advantage of using soft information on a sound theoretical basis.

Brus et al. (2008) used BME to estimate the probabilities of occurrence of soil categories in the Netherlands. They used "hard" observations from 8,369 soil profile descriptions, and the soil map of the Netherlands 1:50,000 as "soft" information. They concluded that BME is a valuable method for spatial prediction and simulation of soil categories when the number of categories is rather small (say <10). For larger numbers of categories, the computational burden becomes prohibitive, and large samples are needed for calibration of the probability model.

37.3.5 Recent Studies

Although the *scorpan* model was formalized in 2003, various authors have fitted parts of it. McBratney et al. (2003) summarized the work of a large number of studies. Their analysis showed that soil attributes have been estimated more often (70% of studies) than soil classes (30% of studies). The key predictor factors are r (80% of studies) followed by s (35%), o and p (both 25%), n (20%), and c (5%) whereas a does not seem to have been used as a factor. The most common combination was r and s. Most studies used a DEM as the main source of ancillary data, followed by remotely sensed imagery and preexisting soil coverage.

Grunwald (2009) reviewed more recent studies in digital soil mapping during the period of 2007–2008. Of the 90 reviewed studies, 40% focused on predictions of basic soil properties such as texture, bulk density, structure, 31% on soil carbon, 24% on environmental quality assessment, 17% on hydrologic properties, 9% on soil degradation (salinity, acidity, and erosion), and 16% on mapping of soil taxonomic/ecological classes. In particular, mapping of soil organic carbon was prominently represented. In these recent studies, about 40% utilized proximal or remote sensors, and 23% of the studies used sensor data to complement analytical soil data, which are more costly and labor-intensive to derive. The sensors include visual and near-infrared, mid-infrared, and apparent electrical conductivity.

37.4 Sources of Data: The Seven Scorpan Factors

There are seven factors or sets of variables in the *scorpan* model, which makes it different from Jenny's model. The aim is to obtain information on all of these. It will be a matter of convenience (access to data sources) and scientific contention, which variables are used to represent the factors. Indeed, this is an area that has not been well enough studied. The creation of these digital maps of the input environmental variables representing the six factors in the *scorpan* model is seen as an integral part of the digital soil resource assessment approach and a very valuable, environmentally useful, by-product of the new approach. The layers can be used for other modeling purposes. Much of the earth science and ecological research of the last 20 years has been contributing toward the creation of these layers.

These digital surfaces themselves will be created using surface modeling procedures; regression-kriging or Laplacian smoothing splines or TINS. For D3 surveys, they should probably be produced on a 100 m raster with a block size of say 100 m × 100 m.

Figure 37.1 highlights the useful parts of the electromagnetic spectrum for obtaining information on soil and environmental variables through remote and proximal sensing. Matter emits electromagnetic radiation in different parts of the spectrum, and this radiation may be measured by different types of spectroscopy depending on the wavelength. This provides a basis for remote sensing of the properties of matter. A sensing system might measure the radiation emitted by an object after the object has itself been irradiated. Two examples of this are the optical remote sensing systems that measure the solar radiation reflected by an object and the synthetic aperture radar systems (SAR) that measure deliberately long-wave radiation backscattered by an object. Alternatively, it may be possible to measure radiation emitted by an object because of its temperature (emitted in thermal infrared frequencies) or because of radioactive decay (decay of uranium, thorium, and potassium isotopes is widely measured by "passive" gamma radiometry in geophysics). The electromagnetic radiation emitted from an object will therefore depend on its physicochemical properties, some of which are of direct interest in the study of soil (the temperature, mineralogy, organic content, physical structure, or the chlorophyll content of overlying vegetation).

37.4.1 "s" Factor

Remote and proximal active and passive sensing gives detailed information on the soil itself—these reflections or emissions or transmissions are intrinsic properties of the soil material and profile they may indicate other soil attributes like texture or mineralogy. This factor is likely to becoming increasingly important as technology advances.
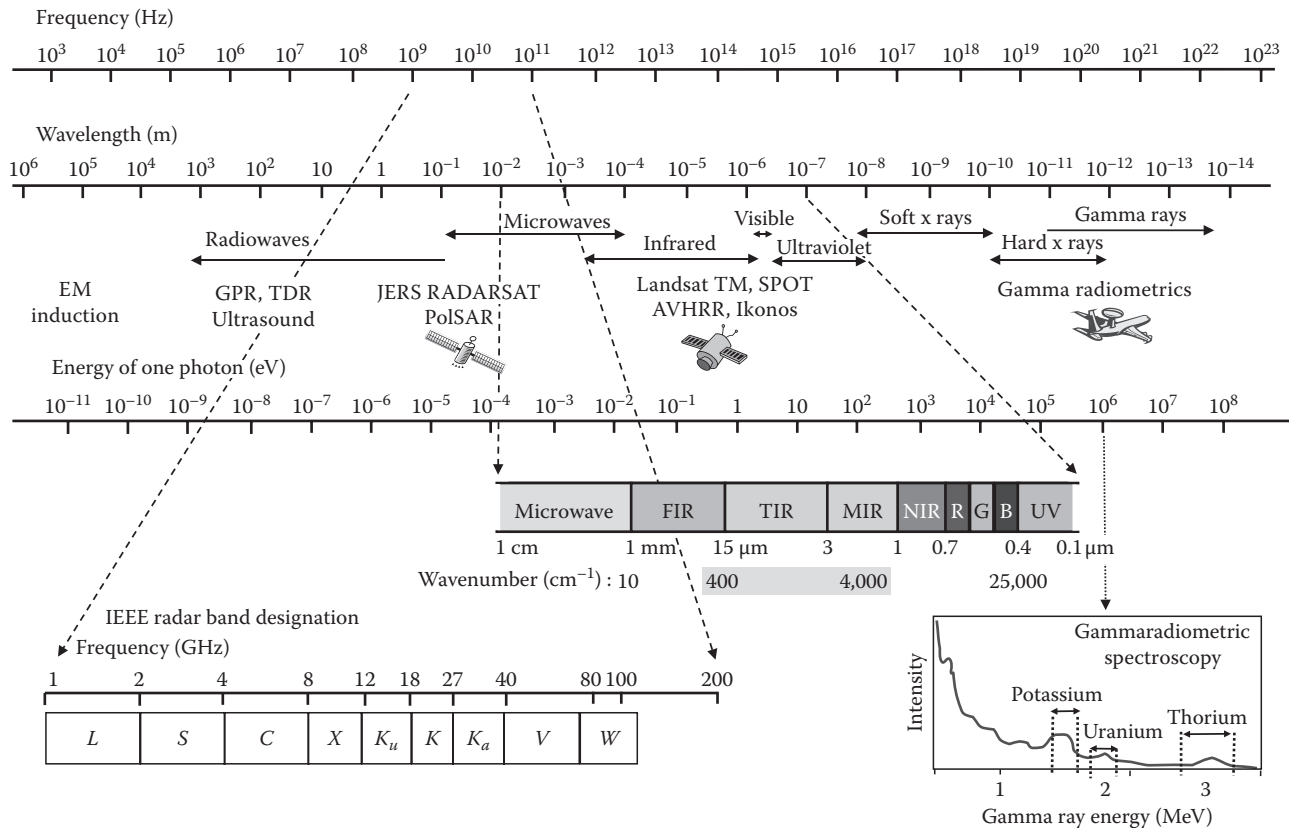

FIGURE 37.1   Electromagnetic spectrum showing various parts important for soil prediction.

37.4.1.1 Surface Multi- and Hyperreflectance

Hyperspectral sensors are those that measure a large amount of bands with spectral resolutions less than 20 nm (Palacios-Orueta and Ustin, 1998). In terms of mapping soil information, hyperspectral sensors have been found to be useful in mapping mineralogical features such as iron oxides (King et al., 1995), carbonates, and sulfates (Crowley, 1993). The amount of information generated by hyperspectral sensors poses computational problems in terms of extracting useful information in an efficient manner (see Section 37.3.3.1.2).

In a general way, the use of digital remotely sensing imagery for mapping soil has been problematic because vegetation cover obscures much of the soil response making it necessary to search for indirect evidence that may be visible at the surface (Campbell, 1987). Thus, remote sensing cannot be applied alone to soil studies (Lee et al., 1988). The use of proxies (such as topography, vegetation, drainage patterns) and field observations are important approaches for inferences about soil. In fact, mapping forest soil directly from remotely sensed data is difficult because of the complexity of environmental factors contributing to the spectral reflectance measured by a sensor. Nevertheless, forest soil was correctly mapped only where it was correlated with species or when vegetation is sparse or absent as a result of cultivation or drought (Post et al., 1994). The principal problems to soil delineation from such imagery, in addition to vegetation cover, are as follows:

- Soil moisture content can interfere with the spectral reflectance (specially in the infrared, thermal, and microwave regions (Obukhov and Orlov, 1964)
- Atmospheric effects (Cipra et al., 1980)
- Physical soil characteristics or disturbance patterns (Huette, 1988) and observation conditions (e.g., intensity and direction of illumination)

However, some research has been done in adjusting some of these drawbacks, by creating some indices with the main purpose of removing the effect of soil spectral influence (Huette, 1988) or even adjusting the images for vegetation interference, in combination with other information, and developing prediction models for improving soil mapping (Odeh and McBratney, 2000). In this way, Dobos et al. (2001) used satellite data complemented with DEM data, in order to correct the distortions caused by topographic variations of the landscape and provide additional data for soil–landscape modeling.

The presence of vegetation cover attenuates electromagnetic radiation at most wavelengths (Skidmore et al., 1997), and the reflectance at the soil surface does not always reflect soil variation at depth (Agbu et al., 1990). While infrared and visible sensors only measure surface characteristics, radar and gamma radiometry can provide spectral information beyond the vegetative cover and the soil surface.

37.4.1.2 Radar Attenuation

Information can be obtained by radar especially if there is a light-textured soil (low dielectric constant) over a heavier textured horizon (high dielectric) or a water table (very high dielectric),

and radar sensors can penetrate through soil to a depth that is equal to 10%–25% of their wavelength (Lascano et al., 1998). The longest wavelength radar sensor available from space platforms is 23.5 cm (L-band) on the JERS-1 satellite.

So far, we have only considered empirical methods where the multivariate prediction methods have been used to incorporate the remotely sensed imagery into prediction models. In the case of SAR, the backscattering signal is largely dependent on the dielectric properties of the media that is reflecting the signal, in the case of soil this is the volumetric moisture content (Moran et al., 1997). Therefore, physical models based on the theory of the diffraction of electromagnetic waves have been developed, an example being the integral equation model (IEM) (Fung et al., 1992). The model calculates a backscattering coefficient that is based on the following:

- Radar sensor configuration, for example, observation frequency, polarization, and incidence angle
- Surface characteristics, for example, roughness and dielectric properties

The model was adapted successfully to predicting soil moisture from bare soil surfaces by Altese et al. (1996). The presence of vegetation complicates matters, and it is currently too difficult to create models describing the interaction between soil–vegetation layers and microwaves in real world applications. Therefore, in the presence of vegetation, empirical approaches are required (e.g., Dobson and Ulaby, 1986; Wood et al., 1993).

37.4.1.3 Electrical Conductivity

Soil bulk electrical conductivity (or its reciprocal soil electric resistivity) reflects a combination of soil mineralogy, salts, moisture, and texture; hence, it is a good compound measure of soil. Two commonly used kinds of instruments are electromagnetic induction (EMI) and electrical conductivity/resistivity based on rolling electrodes (ECRE). The most widely used instruments for soil studies are the EMI devices from Geonics in Canada and two types of ECRE devices, a U.S. design (Lund et al., 1999) and a French one (Tabbagh et al., 2000). These instruments have been used extensively in precision agriculture for mapping soil types and properties (such as Bishop and McBratney, 2001; Sudduth et al., 2001; Anderson-Cook et al., 2002). Such proximal sensing offers the possibility of producing high-resolution maps of soil properties (D1 surveys of Table 37.1). Regression equations have been developed to predict moisture content, topsoil thickness, and clay content. EMI instruments can be placed in airborne platforms for catchment and regional mapping.

37.4.1.4 Gamma Radiometrics

Gamma-ray spectrometry (GRS) provides a direct measurement of natural gamma radiation from the top 30–45 cm of the soil (Bierwirth, 1996). A gamma-ray spectrometer is designed to detect the gamma rays associated with radioactive elements and to accurately sort the detected gamma rays by the respective energies (Grasty et al., 1991). Airborne radiometrics survey measures the radiation naturally emitted from the earth surface,

using gamma emitters like ^{40}K and daughter radionuclides of ^{238}U and ^{232}Th. K is a major constituent of most rocks and is the predominant alteration element in most mineral deposits. Uranium and thorium are present in trace amounts, as mobile and immobile elements, respectively. As the concentration of these radioelements varies between different rock types, we can use the information provided by a gamma-ray spectrometer to map rocks. Airborne methods (air gamma-ray spectrometer [AGRS]) provide valuable, systematic coverage of large areas, by providing information about the distribution of K, U, and Th that is directly interpretable in terms of surface geology. Nevertheless, AGRS is a surface technique only—interpretation requires an understanding of the nature of the surficial materials and their relationship to bedrock geology.

Although this technique has been employed for geological and mineral resource mapping for over 20 years, it has just become an interesting tool in soil science for detecting spatial variation of soil-forming materials (the *p* factor). It can also be considered as a direct, albeit compound, measure of the mineralogical and textural composition of the soil itself (*s*). Gamma-radiation data are usually provided in three channels—corresponding to spectral windows for K, U, and Th radiation. The apparent K concentration is likely to be most easily interpreted by pedologists. The value of gamma-radiometric data is increasing with the knowledge of their relation with soil-forming materials and when considered jointly with other information such as terrain models or aerial photography (Cook et al., 1996b) has become an important source of data for digital soil mapping. This technique has also been applied to estimate variation in surface soil moisture content (Carrol, 1981).

Rawlins et al. (2009) used airborne radiometric data and digital elevation data as covariates to map SOC distribution from an intensive survey in Northern Ireland. Radiometric data (K-band) and, to a lesser extent, altitude are shown to increase the precision of SOC predictions when they are included in linear mixed models of SOC variation. Division of the soil in Northern Ireland into three classes (mineral, organo-mineral, and peat) leads to a further increase in the precision of SOC predictions.

37.4.1.5 Existing Soil Class or Property Maps or Expert Knowledge

An existing soil map for parts of an area can be used to build a prediction model, or an experienced surveyor's expertise can be used to make simple rules that can be applied to a DEM, etc. These soil layers should be used as part of the information to build the new model—they should not really be the model itself. This is arguable, and there is further discussion in Section 37.5.12.

37.4.2 *"c"* Factor

Climate could be represented by mean annual temperature (*T*) and mean annual rainfall (*P*) and perhaps some measure of potential evapotranspiration (*E*). In the old literature, *P/E* was used to separate pedalfers from pedocals. A *P/E* ≪ 1 would imply semiarid conditions and precipitation of carbonates at depth.

Climate surfaces can be produced from meteorological stations interpolated by Laplacian smoothing splines (Hutchinson, 1998a, 1998b). This has been implemented in a program called ANUCLIM. The climate variables used are monthly mean values for minimum temperature, maximum temperature, precipitation, solar radiation, evaporation, and others. The climate surfaces can be used to generate secondary information, for example, bioclimatic parameters such as mean temperature of warmest period, precipitation of driest quarter, etc., which are useful in determining the climatic envelope for plant and animal species.

Published work suggests that remote sensing analysis can be used for estimating and mapping air temperature, soil moisture, and atmospheric humidity at regional to global scales. Air temperature can be inferred from normalized difference vegetation index (NDVI) data from the NOAA advanced very high-resolution radiometer (AVHRR). Several studies have shown that the surface albedo can be estimated using remote sensing data (i.e., Brest and Goward, 1987), and that net radiation can be calculated with sufficient accuracy (Kustas and Norman, 1996; Boegh et al., 2002).

37.4.2.1 Temperature

Surface temperature can be derived from remote sensing such as AVHRR, geostationary orbiting earth satellite (GOES) (Diak et al., 1998), and TIROS operational vertical sounder (TOVS) (Susskind et al., 1997). The TOVS has two sensors: the high-resolution infrared sounder and the brightness temperatures of the microwave sounding unit. These data can provide estimates of daily air temperature, humidity profiles, and surface temperature (Susskind et al., 1997).

Goetz et al. (1995) compared surface temperature derived from a multispectral radiometer (MMR) mounted on a helicopter (resolution ~5 m pixel), a C-130-mounted thematic mapper simulator (TMS) (~20 m pixel), and the Landsat 5 thematic mapper (120 m pixel). Differences between atmospherically corrected radiative temperatures and near-surface measurements ranged from less than 1°C to more than 8°C. Corrected temperatures from helicopter-MMR and TMS were in general agreement with near-surface infrared radiative, thermometer measurements collected from automated meteorological stations while the Landsat 5 TM systematically overestimated surface temperature.

37.4.2.2 Precipitation

Spatially distributed precipitation estimates can be derived from rainfall gauge measurements (interpolated using splines or other techniques) or by remote sensing. Records of gauge measurements of monthly precipitation are available throughout the entire twentieth century, while satellite estimates can provide monthly to hourly resolution since 1974. A review has been presented by New et al. (2001).

37.4.2.3 Evapotranspiration

Li and Lyons (2002) estimated regional evapotranspiration in central Australia, using limited routine meteorological data and the AVHRR data. Their model attempts to minimize the

difference between model-predicted surface temperature and satellite-derived temperature to adjust the estimated soil moisture. They suggested that radiometric surface temperature can be used to adjust simple water balance estimates of soil moisture providing a simple and effective means of estimating large-scale evapotranspiration in remote arid regions.

Boegh et al. (2002) used Landsat-TM data to estimate a composite evaluation of atmospheric resistance, surface resistance, and evapotranspiration. The input parameters were surface temperature, net radiation, soil heat flux, air temperature, and air humidity. The application of the technique in a remote sensing monitoring context was demonstrated for a Danish agricultural landscape containing crops at different stages of development.

37.4.2.4 Water Balance Components

The familiar water balance formulation is

$$P + I = \Delta S + E + T + R + D, \tag{37.20}$$

where

P represents precipitation
I represents irrigation
ΔS represents change in soil moisture
E represents evaporation
T represents transpiration
R represents surface runoff
D represents deep drainage

Precipitation and evapotranspiration can be estimated from remote sensing data. The most important component relating to soil itself is soil moisture, and much research has been focused on estimating spatially distributed soil moisture (see also Section 37.4.1.2). Jackson et al. (1996) gave an overview of remote sensing techniques for estimating soil moisture. Many studies have successfully demonstrated the use of infrared, passive, and active microwave sensors to estimate soil moisture (Hoeben and Troch, 2000). Microwave remote sensing of soil moisture is based on the soil's dielectric properties. The large difference between the dielectric properties of dry soil and moisture enables good calibration. The analysis is based on a model that simulates radar backscattering given known surface characteristics such as moisture and roughness. Passive microwave sensors have the advantage of less dependence on soil surface roughness. The main disadvantage with spaceborne sensors is that they produce coarse-resolution images. This problem is overcome in active microwave sensing through the use of SAR sensors (10–100 m). This has been used to monitor spatial and temporal soil moisture at catchment scale (10–1000 km^2) both in vegetated and nonvegetated areas (Lin et al., 1994; Su et al., 1997). Mancini et al. (1999) evaluated the use of multifrequency radar observations in the laboratory for estimating soil moisture.

Murphy et al. (2009) used a fine resolution digital elevation model to model soil moisture conditions in Alberta, Canada. They compared the compound topographic (wetness) index (CTI) with a new algorithm that produces a cartographic depth-to-water (DTW) index based on distance to surface water and slope. They found that their DTW model was closer to field-mapped conditions. All major wet areas and flow connectivity were reproduced and a threshold value of 1.5 m DTW accounted for 71% of the observed wet areas.

The soil moisture and ocean salinity (SMOS) satellite mission was launched in November 2009. It operates under the L-band (1.4 Ghz) and was designed to observe soil moisture over the Earth's landmasses and salinity over the oceans. A global soil moisture map with accuracy better than 0.04 m^3/m^3 can be produced every 3 days; however, the spatial resolution is very coarse (50 km or better).

37.4.3 *"o"* Factor

The "organisms" factor is a basic term encompassing biota composition or the flora and the fauna of specific region. Flora stands for vegetation description whereas the fauna means the animals, microorganisms, and even human activities. In pristine or newly developed environments, the "natural" vegetation class should represent some kind of equilibrium relation with soil type. In that case, the vegetation is an indicator of the underlying soil type. However, "natural" vegetation is quite uncommon to find since it has frequently been disturbed by human activities (generally called "land uses"). Vegetation and land uses, generally provided by land cover analysis, can then be either an indicator of the underlying soil type or a determinant of the soil type. After defining land use and vegetation, their source of data, their relationships with soil, as determinant or indicator, can be explained.

37.4.3.1 Land Use Information and Relationships with Soil

Three kinds of land uses are generally distinguished: settlements (cities, mines, industrial areas, etc.), agricultural activities (arable lands, pastures, plantation, etc.), and natural or seminatural areas (natural forests, natural grasslands, bare rocks, and water). This information is generally provided by remote sensing images providing access to land cover. While settlements are usually easy to distinguish from above, it is often difficult to punctually differentiate agricultural pastures from natural grasslands, bare soils of arable lands from bare rocks, and plantations, or harvested areas, from natural forests. To this end, historical ground inventories, long-term imageries, or time-series analysis of remotely sensed images should be included in the mapping process of land uses. In that sense, land use information can be crucial for assessing the "*a*" (or time/age) factors. Schulp and Veldkamp (2008) made an inventory of historical land use. They showed that coarse resolution (500 m) historical patterns can explain up to 75% of the soil organic matter variability, whereas actual land use only explains 2% of the variability. Moreover, the IPCC report (2006) provided increase or decrease coefficients of soil organic carbon stocks according to land use or land cover conversions. This information is then crucial at the global level.

37.4.3.2 Vegetation Definitions, Data, and Relationships with Soil

Vegetation information is one of the most used "scorpan" factors along with landform. It is usually obtained from intensive ground survey and/or from visible and infrared reflectance by remote sensing being enhanced more recently by microwave imagery (Clevers and van Leeuwen, 1996). It comprises biomass deriving from agricultural activities and from natural or semi-natural areas. Three kinds of components should be described: the type or species of vegetation, the intensity of vegetation or biomass quantity, and the area covered by vegetation usually called "vegetation coverage."

37.4.3.2.1 Vegetation Type or Vegetation Land Cover

The vegetation type is usually assessed by doing a classification of remote sensing multispectral imagery on the spectral signatures of the image pixels (Figure 37.2). Vegetation is mainly discriminated by wavelengths at 0.55 μm in the green region, at 0.67 μm in the red region, and at 0.87 μm in the NIR region (Bunnik, 1978). Classification can be supervised using ground measurements or unsupervised when applied directly on the pixels. For tree species, and for very fine resolution, when there is a priori knowledge of the species, it is possible to use airborne laser scanning (Lidar) for differentiating the trees according to their morphological attributes such as tree height, stem diameter, and crown radius (Ørka et al., 2009). Carré and Girard (2002) used information on spectral bands for assessing the spatial distribution of soil types using linear regression. Hansen et al. (2009) used multispectral data and topography for assessing soil–landscape units.

37.4.3.2.2 Vegetation Intensity or Biomass Quantity

According to vegetation species and its state of development, chlorophyll absorbs more or less solar energy. The absorbed or transformed energy that represents biomass production is usually modeled by combining different visible and near-infrared bands. The produced indices are then not only indicating vegetation types but also primary production. These indices are NDVI (Rouse et al., 1973), EVI (Waring et al., 2006), and FAPAR (Gobron et al., 2000).

Furthermore, concerning crops, the use of yield monitors on harvesting machines also provides a source of spatial biomass information (Stafford et al., 1996). Bishop and McBratney (2001) used yield-monitored wheat yield to aid in the prediction of soil clay content. Yield-monitored data are currently useful for D1 mapping (Table 37.1).

37.4.3.2.3 Vegetation Coverage or Area Covered by Vegetation

The area covered by vegetation, whatever the vegetation type, can be assessed using the indices explained before. This allows for usually discriminating bare soil from lands covered by vegetation and thus assess potential risks related to soil vulnerability like wind erosion (Skidmore, 2006). Many publications describe the role of vegetation on different soil processes, on the spatial distribution of soil at different scale, and for different time periods (months, year, and decade). It is then important to take this factor into account for digital soil mapping.

37.4.4 "r" Factor

Topography has long been recognized as one of the main soil-forming factors (Jenny, 1941). Milne (1935) introduced the concept of a catena to describe a predictable and recurring pattern of soils that occupied a topographic sequence along a hillslope from divide to channel to divide. Aandahl (1948) was the first to quantitatively relate landscape attributes to soil properties. Troeh (1964) fit a cylindrical parabola to contour lines to derive
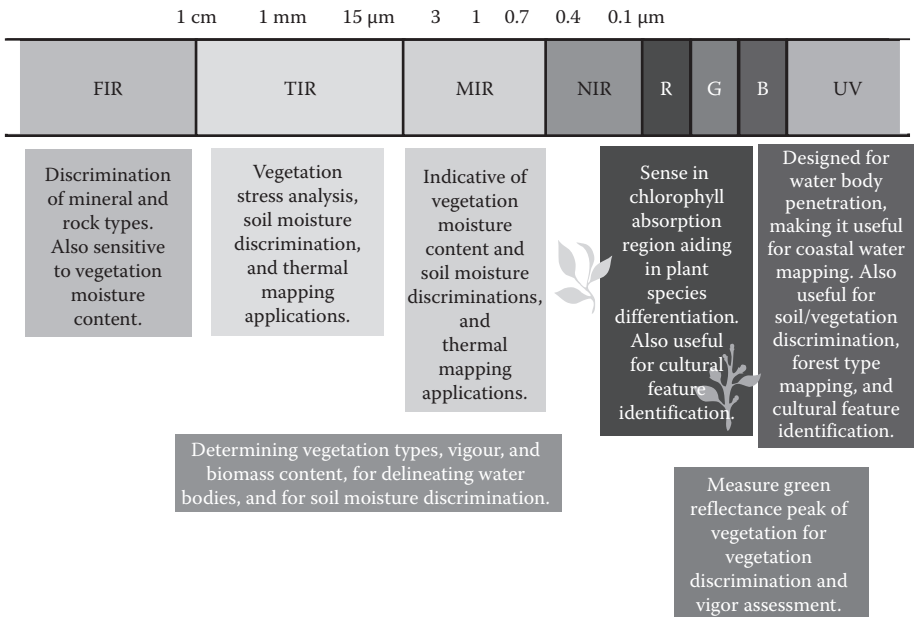

FIGURE 37.2 Ultraviolet-infrared spectrum showing important parts for sensing vegetation indices.

slope and curvatures. These landform parameters were derived to correlate to soil drainage classes.

Almost all DSM studies reflect a recognition that changes in the value of soil properties or the classification of soils occur in response to environmental gradients and that many environmental gradients of interest are strongly related to elevation, moisture, and energy gradients that are defined and controlled by local and regional topography (Milne, 1935; Burrough et al., 2000; McKenzie et al., 2000).

The principal assumption underlying the soil–landform paradigm is that the distribution of soils in the landscape is predictable and that, at least at a local scale, this distribution is largely controlled by how topography influences the distribution and redistribution of moisture, energy, and materials in the landscape. Terrain derivatives computed from digital elevation data can provide useful measures for predicting and quantifying how topographic controls influence this distribution and redistribution. They therefore represent a primary source of explanatory variables for most efforts to implement automated predictive soil mapping. Many would argue that digital terrain modeling is the most useful and quantitatively developed factor for predicting soil attributes and soil classes, at least at a local (hillslope) scale (McKenzie et al., 2000).

Topography has increasingly come to be described and quantified by means of digital elevation models (DEMs). Initially, DEMs were produced by digitizing existing contour maps or by extracting spot elevations and stream-line data from traditional land surveys and then interpolating these contour or spot elevations to a regular grid (or using them to create a triangular irregular network or TIN). More recently, DEMs have been produced more rapidly, cheaply, and at a finer spatial resolution by means of automated stereo autocorrelation of stereo image pairs obtained from airborne or satellite sensors, by interpolation of dense ground point observations obtained from vehicle-mounted high-resolution GPS receivers, or from very fine resolution clouds of point elevation data obtained using airborne lasers (LiDAR) or synthetic aperture radar (IfSAR or InSAR). See Nelson et al. (2009) for a thorough review of DEM production methods and sources.

The use of elevation data in digital form involves quantitative parameterization of the surface model or the numerical description of continuous surface form (Pike, 1988; Wood, 1996). Pike et al. (2009) observed that geomorphometry is fundamentally concerned with the quantitative measurement and extraction of both land surface parameters (LSPs) and land surface objects (LSOs). An LSP is a descriptive measure of a surface attribute (e.g., slope, aspect, and wetness index), usually arrayed as a continuous field of values and usually organized as a raster image or map. An LSO is a discrete surface feature (e.g., a watershed line, cirque, alluvial fan, and drainage network) that is usually represented as a vector object defined by points, lines, or polygons (see also Deng, 2007).

Parameterization means to quantitatively measure continuous properties of a landscape that can be used to describe form (Wood, 1996). Geomorphometry involves quantification of form but can also include consideration of different measures of pattern, texture, and hydrological, gravitational, or spatial context, as well as the extraction of geomorphologic objects. A wide range of LSPs can be extracted from a DEM, such as altitude, slope, aspect, different curvatures, upslope contributing area, and compound topographic index (CTI). Similarly, DEMs can be processed to identify and delineate a wide range of discrete LSOs.

For discussion purposes, LSPs have frequently been grouped into primary and secondary topographic attributes (Moore et al., 1993; Wilson and Gallant, 2000). Primary attributes are measures of a single aspect of the land surface whose calculation does not require combination of two or more separately computed values. Primary attributes are often further differentiated into local attributes, such as slope, which are computed within a neighborhood window of fixed (but now increasingly of variable) dimensions and regional attributes, such as contributing area, whose calculation extent is not predetermined or fixed but varies according to conditions imposed by a particular set of land surface data. Secondary terrain attributes are any attributes (such as wetness index) whose values are computed using a combination of two or more primary attributes (Wilson and Gallant, 2000).

37.4.4.1 Primary Terrain Attributes

Primary attributes have been used successfully in numerous studies to predict different soil attributes and classes. Maintaining the above distinction of local versus regional primary attributes, we observe that local attributes mostly describe the height, orientation, and shape of the land surface at a point or, more correctly, within a region defined by a rectangular (or sometimes circular) window surrounding a point. The main primary local attributes are elevation, slope, aspect, and various measures of curvature.

Elevation in its raw form mainly influences regional and local climate through its effect on surface temperature and precipitation (Deng, 2007). Elevation has also been used in numerous DSM studies as a local indictor of relative slope position within restricted areas (Shi et al., 2004). Slope and aspect describe the orientation of the local land surface within a neighborhood window. Curvatures primarily describe the shape of the local land surface within the window.

Shary et al. (2002) highlighted the importance of recognizing that many terrain attributes are field specific, insofar as they are computed with reference to a particular vector field, such as the gravitational field. Slope, aspect, profile and plan curvature are all computed relative to the orientation of the normal to the gravity vector. When interpreting these measures, the implicit assumption is that the current orientation of the land surface reflects the conditions that existed over the period of time that a soil was formed. For example, a closed depression (nondraining) can change to an open depression (draining) if the land surface tilts at some point in time after the formation of the closed depression.

Evans (1972, 1998) provided an overview of primary terrain attributes in relation to their geomorphologic meaning. Slope is

frequently interpreted in terms of the amount of energy associated with a particular slope gradient that is available to drive soil-forming processes such as runoff, erosion, and deposition. Aspect, in combination with slope, is often interpreted in terms of exposure to incoming solar radiation, drying prevailing winds, or accumulation of snow or sediments in sheltered lee slopes. Curvature is generally interpreted in terms of how local shape affects surface runoff, erosion, and deposition. Areas that are convex in profile or plan are considered likely to shed water, to cause divergence of surface flow (in plan) and acceleration of surface flow (in profile), and to consequently be drier than normal sites. Areas that are concave in plan (across slope) are considered to favor convergence of surface flow and development of moister conditions. Areas that are concave in profile (downslope) are considered to favor deceleration of surface flow and deposition or accumulation of both moisture and sediments. Curvature also has often been interpreted as an indication of relative position on a hillslope (see Pennock et al., 1987). Areas that are convex in profile were interpreted as indicative of crests and shoulders while areas concave in profile were interpreted as indicating decelerating positions in toe slopes. Areas that are planar in profile have been associated with back slopes.

The majority of DSM studies appear to have preferred to make use of the most commonly computed local terrain attributes slope, aspect, profile curvature, and plan curvature computed using the most commonly available algorithms. Shary et al. (2002) described instabilities in the calculation of plan curvature and suggested that it not be used and instead replaced with its close equivalent of horizontal curvature. Many subsequent studies adopted the use of either horizontal or tangential curvature in preference to plan curvature to characterize curvature in the across slope direction (Schmidt et al., 2003; Schmidt and Hewitt, 2004).

Local terrain attributes are typically taken to provide an indication of the shape or orientation of the land surface only, that is, its form. However, the increasing trend to compute these variables over a range of grid resolutions and window dimensions produces information that can be interpreted to quantify characteristics of the land surface that describe texture, pattern, scale, and context. The magnitude and frequency of variation in these local terrain attributes within expanding neighborhood windows can provide useful measures of pattern and texture in the landscape. It can also help to establish the characteristic distances over which variation in these properties occurs. This can be interpreted as a measure of the size and scale of the landscape. Finally, multiple measures of slope, aspect, and particularly curvatures within expanding calculation windows can be interpreted in terms of the position or context of a grid cell in the landscape (e.g., a ridge in a pit located on a larger ridge) (Wood, 1996).

Many primary regional topographic attributes provide information about some aspect of topographic or hydrological position or context. The most commonly used primary regional attribute is upslope contributing area or drainage area also rescaled to compute specific catchment area (SCA). Calculation of this attribute involves tracing down flow directions and accumulating a sum for the area upslope of every cell that contributes drainage to each downslope cell. This calculation can involve using a single steepest flow direction algorithm (the so-called D8 algorithm) to assign all flow from each cell into its steepest downslope neighbor, or it can utilize one of several different algorithms that partition flow from each cell according to some estimate of the proportion of flow that might logically be expected to pass from a given cell into each of its neighbors of equal or lower elevation. This attribute provides a measure of relative position in the landscape within the context of simulated hydrological flow of surface water.

Many primary regional attributes represent efforts to capture the relative position in the landscape of a point or the amount of relief (and therefore potential energy) associated with the area upslope (or downslope) of a point with respect to some other significant reference point, such as a channel or a ridge. Measures of relative landform position have been reported that are based on calculations of the variation in elevation (Pike, 1988) or mean slope (Fels and Matson, 1996) within fixed windows of various dimensions. Wilson and Gallant (2000) identified a number of measures of variation in elevation within circular windows centered at a grid cell.

Skidmore (1990) introduced one of the first algorithms for computing a measure of relative slope position. It involved first classifying cells as either ridges (convex) or channels (concave), based on their local surface shape. An expanding search window centered on each grid cell was then used to determine the Euclidian distance from each cell to the closest ridge and channel cells. This approach has the potential disadvantage that the closest channel or ridge cell may not be the one to which a given cell is connected by a path of surface water flow (see Rennó et al., 2008). MacMillan et al. (2000a, 2000b) implemented an algorithm for computing a variety of different measures of absolute and relative landform position and relief that was based on tracing both upslope and downslope along simulated paths of surface flow to locate and measure the horizontal, flow length, and vertical distances to critical cells designated as belonging to pits, peaks, channels, or divides. This approach has been proposed or implemented by others as well (Bell et al., 1992; McSweeny et al., 1994; Bui and Moran, 2001). Interesting and useful alternate measures of relative position in the landscape have been described in terms of valley bottom index (Gallant and Dowling, 2002) and openness (Yokoyama et al., 2002). Relative landform position tends to be a key conceptual consideration in many human-devised systems of soil or landform mapping, and measures of landform position and context have increasingly come to be used as explanatory variables in digital soil mapping (Bell et al., 1992; Bui and Moran, 2001, 2003; Schmidt and Hewitt, 2004; Schmidt and Andrew, 2005; Schmidt et al., 2005).

See Wilson and Gallant (2000) for a review of primary and secondary topographic attributes including various measures of slope and area computed for cells upslope of a central cell that contribute flow into it (contributing area) and cells downslope of a given cell into which it contributes flow (dispersal area).

37.4.4.2 Secondary Terrain Attributes

Secondary terrain attributes are computed from the primary attributes. These have been described in detail by Wilson and Gallant (2000). These attributes usually combine two or more primary attributes to characterize the spatial variability of specific processes in the landscape. The most widely used is CTI or also called wetness index:

$$\text{CTI} = \ln\left(\frac{A_s}{\tan\beta}\right), \qquad (37.21)$$

where
A_s is the upslope area
β is the slope

Wilson and Gallant (2000) also provide routines for the calculation of erosion, solar radiation, and dynamic wetness indices.

Wilson and Gallant (2000) list and document a number of secondary topographic indices including three different versions of the wetness index or CTI, a stream power index (SPI), a sediment transport capacity (STC) index, a channel initiation (CI) index, and several radiation and temperature indices. All of these can prove useful in efforts to automatically classify landform-based terrain elements. The secondary topographic attribute that is most widely reported as being used in efforts to automatically classify landform entities is the wetness index or CTI (Beven and Kirkby, 1979; Moore et al., 1991a, 1991b, 1993; Wilson and Gallant, 2000).

37.4.4.3 Terrain or Landscape Classification

In contrast to the preceding discussion of continuous LSPs, landform classification locates and extracts identifiable LSOs from digital elevation data. These objects occur across a range of scales, and their recognition can involve varying degrees of subjectivity.

Within the context of digital soil mapping, there have been three main areas of application of landform classification. The first has been the identification of critical points, or surface-specific points (Peucker and Douglas, 1975), based largely on local surface shape. The second has been the identification and classification of areas of relatively homogeneous terrain conditions, referred to as landform elements, which are usually identified as components of a hillslope along a sequence from crest to channel. These landform elements are usually described as being occupied by a single soil series or a narrow range of similar soil series. The third application has involved recognition of regional landform patterns, which are typically described as being occupied by repeating combinations of different soils, such as soil associations or complexes. Each of these three major approaches is reviewed in brief as follows.

Recognition of critical points, or surface-specific points (Peucker and Douglas, 1975), based primarily on local surface shape has been a key component of automated landform classification since its inception. Pike et al. (2009) cite Cayley (1859)

as having been the first to identify and describe the conceptual importance of these points in describing landforms. Critical points and lines identify local maxima and minima in the landscape and help to establish horizontal and vertical scale and context for landform classification. Their identification is relatively objective and crisp but some ambiguity occurs when local surface shape is not pronounced enough to identify pits, peaks, channels, or ridges definitively. Ambiguity also exists with respect to the scale at which these objects are identified. Wood (1996) showed how extraction of these critical objects within multiple windows of expanding dimensions could result in classification of the same point as, for example, a pit within a peak within a pit. This multiresolution analysis approach has been increasingly recognized as a valuable way to assess context and form across multiple scales (Deng, 2007).

The surface-specific points and lines described above delineate a mesh, or framework, of hillslope entities bounded by ridge lines and course lines. The intersection of ridge lines and course lines delineates areas that represent individual hillslopes that run from local divides down to local channels. Pike et al. (2009) cite Maxwell (1870) as having been the first to observe that the area enclosed by these lines was part of both a *hill*, whose lines of slope run down from the same summit and a *dale*, whose slope lines run down to the same pit. These semantics emphasize the relationship of these points and lines to patterns of local surface water flow. Not surprisingly then, essentially the same points and lines can be extracted by computing networks of simulated surface water flow and extracting pits, peaks, channels and divides from the flow networks. This has been done by Band (1989a, 1989b, 1989c) who used pixel to pixel hydrological connectivity to define fundamental terrain units based on the intersection of complementary divide and channel networks. Hillslopes were defined as the areas between a channel segment and its associated drainage divide.

Identification of these pits, peaks, channels, and ridges plays a significant role in other aspects of landform analysis and classification. First, several useful LSPs, such as horizontal distance or change in elevation to a ridge or channel, need to first identify these significant points and lines in order to be computed. Second, these critical points and lines establish the context and boundaries within which subsequent efforts to identify and classify landform elements operate. Third, many of the reported approaches for classifying landform patterns make use of information on the scale and spacing of ridges and channels to establish measures of context and pattern used in classifying landform patterns.

A landform element is a subcomponent of a landform pattern. Landform elements may be conceptualized as consisting of portions of a landform pattern that are relatively homogeneous with respect to shape (profile and tangential curvature), steepness (gradient), orientation or exposure (aspect or solar radiation), moisture regime, and relative landform position (e.g., upper, mid, or lower). Speight (1990) recognized more than 70 types of landform elements and gave as examples cliffs, foot slopes, and valley flats. Dikau (1990) recognized simpler form units, or

form facets, nested within form patterns, that were described by inherent shape-based attributes such as curvature, gradient, and aspect. Dikau (1989) differentiated form elements with homogeneous plan and profile curvature from even more homogeneous form facets that had uniform gradient, aspect, and curvature.

The majority of applications of landform classification in digital soil mapping have represented efforts to identify and extract landform elements arranged in a topographic sequence along hillslopes from crest to channel (Park et al., 2001). These classifications were usually intended to build upon conceptual classifications of hillslopes into summits, shoulders, back slopes, foot slopes, and toe slopes as proposed by Ruhe (1960) or the 9 unit classification of hillslopes proposed by Dalrymple et al. (1968) and Conacher and Dalrymple (1977) or the 10 types of topographic landform positions described by Speight (1990).

Classification of landform patterns is generally more challenging than classification of landform elements along a single hillslope. Not surprisingly then, there have been fewer approaches described for automatically classifying landform patterns and fewer applications of landform pattern classification in digital soil mapping. Classification of landform patterns has mainly been used to establish domains, or regions, over which a particular recurring pattern of soils and terrain (e.g., a soil association) is observed to occur and within which a particular set of classification rules for a limited number of soil classes is thought likely to apply (Dobos et al., 2000, 2001; Bui and Moran, 2001, 2003; Henderson et al., 2005; Bui et al., 2006; MacMillan et al., 2007).

The de facto standard for classifying landform patterns is Dikau et al.'s (1991, 1995) computerized implementation of the manual system of landform classification proposed by Hammond. The method of Dikau et al. (1991, 1995) follows Hammond (1954, 1964) in recognizing, within a window of fixed dimensions, 4 classes of proportion of gentle slopes, 6 classes of relative relief, and 4 classes of profile type, which, when combined, lead to 22 possible landform subclasses (Brabyn, 1998). These are commonly regrouped into 24 landform classes and 5 major landform types of Plains, Tablelands, Plains with Hills or Mountains, Open Hills and Mountains, and Hills and Mountains (Bayramin, 2000).

Brabyn (1997, 1998) noted two problems with implementation of Dikau's methods. The first was a pattern of progressive zonation that developed in transition zones between areas of high and low relief. The second was that the methods classified areas with quite different macro landform into the same class. Guzzetti and Reichenbach (1994) also identified concerns with imprecise boundary locations and mixed classification and tried to address them by developing and applying a different approach that combined automated classification of DEM data with manual interpretation and digitizing of clear, sharp final boundaries.

There have been relatively few other approaches reported for classifying landform patterns from digital elevation data. Bui and Moran (2003) used the computed values of total distance from ridge to channel to produce a map of relief classes (<9, 9–30, 30–90, 90–300, and >300 m) to define physiographic regions that exhibited signature soil patterns. Similarly, MacMillan et al. (2007)

used the total relief computed for individual hillslopes to define classification domains of different relief within which fuzzy rules were used to classify ecological site types along toposequences from crest to channel.

Iwahashi and Pike (2007) proposed an alternative approach for classifying landform types at multiple scales that is very simple, efficient, and effective. This approach uses only three inputs of slope gradient, local convexity, and surface texture (feature frequency and spacing). The method is locally adaptive as it does not use fixed threshold values to distinguish, for example, steep from gentle slopes or convex from concave areas. Instead, the method uses a nested-means approach in which the mean value within a window of 10×10 cells is used as a threshold to differentiate steep from gentle or convex from concave. The method produces 8, 12, or 16 classes depending upon whether the gentle slope class is further subdivided once (into two subclasses) or twice (into four subclasses). This method delineates undefined classes whose local meaning and description must be assigned by a postclassification review and interpretation. Examples presented for DEM grids of 1 km, 270 m, and 55 m illustrated that the method produced meaningful and interpretable results across a range of scales.

In landform classification, as in conventional image analysis, there has been a recent trend toward investigating and adopting methods that involve image segmentation, or object recognition, that classify the object rather than classifying an individual pixel. A multiscale image segmentation approach is advocated by Burnett and Blaschke (2003) and Drăgut and Blaschke (2006) because "a multivariate analysis of pixels does not include topological relationships of neighborhood, embeddedness, or shape information."

In contrast to the bottom-up agglomerative approach of per-pixel classifications, such top-down divisive approaches identify significant discontinuities or boundaries at locations of maximum change in attributes. These boundaries define the spatial extent of objects whose attributes can then be determined and used to classify the entire object. The object's attributes can include context, shape, and topological information, such as adjacency or connectivity in addition to typical measures such as slope, aspect, or curvatures.

In conclusion, it is useful to reflect upon the observation of Dehn et al. (2001) that landforms are described mainly in two different ways (i) based solely on their geometry or (ii) based on semantics used to express and capture subjective conceptual mental models. Deng (2007) also differentiated five classes of landform entities according to the degree of vagueness or subjectivity involved in their definition and recognition. Landform entities that were discrete, clearly identifiable, and objective were classified as bona fide objects or prototypical objects. Landform entities whose definition and recognition involved a degree of vagueness and subjectivity were classified as semantic or fiat objects, landform classes, or multiscale objects.

37.4.5 *"p"* Factor

Acquisition of suitable information about parent material for use in digital soil mapping often presents a challenge. The attributes

of parent material that are most often of interest are depth or thickness (deep or shallow), texture (e.g., coarse, medium, or fine), lithology or mineralogy (e.g., base status), age, and mode of deposition.

At present, the most common source of information about parent material for use in soil mapping studies has come from digitized versions of published geological or soil maps. Among geological maps, those that focus on lithology, rather than age or stratigraphy, tend to be more useful for soil prediction. Reinterpretation or recompilation of existing geological maps is often necessary to emphasize lithology in place of the original stratigraphic focus (Lawley and Smith, 2007).

Given that existing geological maps are often unavailable or inadequate, many studies have sought to identify alternative approaches for obtaining or inferring the main parent material attributes of interest. Of these, interpretation of airborne gamma-radiometric data in terms of regolith properties seems to have generated the greatest interest and success (Cook et al., 1996; Wilford et al., 1997, 2001; Pickup and Marks, 2000, 2001; Hardy, 2004; Rawlins et al., 2009). Radiometric data highlight differences in the parent material that can be interpreted in terms of depth to bedrock, mineralogy and lithology, age and mode of deposition. Clear distinctions have been made between transported and in situ regolith and between rocks with varying degrees of weathering. Radiometrics can provide an indication of the degree of erosional and depositional activity, which allows actively eroding surfaces with shallow regolith to be separated from stable surfaces with deeper regolith. Differences in mineralogy show up as different colors and patterns in ternary radiometric images, and these patterns can identify the spatial extent of areas of different lithologies (e.g., base rich versus acidic) and mineralogies (potassium rich granite versus depleted quartz sands). Degree of weathering, related to geomorphic age and amount of transportation, can often be interpreted from radiometric data, with highly weathered materials emitting low responses and recently exposed or weathered materials emitting high levels of gamma radiation. Such distinctions can be used to separate, for example, recent eolian deposits from old, well-sorted and intensely weathered fluvial deposits (Bierwirth and Brodie, 2005). Typically, radiometric data are most effective when interpreted in combination with additional data sources, particularly terrain data as extracted from a DEM (Cook et al., 1996b; Wilford et al., 2001).

Radiometric data are not readily available for many areas. In such cases, researchers have investigated a number of different options for inferring specific attributes of the parent material by interpretation of airborne and satellite imagery of various types and resolutions. The use of imagery for inferring attributes of parent material typically focuses on identifying specific parent material characteristics that are associated with distinct and highly contrasting patterns in the imagery. For example, areas of salinity, affected by gypsic or calcic surface materials, often contrast sharply with surrounding unaffected areas and can be isolated using imagery alone or using a combination of imagery and digital elevation data (Fernández-Buces et al., 2006;

Nield et al., 2007; Rodriguez et al., 2007; Boettinger et al., 2008). Similarly, areas of organic soils (Connolly et al., 2007), wetlands (Peng et al., 2003; Landmann et al., 2006; Wright and Gallant, 2007), and African dambos (Hansen et al., 2009) often contrast sharply with their surroundings and can be isolated using multidate, multispectral digital imagery or multidate SAR (Townsend and Walsh, 1998; Townsend, 2001; Lang et al., 2008; Liu et al., 2008a).

Ground-based measurements of soil's apparent electrical conductivity (ECa) have been widely used to infer properties of the parent material for relatively small areas of field to regional extent. Here the assumption has been that electrical conductivity reflects the combined influence of soil texture, mineralogy and moisture content, and areas of different texture will exhibit interpretable differences in electrical conductivity (Liu et al., 2008b). Additionally, other natural fields of the earth, gravitational, magnetic (Galdeano et al., 2001), and electromagnetic (Beard, 2000), have also been used to provide information on underlying geological structure.

Recent developments have seen increasing interest in modeling spatial variation in parent material classes or parent material properties in a manner that is analogous to *scorpan* modeling of soil spatial patterns (Hardy, 2004; Lawley and Smith, 2007). Pain et al. (2001) and Clarke and Pain (2004) described regolith–landform studies in Australia as being concerned with understanding the 3D distribution of regolith materials, and its representation in map form, rather than vertical profiles or cross sections. They note that regolith is very closely associated with landforms and once the interrelationships between regolith and landforms are understood, landforms can largely be used to predict regolith patterns. Thus, landforms are used as a proxy for the largely hidden regolith. They report that landforms can be identified and delineated in a number of different ways, including conventional air photo interpretation, manual or automated interpretation of remotely sensed imagery, automated analysis of airborne radiometric data, and 3D mapping of the regolith using such techniques as air- and ground-based electromagnetic, ground-penetrating radar, and shallow seismic magnetic surveys. They inferred regolith properties such as texture, mineralogy, depth, and geochemistry by interpretation of the mapped landform classes. It is anticipated that an increasing number of digital soil mapping studies will utilize predictive modeling methods to produce separate predictions of parent material attributes such as depth, texture, and lithology for subsequent use in digital soil mapping.

37.4.6 *"a"* Factor

a represents age or elapsed time. This will give limits on how long pedogenesis has been occurring and should differentiate soil classes and properties. One useful estimate of *a* is the age of the ground surface, which may be very old indeed (Twidale, 1985). Alternatively, *a* can be represented by the age of the material in which soil has developed, suggesting that the *scorpan* approach will not deal well with polycyclic soils. It is theoretically likely

that soil development will follow some logarithmic or square-root time relationship, suggesting more need to differentiate between younger materials than older ones. Schaetzl et al. (1994) discuss the form of soil chronofunctions.

Geomorphologists and stratigraphers can presumably draw maps of *a* independent of soil maps. In fact, such "gues(s) timated" maps along with an estimate of uncertainty could be used to represent this factor. There are methods for soil and material dating of course, for example, ^{14}C, $\delta^{18}O$, thermoluminescence (inter alia, Matt and Johnson, 1996), and $^{40}Ar/^{39}Ar$ (inter alia, Van Niekerk et al., 1999). None of these are, as far as we are aware, capable of scanning and producing full coverage in a true remote sensing fashion. Ground electromagnetic methods have been used for stratigraphic mapping, for example, Sinha (1990). *a* remains difficult to characterize well. Indeed, it seems that more than any factor expert knowledge is still needed to derive *a*. Considerable advances in technology and knowledge are needed.

Scull et al. (2005) used decision tree in the southwestern Great Basin area for predicting soil class. Age of landform is inferred from remote sensing data, which can indicate surface age by detecting desert varnish. They found that the age of landform covariates is an important predictor in the decision tree model. Noller (2010) presented an explicit use of the age factor in digital soil mapping. He estimated geochronological information for Malheur County, Oregon based on independently compiled Quaternary geological maps, age point data, and remotely sensed data. These data were incorporated in decision tree analysis. Expert soil survey maps are used as reference in making predictions with and without implicit or explicit age information. Addition of geochronological data produces digital soil maps that are most closely aligned with expert maps. Improvements in the map are greater in areas where the age stratification of the landscape is not obvious to the soil surveyor.

37.4.7 *"n"* Factor

As was discussed in Section 37.2, soil can be predicted from spatial coordinates alone. Obtaining these is now much easier due to the advent of GPS with 5 m accuracy receivers costing less than $1000. This may indeed reflect some other environmental variable such as climate, and because of this, it can be argued that *n* is not really a factor, but simply organizes the coordinates is a simple way to ensure that spatial trends not included in the other environmental variables are not missed. Therefore, *n* could also be described by some linear or nonlinear (nonaffine) transformation of the original spatial coordinates, for example, a new coordinate could be the closest distance of each location to the coast (Webster, 1977, p. 201), or contextual representation of space, for example, distance uphill from the nearest discharge area (Bui and Moran, 2000), distance from a watershed, distance from roads, distance from a point source, etc. This factor is potentially a valuable yet cheap source of environmental information and should never be disregarded.

37.5 Applications

Having presented a review of what has been done by others and having suggested and reviewed a generic predictive model and potentially useful environmental data layers, we now put this into a framework for soil mapping based on *scorpan* and Soil Spatial Prediction Functions (SSPFs) and spatially autocorrelated errors (**e**). The *scorpan*-SSPFe approach is now outlined with some brief discussion of each of the steps. Its uses, problems, and other implications of the *scorpan*-SSPFe approach are discussed subsequently.

37.5.1 Summary of the Scorpan-SSPFe Method

The *scorpan*-SSPFe method essentially involves the following steps.

   i. *Define soil attribute(s) of interest and decide resolution* ρ *and block size* β

These are the design specifications for the survey. Define soil attribute(s) of interest, that is, a soil property or set of soil properties and/or a set of soil classes, usually from some predefined classification system. The resolution may be defined by the resolution of the environmental variables, for example, 30 m Landsat pixels, but should be a design specification from the intended use of the information. Referring back to Table 37.1, we believe that the methodology discussed here is most appropriate for D3 surveys, therefore pixel or block size ρ, is equal to pixel spacing β and will be in the range 20–200 m. (The linking of ρ and β is a simplification and is a point that requires further study—see Bishop et al. (2001) for further discussion.) At this stage, the uncertainty limits that can be tolerated may be also be defined.

   ii. *Assemble data layers to represent Q*

Assemble the data layers with consideration of the number of layers describing each factor and any prior evidence as to the importance of each factor. This was discussed in detail in Section 37.4. At this stage, we do not know the relative importance of the data layers. Balance is probably important. At this phase of development, because of the relative availability of DEMs, it would be easy to obtain 15 or 50 terrain attributes (*r*), for example, as described in Shary et al. (2002), and rather difficult to represent *a* or *c*. An attempt should be made to represent all the factors however.

   iii. *Spatial decomposition or lagging of data layers*

This is suggested as a step because it is felt that predictions might be scale dependent, and it is important to find the appropriate spatial associations. This can be achieved either by a wavelet decomposition (e.g., Mendonça-Santos et al., 2006), or geostatistically. The geostatistical approach involves modeling the correlation structure in the imagery by decomposing the variogram into independent spatial components, and then taking each component in turn and kriging it, thereby separating it from the others. Oliver et al. (2000) used this approach on SPOT imagery. Both of these

methods will allow the removal of short-range uncorrelated noise components from subsequent sampling, modeling, and prediction stages. The spatial decomposition of the environmental variables begs the question of whether the target attribute should also be spatially decomposed. In most cases, there would probably be insufficient observations to do this effectively, but it could be done where this is not the case.

An alternative approach to spatial decomposition is that of spatial lagging, that is, to fit a model such as,

$$S(x, y) = f(s(x + u, y + v), c(x + u, y + v),...) + g(x + u, y + v),$$

$$(37.22)$$

where the soil attribute of interest is "regressed" on the layers representing the *scorpan* attributes and on spatially lagged ($+u$, $+v$) copies of them, with u and v variable. This approach seems somewhat more cumbersome than the spatial decomposition approach.

iv. *Sampling of assembled data (Q) to obtain sampling sites*
In most cases, soil sampling will be required to set up the model (except perhaps when the aim is map updating). We have a lot of prior information on the environmental variables, which we can use to guide the sampling. The aim is to construct predictive equations for the soil attributes of interest in terms of the environmental variables. This is a kind of calibration exercise, and, therefore, it would seem wise to span the range of values of each variable, so that the prediction model will not be required to extrapolate beyond its bounds. Hengl et al. (2003) proposed sampling along the principal components of the *scorpan* factors; the number of samples taken from each of the components is the proportion of the total variance described by each of the principal components.

Minasny and McBratney (2006) argued that, for the purpose of spatial regression calibration, it would be beneficial to cover the range of values of each of the covariates using LHS (McKay et al., 1979). LHS is a procedure that ensures a full coverage of the range of each variable by maximally stratifying the marginal distribution. LHS involves sampling n values from the prescribed distribution of each of k variables $x_1, x_2,..., x_k$. The cumulative distribution for each variable is divided into n equiprobable intervals. A value is selected randomly from each interval. The n values obtained for each variable are matched randomly with those of the other variables. This method does not require more samples for more dimensions (variables). A conditioned Latin hypercube sampling (cLHS) was proposed by Minasny and McBratney (2006) for sampling of existing covariates or *scorpan* factors. We cannot directly apply conventional LHS to the multivariate distribution of covariates. Sample points selected by conventional LHS may represent combinations of the variables that do not exist in the real world. Randomization is used in this case as the distribution among the variables is not even, and some parts of the variable space might not correspond with existing soil. The conditioned LHS algorithm

attempts to select n observations (sites) from the covariates, which can form a Latin hypercube in the feature space. The algorithm solves an optimization problem: given M sites with covariates (X), select m sample sites ($m \ll M$) so that the sampled sites x form a Latin hypercube. The method is a search algorithm based on heuristic rules combined with an annealing schedule. The objective function to be minimized is an error criterion that counts the occupancy of the hypercube. This kind of sampling should produce a reasonably efficient way of sampling the soil and its environment, so that the range of conditions are encountered, ensuring a good chance of fitting relationships *if they exist*.

An alternative procedure is suggested by the work of Lagacherie et al. (2001) that define a reference area (Favrot, 1989), which through the spatial data layers of the environmental variables extrapolates well to a larger region. Sample the reference area purposively or systematically (fit the model and extrapolate to the rest of the area). This might give a better chance of fitting local relationships with a given sampling effort and should be more efficient in field time—the advantage hinges however on how well the extrapolation can be done.

v. *GPS-field sampling and laboratory analysis to obtain soil class or property data*
Step (iii) yields a set of m spatial coordinates at which the observations of the soil attribute(s) are to be made in the field. These can be located with a GPS receiver, and samples taken for subsequent laboratory analyses, in the usual manner. At a subset of these locations, say 5%, a duplicate observation should be made at a distance (say) half the resolution (ρ) to get an estimate of the short-range field variability in the soil attribute(s). This will help in subsequent spatial modeling. (If a specific purpose or numerical classification or allocation to a conventional classification system is required then the observed soil data can be processed to obtain discrete or continuous class labels for each observation site.)

vi. *Fit quantitative relationships (observing Ockham's razor)*
We can now assemble the soil data for the left side of the *scorpan* equation and the environmental data for the right side. We can now fit the model $f()$ representing the m locations using any of the techniques described in Section 37.3.3. Ockham's razor (the principle that states that "Entities should not be multiplied unnecessarily") should be applied to find the model (with the least number of parameters) that fits best. The residuals (e) of the soil property or class probability or membership at each of m sites should be estimated also (and kriged using either *scorpan*-simple or *scorpan*-compositional kriging as mentioned in Section 37.3.4.2.2).

A single model may not be adequate especially if there are strong pedogeomorphic or geological contrasts within the study area—each subregion may show quite different relationships between soil and environmental variables. In that case, it may be necessary to develop a *scorpan*-SSPFe model for each pedogeomorphic or geological subregion. Care needs to be taken that the prediction surfaces

produced are realistically continuous. This approach demands a higher data burden than a single model for a whole geographic region. Once again the total number of parameters in the final model needs to be considered.

If we have a lot of local information for both sides of the equation (like an existing soil map, or proximally sensed soil data), we can fit a local model,

$$S[x, y] = f_l([x, y]) + e_l(x, y), \qquad (37.23)$$

where the l subscript on $f_l([x, y])$ and $e_l(x, y)$ refers to models fitted to a local (moving) neighborhood centered on $[x, y]$ rather than to the whole area to be mapped. It will be rare that m will be large enough to fit this model. This kind of model was used geostatistically, that is, $S[x, y] = e_l(x, y)$ for soil salinity mapping by Walter et al. (2001).

vii. *Predict digital map*

We now have a model $f()$ fitted to the m locations, which we can now apply to the $(M - m)$ locations where we have no soil observations, but have environmental observations. Additionally, kriging of the residuals (e) at is also done at the $(M - m)$ locations and the results added together. Additionally, the uncertainties of the predictions of $f()$ and e at the M locations should be evaluated. Raster maps can then be made of the soil attribute(s) and their associated uncertainties.

viii. *Uncertainty analysis*

A principal feature of digital soil mapping is the ability to present an accompanying map of uncertainty to the prediction map. However, uncertainty is seldom quantified in digital soil maps. There are two important sources of uncertainty: the uncertainty in the input data and the uncertainty of the spatial prediction techniques.

Uncertainty in the input data can be from the soil data and *scorpan* factors. The uncertainty in soil data is usually much larger than in the covariates. Usually, digital soil mapping uses legacy soil data. Legacy soil data arise from traditional survey, which is generally empirical and based on the mental development of the surveyor. There are no statistical criteria for traditional soil sampling, which may lead to bias in the areas being sampled. Carré et al. (2007) discusses some techniques for evaluating the quality of the legacy soil data based on the *scorpan* factors.

Conventional polygon-based maps often have no uncertainty information. In the Australian Soil Resource Information system, the uncertainty of a soil property is derived from the polygon-based map (McKenzie et al., 2005). This is calculated from a standard uncertainty, which is based on scientific judgment using all the relevant information available, which may include previous measurements on related soils, experience of lab or field or prediction techniques, and assumed from spatial variation.

The major contributing factors to uncertainty in spatial prediction are the neighborhood points used for interpolation: the number and proximity of the samples, clustering of samples, and continuity of the variables. In geostatistics, the usual approach for calculating the uncertainty is by computing a kriging estimate and the associated error variance, which are then combined to derive a Gaussian-type confidence interval. For the regression-kriging approach, the uncertainty has two parts, from the regression model and the kriging of the residuals. See Minasny and McBratney (2007a, 2007b) for mathematical details on this topic.

The uncertainty for prediction of soil classes is more difficult to quantify. Kempen et al. (2009) assessed the uncertainty of the multinomial logistic regression model prediction of soil classes by using Shannon's information entropy. It is calculated based on the probability of the predicted soil classes, the maximum entropy is 1, which occurs when all outcomes have equal probability. The minimum value for the entropy is 0, which occurs when there is no uncertainty and one of the outcomes has probability 1. It should be noted that the entropy indicates whether the predicted soil group has a large probability, it does not indicate that the prediction itself is correct. What is really required is a prediction interval on the prediction probability statement.

Bui and Moran (2003) used a heuristic approach for uncertainty of soil classes. The training area was randomly sampled 10 times. Each time a decision tree was built, and soil classes were predicted into unmapped areas. Where the same class was consistently predicted by several trees, the likelihood is greater that the prediction is consistent (not necessarily correct). The modal class prediction from the 10 trees was used in making the final map. The proportion of predictions resulting in the modal class was used to create a map of uncertainty of the predictions. The map of uncertainty of predictions indicates where the trees have difficulty in capturing the spatial pattern of soils in the landscape.

ix. *Field sampling and laboratory analysis for corroboration and quality testing*

It must not be assumed that the digital information is perfect with minimal information on the quality of the information. Indeed, we cannot expect this kind of map to be more accurate than conventional ones. There are two reasons for this: (i) local variation (at whatever resolution) is a limiting factor, there is a lot of soil variation within 10 m, or 100 m, or 1 km; and (ii) there is uncertainty in environmental layers and this can propagate errors (Lagacherie and Holmes, 1997; Florinsky, 1998; Heuvelink and Burrough, 2002).

Like conventional soil maps, digital ones should also be evaluated for quality. This is the same as ground truthing in remote sensing where many measures have been advocated (see Story and Congalton, 1986; Liu et al., 2006). The main quality criteria for soil class maps are purity (user's accuracy) and producer's accuracy, and root mean square error for continuous soil property maps. These can be evaluated by a number of strategies.

One approach is by cross-validation. Some proportion of the available observations are put to one side, a common

proportion is one-third. Prediction modeling is done with the remaining two-thirds of the data and the map produced. The one-third put aside are now considered test sites and compared with model predictions This can give an indication of producer's and user's accuracy for categorical maps and average bias and root mean square error for maps of continuous soil properties. The results are only indicative because the sample was probably not designed for the express purpose of quality assessment. The results might be somewhat biased for legacy data (Carré et al., 2007), and less so for observations from LHS (Minasny and McBratney, 2006), but still not optimal.

It is better to construct a new sample. Cluster sampling and stratified random sampling have been used in the past for testing the accuracy of conventional maps (De Gruijter and Marsman, 1985; Marsman and De Gruijter, 1986). Kempen et al. (2009) carried out an accuracy assessment of their digitally updated soil map of part of the Netherlands. They observed 150 new points based on a stratified random sampling design with strata based on the original map and found a small improvement in purity from 52% to 58%. They also estimated the purity and producer's accuracy (which they called "sensitivity") for all the units of the new digital soil map.

One of the principal advantages of digital soil maps is the promise of spatialized uncertainty estimators. The quality of the combination of the estimate and its associated uncertainty needs to be tested to properly evaluate digital soil maps. Thus far, this has not been attempted.

x. *If necessary simplify legend, or decrease resolution by returning to (i) or, improve map by returning to (v)*
   If we find the map does not meet design specifications, that is, class purity is less than x%, or the confidence intervals for a soil property are too wide over parts of the map, then we can simplify the legend or decrease spatial resolution (i)—but these should be design specifications, or more sensibly, target further sampling (v) in areas where predictability appears to be poor, and recalculate the maps.

37.5.2 Uses

There are various potential uses of the digital soil mapping approach. Among them are production of digital soil maps as a replacement for the paper-based choropleth soil map of the past, updating existing soil maps, extrapolating existing soil maps into unmapped areas, construction of dynamic soil maps, digital soil assessment, and global soil map.

37.5.2.1 Digital Soil Maps

The main use of the *scorpan*-SSPF approach is to replace the polygon-based soil maps of the past with digital maps of soil properties and classes and their associated uncertainties for areas previously mapped, or for new areas. These maps will be stored and manipulated in digital form in a GIS creating the possibility of vast arrays of data for analysis and interpretation.

The first digital soil maps were simply representations of the observations without interpolation or relation to the environment (e.g., Webster et al., 1979). Some authors have worked on better ways to present digital spatial soil information, chronologically, De Gruijter and Bie (1975), De Gruijter et al. (1997), and Grunwald et al. (2001). This is an area requiring considerable research. One goal must be fully operational multiresolution digital soil maps.

While we cannot necessarily expect the maps to more accurate than conventional ones, we can expect to have a quantitative estimate of the uncertainty (Section 37.5.1); sampling effort should be expended to achieve this. Laba et al. (2002) compared conventional (Congalton and Green, 1999) and fuzzy (Gopal and Woodcock, 1995) methods to assess the accuracy and uncertainty of land cover maps produced at high taxonomic resolution. These methods could be applied to digital soil maps.

Survey commissioners, decision makers and users in general would perhaps be more comfortable with a concept of certainty rather uncertainty. This answers the question, "how well do we know the value at some location?" rather than concentrating on "how badly we know it." A potentially adequate standardized $(0 \rightarrow 1$ or $0\% \rightarrow 100\%)$ measure of certainty is $f = 1 - \min(2s/V, 1)$, where s is the standard deviation of the estimate. For example, if we have an estimate V of clay content of 50% and an s of 5% then $f = 0.8$ or $f\% = 80\%$. More sophisticated measures may be required, such as a certainty characteristic—the probability that a statement C is true within a distance d or an increasing neighborhood A. Clearly, more work is needed on standards for digital soil maps.

A recent advancement in digital soil mapping is the capability to map the distribution of soil properties over the whole soil profile. This created a pseudo 3D distribution of soil properties. Minasny et al. (2006) fitted an exponential decay depth function for soil organic C distribution data in the Edgeroi area in NSW, Australia. Parameters of the exponential function were then mapped using neural networks with some *scorpan* factors as covariates. Carbon storage for any point in the landscape then can be obtained by integrating the exponential function to the desired depth. A more general depth function can be obtained using an equal-area spline (Bishop et al., 1999). Parameters of the spline function can be mapped creating a whole soil profile distribution (Malone et al., 2009).

37.5.2.2 Updating Soil Maps

Digital soil mapping can be used to update soil maps from previous surveys. Models predicting soil classes or series can be built based on legacy soil data (existing soil maps or soil profile descriptions) with *scorpan* factors as covariates. The spatial prediction models can be applied to the entire data. Kempen et al. (2009) used digital soil mapping principles to update the national 1:50,000 soil map for the province of Drenthe (2680 km²) using legacy soil data. Multinomial logistic regression was used to quantify the relationship between ancillary variables and soil groups. The model-building process was guided by pedological expert knowledge to ensure that the final regression model was not only statistically sound but also pedologically plausible.

They estimated actual purity of the updated map, as assessed by the validation sample, was 58%.

37.5.2.3 Interpolation or Extrapolation of Existing Soil Maps

If *s* for a previously mapped region is put on the right of *scorpan* equation, the legend is retained largely, and new samples are collected, this might be considered by some to be map updating. There is another possibility; this is where the previous *s* is put on the left rather than the right side of the *scorpan* equation (an example is given in Bui et al., 1999). The advantage of this is that no new sampling is required for fitting—although corroboration sampling should be done. This would allow quantitative elaboration of the existing (but unknown) models, for classes this is what Girard (1983) referred to as (the usually unknown) "chorological rules," and their subsequent extrapolation to new areas. Possible problems include repeating old models that may be wrong, or extrapolation outside the range of associated environmental data sets—Lagacherie and Holmes (1997) and Lagacherie and Voltz (2000) discuss this for landscapes, while McBratney et al. (2002) were concerned about this for pedotransfer functions. It is also important to establish for which purpose existing data have been collected. Some soil surveyors use auger observations to confirm their mental models, while others use them to find inclusions or boundaries. Therefore, this approach should be used with a deal of caution.

Extrapolation of soil types from one small area to a larger extent can be done once the digital soil mapper knows the initial area is representative of the larger extent (that is why it is usually called "reference area"). This idea, initially proposed by Favrot (1981), has been applied by Lagacherie et al. (2001) for extrapolating French Mediterranean soilscapes. Soilscapes of reference defined as a combination of soil-forming factors in a buffer neighbor can be expressed as a vector composition of elementary landscape classes of different sizes. Since the landscape classes are known to be representative of the overall area, Manhattan distances between soilscapes of reference and composition of landscape classes in running neighboring windows can be calculated. The final allocated soilscape is the one that minimizes the Manhattan distance. Two outputs are then produced in that case:

- The final soilscape map
- The corresponding allocation distance giving access to the accuracy of the final map

This technique allows for minimizing field sampling. However, the main scientific question is where to stop the boundaries of the bigger extent to map?

Bui and Moran (2001, 2003) built decision tree rules in training areas where detailed soil maps were available, and the rules were extrapolated to larger areas where detailed maps were not available. Grinand et al. (2008) tested this approach and observed marked differences in accuracy between the training area and the extrapolated area. They found that sampling intensity did not appear to influence the accuracy of prediction, and spatial context integration by the use of a mean filtering algorithm on the covariates increased the accuracy of the prediction on the extrapolated area.

37.5.2.4 Environmental Change: Partially Dynamic "Scenario" Soil Maps

One major criticism of conventional soil maps is that they are essentially static statements. Digital soil maps created with the *scorpan*-SSPF methodology offer new and necessary possibilities. It is becoming increasingly important for environmental reasons to know not just $S[x, y]$ but $S[x, y, t]$. If we know any of the partial differentials, $\delta s/\delta t$, $\delta c/\delta t$, $\delta o/\delta t$, etc., the first two perhaps being the more important—we can project the existing soil map forward by some time *u* by calculating most simply say $c + u(\delta c/\delta t)$ for all points and running the new *c* layer(s) through the prediction function. "Change-detection analysis" (Mücher et al., 2000) is well developed for land use and vegetation change (components of $\delta o/\delta t$) using remotely sensed imagery and/or aerial photographs average, and localized values of *o* can be estimated from rasterized images taken at two or more times (Munyati, 2000). Other derivatives may be obtained from models (e.g., temperature and rainfall changes) or from a few monitoring stations (e.g., soil changes) within the area of interest (Mendonça-Santos et al., 1997).

This potential approach has limitations compared with a fully fledged dynamic simulation model, such as lack of feedback and possible extrapolation problems, where, for example, $c + u(\delta c/\delta t)$ takes us (well) outside the range of the original training data. Nevertheless, we still have a relatively quick and easy way to produce first-cut "scenario" soil maps of both properties and classes.

The European Parliament is currently implementing a new Directive on the use of renewable energy in Europe. Ten percent share of renewable energy in transport should be achieved by each Member State by 2020. The main source of renewable energy for transport is planned to be biofuel. One scientific aim addressing the sustainability question is then to assess the impacts of such a Directive on global land (soil and above-ground vegetation) carbon stock changes due to indirect land use conversions. After quantifying the biocrops needed to achieve the 10% biofuel share target, land use conversion due to biocrops and transfer of food/feed/fiber crops will be spatially estimated at the global scale. Overlaying the current stock of land carbon with the current land use and targeted land use conversion, climate and land management practices information, it is possible to spatially assess the carbon stock changes within the next 20 years by following the Tier 1 approach of volume 4 of the *IPCC Guidelines for National Greenhouse Gas Inventories* (IPCC, 2006). The main difficulty is not only to estimate uncertainties coming from the modeling (this is provided by the IPCC report) but also to quantify the uncertainties of the different sources of inputs, usually not specified.

37.5.2.5 Digital Soil Assessment

Digital soil assessment is the step following digital soil mapping (Carré et al., 2007). It is the quantitative modeling of difficult to measure soil attributes, necessary for assessing threats to soil (e.g., erosion, decline of organic matter, and compaction) and soil functions (biomass production and environmental interactions) using outputs from digital soil maps, that is, the soil

properties map and the accuracy of the predictions. The threats to soil and the soil functions are generally aspatial models. They require not only data on soil (as provided through digital soil mapping) but also external data that can be also considered as soil-forming factors, and already used as covariates. These can be climate, land cover, or land management practices. DSA requires running the aspatial models, and as for DSM, the prediction of uncertainties coming from input data and the models.

Uncertainty prediction is usually performed using global sensitivity analysis of the models and simulating the distribution of the uncertainties with Monte Carlo simulation (Heuvelink, 1998). Three exercises have been tested for digital assessment problems. The first one deals with soil pollution issue (Romić et al., 2007) using regression-kriging approach and continuous limitation scores for assessing global soil pollution. The second one deals with wind erosion (Reuter et al., not published) combining regression-kriging for the soil properties prediction (used as output for wind erosion). The third one deals with estimation of biomass production in England & Wales (Mayr et al., private communication).

37.5.2.6 Global Soil Map

A Digital World Soil Properties Map consortium (www.global-soilmap.net) has been formed and comprises representatives from universities, research centers, development organizations, and private enterprises around the world. This is a response to the urgent need for accurate, up-to-date, and spatially referenced soil information, at a global scale, as expressed by the modeling community, farmers and land managers, and policy and decision makers. The objective of this consortium is to create a digital map of the world's soil properties at a resolution of 90 m × 90 m to a depth of 1 m based on legacy and newly collected soil data.

Making a digital global soil map is a challenging task, methods for mapping soil properties at a continental or global scale are not straightforward. Moreover, different parts of the world have varying soil information of varying qualities. The soil information can be from legacy soil profile data, existing soil maps, but also data from reflectance spectra. The soil–landscape model will vary from place to place and the knowledge and techniques for regional soil mapping may not be applicable at a global scale.

Minasny and McBratney (2010) summarized a methodology for global digital soil mapping based on legacy soil data. For an area of interest, we assemble all the *scorpan* or environmental covariates and existing soil data. The second step is to check how the soil data cover the covariate space and to select possible training areas.

Methods used for digital soil mapping depend on the availability of soil data, and the possibilities in the order from the richest to the poorest soil information are as follows:

1. *Detailed soil maps with legends and soil point data*
   This is the richest information that can give the best prediction of soil properties. Soil properties can be derived from both soil maps and soil point data. The available methods are extracting soil properties from soil map using a spatially weighted measure of central tendency, for example, the mean, spatial disaggregation of soil maps, *scorpan*

kriging, and combinations of these. An example of such an application is given in Henderson et al. (2001, 2003).

2. *Soil point data*
   When soil point data are available, soil properties can be interpolated and extrapolated to the whole area by using a combination of empirical deterministic modeling and a stochastic spatial component. We called this the *scorpan* kriging approach.

3. *Detailed soil maps with legends*
   When only soil maps are available, soil properties need to be extracted from the maps using some central and distributional concepts of soil mapping units.

4. *No data*
   When no data or soil maps exist in area, we will use an approach we called *Homosoil*, which means that we need to estimate the likely soil properties under the observed soil-forming factors or *scorpan* factors.

37.5.3 General Discussion

We now discuss some general points relating to the *scorpan*-SSPF approach to making digital soil maps.

37.5.3.1 Is This the Right Approach?

As stated at the beginning of Section 37.5.1, this is a proposal, not a fait accompli, albeit based on a significant amount of work and experience worldwide. Whether or not this turns out to be the right approach hinges on a number of factors, not all of them scientific, but we shall deal with those first. Scientifically one could ask a number of questions. Are there other soil-forming factors? Are we missing key variables? For example, have we successfully incorporated hydrological effects? Is the underlying idea of a soil somehow in equilibrium with its environment reasonable enough to be predictive? Or, is the soil too chaotic for prediction by other factors? Will the proposed methodology give similar answers as traditional approaches, and do we want it to? Further experiment and experience will undoubtedly answer these questions.

There are other socio-economico-political factors that will have a bearing also. The sociopolitical factors demand recognition of, and solutions to, environmental problems. We believe that the approach has the right kind of economics; it is potentially cheaper than traditional approaches and gives the desired kinds of information. Most of the hardware and software tools are in place to put this approach into practice. Clearly, integrated systems have to be devised. Research into aspects is always needed, principally efficient sampling designs and useful certainty estimation. The biggest stumbling block is identifying and assembling the teams of personnel with the skills required to complete the task. Education of skilled and knowledgeable personnel for those teams is a key priority.

37.5.3.2 Dangers: Let the User Beware!

There are real dangers with this, or any new approach, if it is misused or abused. Here, we outline briefly some obvious pitfalls.

1. *Data quantity and quality.* The first danger is not using enough real soil observations to fit the models, or with using poor quality (missing or noisy) predictor variables. This can, to a degree, be handled by uncertainty analysis—a large topic (Heuvelink, 1998), which has not been discussed formally in this paper. There is a lower limit below which any fitted models will be meaningless.

2. *Overfitting the data.* It is easy to overfit models; this could be because of lack of observations but is more often due to a lack of parsimony, especially a problem for tree-based methods. Overfitted models predict poorly. It is imperative to apply Ockham's razor—this will help with evaluating poorly fitting or overfitted models. The use of cross-validation, pruning, and boosting methods (Hastie et al., 2009) might also help.

3. *Circularity.* A third hazard comes from the possible circularity of the model, for example, a DEM producing climate surfaces producing soil variables as an input to soil class prediction. Once again uncertainty analysis will help.

4. *Databases and data mining.* During the past decade, soil scientists have created regional, national, continental, and worldwide databases. Data mining is a phrase for a class of database applications that look for hidden patterns in such groups of data (Hastie et al., 2009). (Unfortunately, the term is sometimes misused to describe software that presents data in new ways.) Proper data mining software attempts to discover previously unknown relationships among the data. Data mining is a broad concept from supervised learning (prediction) to unsupervised learning and includes all the methods described in Section 37.3.3—neural networks, classification trees with boosting. There are a large number of commercial software products available to do this. They incorporate one, or often several, of the methods described in Section 37.3.3. This will make evaluation difficult as different soil science groups use different software products for fitting $f()$; therefore, comparative studies will be important to evaluate the best approaches. In addition, large national or international databases of legacy soil data will be available (e.g., Bui and Moran, 2003); they also have potential problems because of their unknown site selection probabilities—which are not usually equal—some of the data from Britain where a 5 km grid survey has been completed, are an exception!

37.5.3.3 A New Paradigm?

Hudson (1992) described soil survey as a paradigm-based science. Paradigm is a much overused and hackneyed word these days but it has a precise philosophical meaning. (Much of the following two paragraphs is paraphrased from Rosenberg, 2000.) Paradigm is a term employed by Kuhn (1996) to characterize a scientific tradition, including its theory, apparatus, methodology, and scientific philosophy. The soil scientist's task is to apply the paradigm to the solution of problems. Failure to solve problems is the fault of the scientist not the paradigm. Persistent failure makes a problem an anomaly and threatens a revolution, which may end the paradigm's hegemony.

What is the difference between the *scorpan*-SSPFe approach and the conventional Jenny-landscape model? Both are models—they are simplified descriptions of regularities governing a natural process, usually mathematical and sometimes derived from a more general or simplified theory. Ontologically, they are similar—they both require soil objects and attributes, which are a function of their environment. The conventional paradigm is a qualitative theory. The approach outlined here is a quantitative, partially deterministic, partially probabilistic, empirical theory. So methodologically, they are quite different. The apparatus is also different; here, we require digital information, computers, GIS, etc. The Jenny-landscape model may fall under the *deductive-nomological* model of scientific explanation but, because of its somewhat probabilistic nature, the *scorpan*-SSPFe approach may fall under the *inductive-statistical* model of explanation (Rosenberg, 2000). So the *scorpan*-SSPFe approach to soil mapping probably represents an emerging paradigm eventually leading to a complete paradigm shift.

This begs the question, does $f()$ have to be empirical? The Vienna school of logical empiricists would be generally happy with *scorpan*-SSPFe approach, although perhaps they would have difficulties with its partially probabilistic nature. The lack of a mechanistic theory for predicting soil tugs at the soil scientist's cloak of explanation. Perhaps this is an unnecessary concern; philosophical empiricists believe that there is nothing to causation beyond a regular sequence. Any testing of the mechanistic theory will require empirical observation of the real world. The first attempts at a mechanistic approach have begun but it will be a long time before the mechanistic theoretical approach will be competitive in the predictive sense.

37.5.4 Working Group on Digital Soil Mapping

The Working Group on digital soil mapping (www.digitalsoilmapping.org) was formed in 2005. It operates under the auspices of the Commissions on Soil Geography (C1.2) and Pedometrics (C1.5) of the International Union of Soil Sciences (IUSS). The workgroup has biennial workshops on Digital Soil Mapping. The first workshop was held in Montpellier, France in 2004. The outcome of the workshop is published as a book: Digital Soil Mapping: An Introductory Perspective (Lagacherie et al., 2006). The second workshop was in Rio de Janeiro, Brazil in 2006. The outcome of the workshop is published as a book: Digital Soil Mapping with Limited Data (Hartemink et al., 2008). The third workshop was held in Logan, the United States, in 2008 with an outcome: Digital Soil Mapping: Bridging Research, Environmental, Application, and Operation (Boettinger et al., 2010). The fourth workshop was in Rome in 2010.

37.6 Conclusions

We have reviewed various approaches to predictive modeling and data acquisition and proposed a framework and a methodology for producing digital soil maps.

37.6.1 Summary of the Method

The *scorpan*-SSPFe method essentially involves the following steps:

i. Define soil attribute(s) of interest and decide resolution ρ and block size β
ii. Assemble data layers to represent *Q*
iii. Spatial decomposition or lagging of data layers
iv. Sampling of assembled data (*Q*) to obtain sampling sites
v. GPS-field sampling and laboratory analysis to obtain soil class or property data
vi. Fit quantitative relationships (observing Ockham's razor) including spatially autocorrelated residual errors
vii. Predict digital map
viii. Perform uncertainty analysis
ix. Field sampling and laboratory analysis for corroboration and quality testing
x. If necessary simplify legend, or decrease resolution by returning to (i) or, improve map by returning to (v)

All of the hardware and software tools, technologies and knowledge, are in place to make this approach operational. This is clearly an exciting time for soil resource assessment.

37.6.2 Future Work: Open Questions

Clearly, we need to try out the methodology outlined above and by experience we shall discover the useful forms of *f*() and the serviceable *Q* layers. These are the key open questions. In summary, topics to be further researched include the following:

1. Environmental covariates for digital soil mapping
2. Spatial decomposition and/or lagging of soil and environmental data layers
3. Sampling methods for creating digital soil maps
4. Quantitative modeling for predicting soil classes and attributes (including generalized linear and additive models, classification and regression trees, neural networks, fuzzy systems, expert knowledge and geostatistics)
5. Quality assessment of digital soil maps
6. (Re)presentation of digital soil maps
7. Economics of digital soil mapping

Nevertheless, we believe the methodology can be used now for real-world applications.

References

Aandahl, A.R. 1948. The characterization of slope positions and their influence on the total N content of a few virgin soils in Western Iowa. Soil Sci. Soc. Am. Proc. 13:449–454.

Agbu, P.A., D.J. Fehrenbacher, and I.J. Jansen. 1990. Soil property relationships with SPOT satellite digital data in east central Illinois. Soil Sci. Soc. Am. J. 54:807–812.

Ahmed, S., and G. DeMarsily. 1987. Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. Water Resour. Res. 23:1717–1737.

Aiken, R.M., M.D. Jawson, K. Grahammer, and A.D. Polymenopoulos. 1991. Positional, spatially correlated and random components of variability in carbon-dioxide efflux. J. Environ. Qual. 20:301–308.

Altese, E., O. Bolognani, M. Mancini, and P.A. Troeh. 1996. Retrieving soil moisture over bare soil from ERS-1 synthetic aperture radar data: Sensitivity analysis based on theoretical surface scattering model and field data. Water Resour. Res. 32:653–661.

Anderson, K.E., and P.A. Furley. 1975. An assessment of the relationship between surface properties of chalk soils and slope form using principal component analysis. J. Soil Sci. 26:130–143.

Anderson-Cook, C.M., M.M. Alley, J.K.F. Roygard, R. Khosla, R.B. Noble, and J.A. Doolittle. 2002. Differentiating soil types using electromagnetic conductivity and crop yield maps. Soil Sci. Soc. Am. J. 66:1652–1670.

Arrell, K.E., P.F. Fisher, N.J. Tate, and L. Bastin. 2007. A fuzzy c-means classification of elevation derivatives to extract the morphometric classification of landforms in Snowdonia. Wales. Comput. Geosci. 33:1366–1381.

Ballabio, C. 2009. Spatial prediction of soil properties in temperate mountain regions using support vector regression. Geoderma 151:338–350.

Band, L.E. 1989a. A terrain-based watershed information system. Hydrol. Processes 3:151–162.

Band, L.E. 1989b. Automating topographic and ecounit extraction from mountainous forested watersheds. Artif. Intell. Appl. Nat. Resour. Manag. 3:1–11.

Band, L.E. 1989c. Spatial aggregation of complex terrain. Geogr. Anal. 21:279–293.

Barshad, I. 1957. Factors affecting clay formation, Clays and Clay Minerals 6:110–132.

Bayramin, I. 2000. Using geographic information system and remote sensing techniques in making pre-soil surveys, p. 76–33. *In* Proc. Int. Sympos. Desertification. Soil Science Society of Turkey, June 13–17, Konya, Turkey.

Beard, L.P. 2000. Comparison of methods for estimating earth resistivity from airborne electromagnetic measurements. J. Appl. Geophys. 45:239–259.

Behrens, T., H. Förster, T. Scholten, U. Steinrücken, E.-D. Spies, and M. Goldschmitt. 2005. Digital soil mapping using artificial neural networks. J. Plant Nutr. Soil Sci. 168:1–13.

Behrens, T., K. Schmidt, and T. Scholten. 2007a. Multi-scale digital terrain analysis and feature selection in digital soil mapping. Biannual Conf. Commission 1.5 Pedometrics, Division 1 of the International Union of Soil Sciences (IUSS), August 27–30, Tübingen, Germany.

Behrens, T., and T. Scholten. 2006. Digital soil mapping in Germany—A review. J. Plant Nutr. Soil Sci. 169:434–443.

Behrens, T., A.-X. Zhu, K. Schmidt, and T. Scholten. 2007b. Multiscale digital terrain analysis and feature selection for digital soil mapping. Geoderma 155:175–185.

Bell, J.C., R.L. Cunningham, and M.W. Havens. 1992. Calibration and validation of a soil–landscape model for predicting soil drainage class. Soil Sci. Soc. Am. J. 56:1860–1866.

Bell, J.C., R.L. Cunningham, and M.W. Havens. 1994. Soil drainage probability mapping using a soil–landscape model. Soil Sci. Soc. Am. J. 58:464–470.

Beven, K., and M.J. Kirkby. 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24:1–10.

Bierwirth, P.N. 1996. Gamma-radiometrics, a remote sensing tool for understanding soils. ACLEP Newslett. 5:12–14.

Bierwirth, P.N., and R.S. Brodie. 2005. Identifying acid sulfate soil hotspots from airborne gamma-radiometric data and GIS analysis. Australia Government Bureau of Rural Sciences.

Bishop, T.F.A., and A.B. McBratney. 2001. A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma 103:149–160.

Bishop, T.F.A., A.B. McBratney, and G.M. Laslett. 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91:27–45.

Bishop, T.F.A., A.B. McBratney, and B.M. Whelan. 2001. Measuring the quality of digital soil maps using information criteria. Geoderma 105:93–111.

Boegh, E., H. Soegaard, and A. Thomsen. 2002. Evaluating evapotranspiration rates and surface conditions using Landsat TM to estimate atmospheric resistance and surface resistance. Remote Sens. Environ. 79:329–343.

Boettiger, J.L., D.W. Howell, A.C. Moore, A.E. Hartemink, and S. Kienast-Brown (eds.). 2010. Digital soil mapping: Bridging research, production, and environmental application. Springer, New York.

Boettiger, J.L., R.D. Ramsey, J.M. Bodily, N.J. Cole, S. Kienast-Brown, S.J. Nield, A.M. Saunders, and A.K. Stum. 2008. Landsat spectral data for digital soil mapping, p. 193–202. *In* A.E. Hartemink., A.B. McBratney, and M.L. Mendonça-Santos (eds.). Digital soil mapping with limited data. Springer, Heidelberg, Germany.

Bogaert, P., and D. D'Or. 2002. Estimating soil properties from thematic soil maps: The Bayesian maximum entropy approach. Soil Sci. Soc. Am. J. 66:1492–1500.

Boulaine, J. 1980. Pédologie appliquée. Collection Sciences agronomiques. Masson, Paris, France.

Bourennane, H., D. King, P. Chery, and A. Bruand. 1996. Improving the kriging of a soil variable using slope gradient as external drift. Eur. J. Soil Sci. 47:473–483.

Bourgault, G. 1994. Robustness of noise filtering by kriging analysis. Math. Geol. 26:733–752.

Brabyn, L. 1997. Classification of macro landforms using GIS. ITC J. 1997:26–40.

Brabyn, L. 1998. GIS analysis of macro landform, p. 35–48. *In* P. Firns (ed.) The 10th colloquium of the spatial information research centre. University of Otago, Dunedin, New Zealand.

Brest, C.L., and S.N. Goward. 1987. Deriving surface albedo measurements from narrow-band satellite data. Int. J. Remote Sens. 8:351–367.

Breiman, L. 1996. Bagging predictors. Mach. Learn. 26:123–140.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. Classification and regression trees. Wadsworth, Belmont, CA.

Brus, D.J., P. Bogaert, and G.B.M. Heuvelink. 2008. Bayesian maximum entropy prediction of soil categories using a traditional soil map as soft information. Eur. J. Soil Sci. 59:166–177.

Bui, E.N. 2004. Soil survey as a representation of structured knowledge. Geoderma 120:17–26.

Bui, E.N., B.L. Henderson, and K. Viergever. 2006. Knowledge discovery from models of soil properties developed through data mining. Ecol. Model. 191:431–446.

Bui, E.N., A. Loughhead, and R. Corner. 1999. Extracting soil–landscape rules from previous soil surveys. Aust. J. Soil Res. 37:495–508.

Bui, E.N., and C.J. Moran. 2000. Regional-scale investigation of the spatial distribution and origin of soluble salts in central north Queensland. Hydrol. Processes 14:237–250.

Bui, E.N., and C.J. Moran. 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma 103:79–94.

Bui, E.N., and C.J. Moran. 2003. A strategy to fill gaps in soil survey over large spatial extents: An example from the Murray–Darling basin of Australia. Geoderma 111:21–44.

Bunnik, N.J.J. 1978. The multispectral reflectance of shortwave radiation by agricultural crops in relation with their morphological and optical properties. Ph.D. Thesis. Wageningen Agricultural University. Wageningen, the Netherlands.

Burgess, T.M., and R. Webster. 1980a. Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging. J. Soil Sci. 31:315–331.

Burgess, T.M., and R. Webster. 1980b. Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. J. Soil Sci. 31:333–341.

Burnett, C., and T. Blaschke. 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. Ecol. Model. 168:233–249.

Burrough, P.A. 1993. Soil variability: A late 20th century view. Soils Fertil. 56:529–562.

Burrough, P.A., P.F.M. van Gaans, and R.A. McMillan. 2000. High-resolution landform classification using fuzzy k-means. Fuzzy Sets Syst. 113:37–52.

Campbell, J.B. 1987. Introduction to remote sensing. The Guilford Press, New York.

Campling, P., A. Gobin, and J. Feyen. 2002. Logistic modeling to spatially predict the probability of soil drainage classes. Soil Sci. Soc. Am. J. 66:1390–1401.

Carré, F., and M.C. Girard. 2002. Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes. Geoderma 110:241–263.

Carré, F., and M. Jacobson. 2009. Numerical classification of soil profile data using distance metrics. Geoderma 148:336–345.

Carré, F., A.B. McBratney, and B. Minasny. 2007. Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. Geoderma 141:1–14.

Carrol, T.R. 1981. Airborne soil moisture measurement using natural terrestrial gamma radiation. Soil Sci. 132:358–366.

Carvalho, L.M.T., L.M.G. Fonseca, F. Murtagh, and J.G.P.W. Clevers. 2001. Digital change detection with the aid of multiresolution wavelet analysis. Int. J. Remote Sens. 22:3871–3876.

Cayley, A. 1859. On contour and slope lines. The London, Edinburgh and Dublin Philosophical Magazine. J. Sci. Ser. 18:264–268.

Chang, C.-W., D.W. Laird, M.J. Mausbach, and C.R. Hurburgh. 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Sci. Soc. Am. J. 65:480–490.

Chang, D.H., and S. Islam. 2000. Estimation of soil physical properties using remote sensing and artificial neural network. Remote Sens. Environ. 74:534–544.

Christakos, G. 1990. A Bayesian/maximum-entropy view to the spatial estimation problem. Math. Geol. 22:763–777.

Christakos, G. 2000. Modern spatiotemporal geostatistics. Oxford University Press, New York.

Cipra, J.E., D.P. Franzmeir, M.E. Bauer, and R.K. Boyd. 1980. Comparison of multispectral measurements from some nonvegetated soils using Landsat digital data and a spectroradiometer. Soil Sci. Soc. Am. J. 44:80–84.

Clarke, J.D.A., and C.F. Pain. 2004. From Utah to Mars: Regolith–landform mapping and its application. Am. Astronautical Soc. Sci. Technol. Ser. 107:131–159.

Clevers, J.G.P.W., and H.J.C. van Leeuwen. 1996. Combined use of optical and microwave remote sensing data for crop growth monitoring. Remote Sens. Environ. 56:42–51.

Conacher, A.J., and J.B. Dalrymple. 1977. The nine unit land surface model: An approach to pedogeomorphic research. Geoderma 18:1–154.

Congalton, R.G., and K. Green. 1999. Assessing the accuracy of remotely sensed data: Principles and practices. Lewis, New York.

Connolly, J., N.M. Holden, and S.M. Ward. 2007. Mapping peatlands in Ireland using a rule-based methodology and digital data. Soil Sci. Soc. Am. J. 71:492–499.

Cook, S.E., R. Corner, G.J. Grealish, P.E. Gessler, and C.J. Chartres. 1996a. A rule-based system to map soil properties. Soil Sci. Soc. Am. J. 60:1893–1900.

Cook, S.E., R. Corner, P.R. Groves, and G.J. Grealish. 1996b. Use of airborne gamma radiometric data for soil mapping. Aust. J. Soil Res. 34:183–194.

Corner, R.J., S.E. Cook, and G.A. Moore. 1997. Expector: A knowledge based soil attribute mapping method. ACLEP Newslett. 6:9–11.

Cressie, N.A.C. 1993. Statistics for spatial data. John Wiley & Sons, New York.

Cressie, N.A.C., and G. Johannesson. 2008. Fixed rank kriging for very large spatial data sets. J. R. Stat. Soc. Ser. B. 70:209–226.

Crowley, J.K. 1993. Mapping playa evaporite minerals with AVIRIS data: A 1st report from Death Valley, California. Remote Sens. Environ. 44:337–356.

Dale, M.B., A.B. McBratney, and J.S. Russell. 1989. On the role of expert systems and numerical taxonomy in soil classification. J. Soil Sci. 40:223–234.

Davies, B.E., and S.A. Gamm. 1969. Trend surface analysis applied to soil reaction values from Kent, England. Geoderma 3:223–231.

Dalrymple, J., R. Blong, and A. Conacher. 1968. A hypothetical nine unit landsurface model. Z. Geomorphol. 12:60–76.

De Bruin, S., and A. Stein. 1998. Soil–landscape modeling using fuzzy c-means clustering of attribute data derived from a digital elevation model (DEM). Geoderma 83:17–33.

De Gruijter, J.J., and S.W. Bie. 1975. A discrete approach to automated mapping of multivariate systems, p. 17–28. *In* J.M. Wilford-Brickwood, R. Bertrand, and L. van Zuylen (eds.) Automation in cartography. Proc. Tech. Working Session, Commission III. Int. Cartography Association. Enschede, the Netherlands.

De Gruijter, J.J., and B.A. Marsman. 1985. Transect sampling for reliable information on mapping units, p. 150–163. *In* D.R. Nielsen and J. Bouma (eds.) Soil spatial variability. Pudoc, Wageningen, the Netherlands.

De Gruijter, J.J., D.J.J. Walvoort, and P.F.M. van Gaans. 1997. Continuous soil maps—A fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma 77:169–195.

Delhomme, J.P. 1978. Kriging in the hydrosciences. Adv. Water Resour. 1:251–266.

Dehn, M., H. Gartner, and R. Dikau. 2001. Principals of semantic modeling of landform structures. Comput. Geosci. 27:1011–1013.

Deng, Y. 2007. New trends in digital terrain analysis: Landform definition, representation, and classification. Prog. Phys. Geogr. 31:405–419.

Diak, G.R., M.D. Anderson, W.L. Bland, J.M. Norman, J.M. Mecikalski, and R.M. Aune. 1998. Agricultural management decision aids driven by real-time satellite data. Bull. Am. Meteorol. Soc. 79:1345–1355.

Dikau, R. 1989. The application of a digital relief model to landform analysis in geomorphology, p. 51–77. *In* J. Raper (ed.) Three dimensional applications in geographic information systems. Taylor & Francis, London, U.K.

Dikau, R. 1990. Geomorphic landform modeling based on hierarchy theory, p. 230–239. *In* K. Brassel and H. Kishimoto (eds.) Proc. 4th Int. Sympos. Spatial Data Handling. Department of Geography, University of Zurich, Zurich, Switzerland.

Dikau, R., E.E. Brabb, and R.M. Mark. 1991. Landform classification of New Mexico by computer, p. 26. U.S. Department Interior, U.S. Geological Survey. Open-file report 91-634.

Dikau, R., E. Brabb, R. Mark, and R. Pike. 1995. Morphometric landform analysis of New Mexico. Z. Geomorphol. Suppl. 101:109–126.

Dobos, E., E. Micheli, M.F. Baumgardner, L. Biehl, and T. Helt. 2000. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma 97:367–391.

Dobos, E., L. Montanarella, T. Negre, and E. Micheli. 2001. A regional scale soil mapping approach using integrated AVHRR and DEM data. Int. J. Appl. Earth Obs. Geoinf. 3:30–41.

Dobson, M.C., and F.T. Ulaby. 1986. Active microwave soil moisture research. IEEE Trans. Geosci. Remote Sens. 24:23–36.

Drăgut, L., and T. Blaschke. 2006. Automated classification of landform elements using object-based image analysis. Geomorphology 81:330–344.

Edmonds, W.J., and J.B. Campbell. 1984. Spatial estimates of soil temperature. Soil Sci. 138:203–208.

Efron, B., and R.J. Tibshirani. 1993. An introduction to the bootstrap. Monographs on statistics and applied probability 57. Chapman & Hall, London, U.K.

Erskine, R.H., T.R. Green, J.A. Ramirez, and L.H. MacDonald. 2007. Digital elevation accuracy and grid cell size; effects on estimated terrain attributes. Soil Sci. Soc. Am. J. 71:1371–1380.

Evans, I.S. 1972. General geomorphometry, derivatives of altitude, and descriptive statistics, p. 17–90. *In* R.J. Chorley (ed.) Spatial analysis in geomorphology. Methuen, London, U.K.

Evans, I.S. 1998. What do terrain statistics really mean? *In* S.N. Lane, K.S. Richards, and H. Chandler (eds.) Land monitoring, modelling and analysis. John Wiley, Chichester, U.K.

Favrot, J.C. 1989. Une stratégie d'inventaire cartographique à grand échelle: La méthode des secteurs de reférence. Sci. du sol 27:351–368.

Favrot, J.C., and P. Lagacherie. 1993. La cartographie automatisee des sols: Une aide a la gestion ecologique des paysages ruraux. C.R. Acad. Agric. Fr. 79:61–76.

Fels, J.E., and K.C. Matson. 1996. A cognitively-based approach for hydrogeomorphic land classification using digital terrain models. 3rd Int. Conf. Integrating GIS Environmental Modeling. January 21–25. Santa Fe, NM. National Center for Geographic Information and Analysis, Santa Barbara, CA.

Fernández-Buces, N., C. Siebe, S. Cram, and J.L. Palacio. 2006. Mapping soil salinity using a combined spectral response index for bare soil and vegetation: A case study in the former lake Texcoco, Mexico. J. Arid Environ. 65:644–667.

Fidêncio, P.H., I. Ruisanchez, and R.J. Poppi. 2001. Application of artificial neural networks to the classification of soils from Saõ Paulo state using near-infrared spectroscopy. Analyst 126:2194–2200.

Fisher, R.A. 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7:179–188.

Fisher, P., J. Wood, and T. Cheng. 2004. Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. Trans. Inst. Br. Geogr. 29:106–128.

Fisher, P., J. Wood, and T. Cheng. 2005. Fuzziness and ambiguity in multi-scale analysis of landscape morphometry, p. 209–232. *In* M. Cobb, F. Petry, and V. Robinson (eds.) Fuzzy modelling with spatial information for geographic problems. Springer, New York.

Florinsky, I.V. 1998. Accuracy of land topographical variables derived from digital elevation models. Int. J. Geogr. Inf. Sci. 12:47–61.

Freund, Y., and R.E. Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55:119–139.

Fridland, V.M. 1972. Pattern of the soil cover. Israel program for scientific translations, Jerusalem, Israel.

Fung, A.K., Z. Li, and K.S. Chen. 1992. Backscattering from a randomly rough dielectric surface. IEEE Trans. Geosci. Remote Sens. 30:356–369.

Furley, P.A. 1968. Soil formation and slope development. 2. The relationship between soil formation and gradient angle in the Oxford area. Z. Geomorphol. 12:25–42.

Galdeano, A., F. Asfirane, C. Truffert, E. Egal, and N. Debeglia. 2001. The aeromagnetic map of the French Cadomian belt. Tectonophysics 331:99–122.

Gallant, J., and T. Dowling. 2003. A multi-resolution index of valley bottom flatness for mapping depositional areas. Water Resour. Res. 39(12):1347.

Gessler, P.E., I.D. Moore, N.J. McKenzie, and P.J. Ryan. 1995. Soil–landscape modelling and spatial prediction of soil attributes. Int. J. Geogr. Inf. Syst. 9:421–432.

Giltrap, D.J. 1977. Mathematical techniques for soil survey design. Ph.D. Thesis. University of Oxford. Oxford, U.K.

Girard, M.C. 1983. Recherche d'une modélisation en vue d'une représentation spatiale de la couverture pédologique. Application á une région des plateaux jurassiques de Bourgogne. Thèse d'Etat, Sols No. 12. INA-PG.

Gobron, N., B. Pinty, M.M. Verstraete, and J.-L. Widlowski. 2000. Advanced vegetation indices optimized for up-coming sensors: Design, performance and applications. IEEE Trans. Geosci. Remote Sens. 38:2489–2505.

Goetz, S.J., R.N. Halthore, F.G. Hall, and B.L. Markham. 1995. Surface temperature retrieval in a temperate grassland with multiresolution sensors. J. Geophys. Res. 100:25397–25410.

Goldberg, D.E. 1989. Genetic algorithms in search, optimisation and machine learning. Addison-Wesley, New York.

Goovaerts, P. 1999. Geostatistics in soil science: State-of-the-art and perspectives. Geoderma 89:1–45.

Gopal, S., and C. Woodcock. 1995. Theory and methods for accuracy assessment of thematic maps using fuzzy sets. Photogramm. Eng. Remote Sens. 60:181–188.

Goulard, M., and M. Voltz. 1992. Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. Math. Geol. 24:269–286.

Grasty, R.L., H. Mellander, and M. Parker. 1991. Airborne gamma-ray spectrometer surveying. International Atomic Energy Agency, Vienna, Austria.

Grimm, R., T. Behrens, M. Märker, and H. Elsenbeer. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using random forests analysis. Geoderma 146:102–113.

Grinand, C., D. Arrouays, B. Laroche, and M.P. Martin. 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. Geoderma 143:180–190.

Grunwald, S. 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152:195–207.

Grunwald, S., K. McSweeney, D.J. Rooney, and B. Lowery. 2001. Soil layer models created with profile cone penetrometer data. Geoderma 103:181–201.

Guzzetti, F., and P. Reichenbach. 1994. Toward the definition of topographic divisions for Italy. Geomorphology 11:57–75.

Hammond, E.H. 1954. Small-scale continental landform maps. Ann. Assoc. Am. Geogr. 44:33–42.

Hammond, E.H. 1964. Analysis of properties in landform geography: An application to broadscale landform mapping. Ann. Assoc. Am. Geogr. 54:11–19.

Hansen, M.K., D.J. Brown, P.E. Dennison, S.A. Graves, and R.S. Bricklemyer. 2009. Inductively mapping expert-derived soil–landscape units within dambo wetland catenae using multispectral and topographic data. Geoderma 150:72–84.

Hardy, S.M. 2004. The application and analysis of airborne gamma spectroscopy (radiometrics) for mapping soils in central Queensland. 13th Int. Soil Conservation organisation Conf. July 2004. Conserving Soil and Water for Society: Sharing Solutions, Brisbane, Australia.

Hartemink, A.E., A.B. McBratney, and M.L. Mendonça-Santos (eds.). 2008. Digital soil mapping with limited data. Springer, Dordrecht, the Netherlands.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The elements of statistical learning: Data mining, inference and prediction. 2nd edn. Springer Series in Statistics, Springer-Verlag, New York.

Hastie, T.J., and D. Pregibon. 1992. Generalized linear models, p. 195–248. *In* J.M. Chambers and T.J. Hastie (eds.) Statistical models in S. Wadsworth and Brooks, Pacific Grove, CA.

Hastie, T.J., and R.J. Tibshirani. 1990. Generalized additive models. Chapman & Hall, London, U.K.

Hay, R.L. 1960. Rate of clay formation and mineral alteration in a 4000-years-old volcanic ash soil on St. Vincent, B.W.I. Am. J. Sci. 258:354–368.

Henderson, R., and J.M. Ragg. 1980. A reappraisal of soil mapping in an area of Southern Scotland. Part II. The usefulness of some morphological properties and of a discriminant analysis in distinguishing between the dominant taxa of four mapping units. J. Soil Sci. 31:573–580.

Henderson, B., E. Bui, C. Moran, D. Simon, P. Carlile. 2001. ASRIS: Continental-scale soil property predictions from point data. CSIRO Land and Water Technical Report 28/01. CSIRO, Canberra, Australian Capital Territory Australia.

Henderson, B.L., E.N. Bui, C.J. Moran, and D.A.P. Simon. 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124:383–398.

Hengl, T., and R.A. MacMillan. 2009. Geomorphometry—A key to landscape mapping and modelling. Dev. Soil Sci. 33:433–460.

Hengl, T., D.G. Rossiter, and A. Stein. 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. Aust. J. Soil Res. 41:1403–1422.

Heuvelink, G.B.M. 1996. Identification of field attribute error under different models of spatial variation. Int. J. Geogr. Inf. Sci. 10:921–935.

Heuvelink, G.B.M. 1998. Error propagation in environmental modelling with GIS. Taylor & Francis, London, U.K.

Heuvelink, G.B.M., and P.A. Burrough. 2002. Developments in statistical approaches to spatial uncertainty and its propagation. Int. J. Geogr. Inf. Sci. 16:111–113.

Heuvelink, G.B.M., and R. Webster. 2001. Modelling soil variation: Past, present, and future. Geoderma 100:269–301.

Hewitt, A.E. 1993. Predictive modelling in soil survey. Soils Fertil. 56:305–314.

Hoeben, R., and P.A. Troch. 2000. Assimilation of active microwave observation data for soil moisture profile estimation. Water Resour. Res. 36:2805–2819.

Hudson, B.D. 1992. The soil survey as paradigm-based science. Soil Sci. Soc. Am. J. 56:836–841.

Huette, A.R. 1988. A soil adjusted vegetation index (SAVI). Remote Sens. Environ. 25:295–309.

Hutchinson, M.F. 1998a. Interpolation of rainfall data with thin plate smoothing splines. I. Two dimensional smoothing of data with short range correlation. J. Geogr. Inf. Decis. Anal. 2:152–167.

Hutchinson, M.F. 1998b. Interpolation of rainfall data with thin plate smoothing splines. II. Analysis of topographic dependence. J. Geogr. Inf. Decis. Anal. 2:168–185.

Hutchinson, M.F., and P.E. Gessler. 1994. Splines—More than just a smooth interpolator. Geoderma 62:45–67.

Irvin, B.J., S.J. Ventura, and B.K. Slater. 1997. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. Geoderma 77:137–154.

IPCC. 2006. 2006 IPCC Guidelines for National Greenhouse Gas Inventories, prepared by the National Greenhouse Gas Inventories Programme. IGES, Kanagawa, Japan.

Iwahashi, J., and R.J. Pike. 2007. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology 86:409–440.

Jackson, T.J., J. Schmugge, and E.T. Engman. 1996. Remote sensing applications to hydrology: Soil moisture. Hydrol. Sci. J. 41:517–530.

Jenny, H. 1941. Factors of soil formation, a system of quantitative pedology. McGraw-Hill, New York.

Jones, M.J. 1973. The organic matter content of the savanna soils of West Africa. J. Soil Sci. 24:42–53.

Kempen, B., D.J. Brus, G.B.M. Heuvelink, and J.J. Stoorvogel. 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. Geoderma 151:311–326.

Kienzle, S. 2004. The effect of DEM raster resolution on first order, second order and compound terrain derivatives. Trans. GIS 8:83–111.

King, D., M. Jamagne, D. Arrouays, D. Bornand, J.C. Favrot, R. Hardy, C. Le Bas, and P. Stengel. 1999. Inventaire cartographique et surveillance des sols en France. Etat d'advancement et exemples d'utilisation. Étude et Gestion des Sols 6:215–228.

King, T.V.V., R.N. Clark, C. Ager, and G.A. Swayze. 1995. Remote mineral mapping using AVIRIS data at Summitville, Colorado and the adjacent San Juan Mountains, p. 59–63. *In* H.H. Posey, J.A. Pendleton, and D. van Zyl (eds.) Summitville Forum '95. Colorado Geological Survey Special Publication 38. Colorado Geological Society, Boulder, CO.

Kiss, J.J., E. de Jong, and L.W. Martz. 1988. The distribution of fallout Cesium-137 in southern Saskatchewan, Canada. J. Environ. Qual. 17:445–452.

Knotters, M., D.J. Brus, and J.H. Oude Voshaar. 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma 67:227–246.

Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43:59–69.

Kuhn, T.S. 1996. The structure of scientific revolutions. University of Chicago Press, Chicago, IL.

Kustas, W.P., and J.M. Norman. 1996. Use of remote sensing for evapotranspiration monitoring over land surfaces. Hydrol. Sci. J. 41:495–516.

Laba, M., S.K. Gregory, J. Braden, D. Ogurcak, E. Hill, E. Fegraus, J. Fiore, and S.D. DeGloria. 2002. Conventional and fuzzy accuracy assessment of the New York gap analysis project land cover map. Remote Sens. Environ. 81:443–455.

Lagacherie, P. 1992. Formalisation des lois de distribution des sols pour automiser la cartographie pedologique a partir d'un secteur pris comme reference. Ph.D. Thesis. Université Montpellier II. Montpellier, France.

Lagacherie, P., and S. Holmes. 1997. Addressing geographical data errors in a classification tree soil unit prediction. Int. J. Geogr. Inf. Sci. 11:183–198.

Lagacherie, P., and A.B. McBratney. 2006. Spatial soil information systems and spatial soil inference systems: Perspectives for digital soil mapping, p. 3–22. *In* P. Lagacherie, A.B. McBraney, and M. Voltz. (eds.) Digital soil mapping: An introductory perspective. Elsevier, Amsterdam, the Netherlands.

Lagacherie, P., A.B. McBratney, and M. Voltz (eds.). 2006. Digital soil mapping: An introductory perspective. Developments in Soil Science. Vol. 31. Elsevier. Amsterdam, the Netherlands.

Lagacherie, P., J.M. Robbez-Masson, N. Nguyen-The, and J.P. Barthès. 2001. Mapping of reference area representativity using a mathematical soilscape distance. Geoderma 101:105–118.

Lagacherie, P., and M. Voltz. 2000. Predicting soil properties over a region using sample information from a mapped reference area and digital elevation data: A conditional probability approach. Geoderma 97:187–208.

Landmann, T., R. Colditz, and M. Schmidt. 2006. An object-conditional land cover classification system (LCCS) wetland probability detection method for West African savannas using 250-meter MODIS observations. European Space Agency Special Publication, ESA SP (SP-634). ESA Publications Division, Noordwijk, the Netherlands.

Lane, P.W. 2002. Generalized linear models in soil science. Eur. J. Soil Sci. 53:241–251.

Lang, M.W., G.W. McCarty, J.C. Ritchie, A.M. Sadeghi, W.D. Hively, and S.D. Eckles. 2008. Radar monitoring of wetland hydrology: Dynamic information for the assessment of ecosystem services. Int. Geosci. Remote Sens. Symp. 1:I261–I264.

Lark, R.M. 2000. Regression analysis with spatially autocorrelated error: Simulation studies and application to mapping of soil organic matter. Int. J. Geogr. Inf. Sci. 14:247–264.

Lark, R.M., T.F.A. Bishop, and R. Webster. 2007. Using expert knowledge with control of false discovery rate to select regressors for prediction of soil properties. Geoderma 138:65–78.

Lark, R.M., and A. Papritz. 2003. Fitting a linear model of coregionalization for soil properties using simulated annealing. Geoderma 115:245–260.

Lascano, R.J., R.L. Baumhardt, S.K. Hicks, and J.A. Landivar. 1998. Spatial and temporal distribution of surface water content in a large agricultural field, p. 19–30. *In* P.C. Robert, R.H. Rust, and W.E. Larson (eds.) Precision agriculture. ASA-CSSA-SSSA, Madison, WI.

Laslett, G.M., A.B. McBratney, P.J. Pahl, and M.F. Hutchinson. 1987. Comparison of several spatial prediction methods for soil pH. J. Soil Sci. 38:325–341.

Lawley, R., and B. Smith. 2007. Digital soil mapping at a national scale: A knowledge and GIS based approach to improving parent material and property information, p. 173–182. *In* A.E. Hartemink, A.B. McBratney, and M.L. Mendonça-Santos (eds.) Digital soil mapping with limited data. Springer, New York.

Lee, K.S., G.B. Lee, and E.J. Tyler. 1988. Determination of soil characteristics from thematic mapper data of a cropped organic–inorganic soil landscape. Soil Sci. Soc. Am. J. 52:1100–1104.

Legros, J.P. 1975. Occurrence des podzols dans l'Est du Massif Central. Sci. du Sol. 1:37–49.

Legros, J.P., and P. Bonneric. 1979. Modelisation informatique de la repartition des sols dans le Parc Naturel Régional du Pilat. Annales de l'Université de Savoie, Tome 4, Sciences Naturelles, 63–68.

Li, F., and T.J. Lyons. 2002. Remote estimation of regional evapotranspiration. Environ. Model. Softw. 17:61–75.

Lin, D.S., E.F. Wood, P.A. Troch, M. Mancini, and T.J. Jackson. 1994. Comparison of remote sensed and model simulated soil moisture over a heterogenous watershed. Rem. Sens. Environ. 48:159–171.

Liu, C., P. Frazier, and L. Kumar. 2006. Comparative assessment of the measures of thematic classification accuracy. Remote Sens. Environ. 107:606–616.

Liu, J., E. Pattey, and M.C. Nolin. 2008a. Object-based classification of high resolution SAR images for within field homogeneous zone delineation. Photogramm. Eng. Remote Sens. 74:1159–1168.

Liu, J., E. Pattey, M.C. Nolin, J.R. Miller, and O. Ka. 2008b. Mapping within-field soil drainage using remote sensing, DEM and apparent soil electrical conductivity. Geoderma 143:261–272.

Lund, E.D., P.E. Colin, D. Christy, and P.E. Drummond. 1999. Applying soil conductivity technology to precision agriculture, p. 1089–1100. *In* P.C. Robert, R.H. Rust, and W.E. Larson (eds.) Proc. 4th Int. Conf. Precision Agriculture. ASA-CSSA-SSSA, Madison, WI.

MacMillan, R.A., D.H. McNabb, and R.K. Jones. 2000a. Automated landform classification using DEMs: A conceptual framework for a multi-level, hierarchy of hydrologically and geomorphologically oriented physiographic mapping units. 4th Int. Conf. Integrating GIS Environmental Modeling (GIS/EM4). Problems, Prospects and Research Needs. September 2–8. Banff, Alberta, Canada.

MacMillan, R.A., D.E. Moon, and R.A. Coupé. 2007. Automated predictive ecological mapping in a forest region of B.C., Canada 2001–2005. Geoderma 140:353–373.

MacMillan, R.A., W.W. Pettapiece, S.C. Nolan, and T.W. Goddard. 2000b. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. Fuzzy Sets Syst. 113:81–109.

Malone, B.P., A.B. McBratney, and B. Minasny. 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154:138–152.

Mancini, M., R. Hoeben, and P.A. Troch. 1999. Multifrequency radar observations of bare surface soil moisture content: A laboratory experiment. Water Resour. Res. 35:1827–1838.

Marsman, B.A., and J.J. De Gruijter. 1986. Quality of soil maps, a comparison of soil survey methods in a study area. Soil Surv. Pap. No. 15. Netherlands Soil Survey Institute, Stiboka, Wageningen, the Netherlands.

Martens, G., and T. Naes. 1989. Multivariate calibration. John Wiley & Sons, New York.

Matt, P.B., and W.C. Johnson. 1996. Thermoluminescence and new ^{14}C age estimates for late quaternary loesses in southwestern Nebraska. Geomorphology 17:115–128.

Mayr, T.R., R.C. Palmer, and H.J. Cooke. 2008. Digital soil mapping using legacy data in the Eden Valley, UK, p. 291–302. *In* A.E. Hartemink, A.B. McBratney, and M.L. Mendonça-Santos (eds.) Digital soil mapping with limited data. Springer, New York.

Maxwell, J.C. 1870. On hills and dales. The London, Edinburgh and Dublin Philosophical Magazine. J. Sci. Ser. 4. 169:421–427.

McBratney, A.B., and J.J. De Gruijter. 1992. A continuum approach to soil classification by modified fuzzy-k means with extragrades. J. Soil Sci. 43:159–176.

McBratney, A.B., M.L. Mendonça-Santos, and B. Minasny. 2003. On digital soil mapping. Geoderma 117:3–52.

McBratney, A.B., B. Minasny, S. Cattle, and R.W. Vervoort. 2002. From pedotransfer functions to soil inference systems. Geoderma 109:41–73.

McBratney, A.B., and I.O.A. Odeh. 1997. Application of fuzzy sets in soil science: Fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77:85–113.

McBratney, A.B., I.O.A. Odeh, T.F.A. Bishop, M.S. Dunbar, and T.M. Shatar. 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97:293–327.

McBratney, A.B., and R. Webster. 1983. Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. J. Soil Sci. 34:137–162.

McCullagh, P., and J.A. Nelder. 1983. Generalized linear models. Cambridge University Press, Cambridge, U.K.

McKay, M.D., W.J. Conover, and R.J. Beckman. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 221:239–245.

McKenzie, N.J., and M.P. Austin. 1993. A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. Geoderma 57:329–355.

McKenzie, N.J., P.E. Gessler, P.J. Ryan, and D. O'Connel. 2000. The role of terrain analysis in soil mapping, p. 245–265. *In* J.P. Wilson and J.C. Gallant (eds.) Terrain analysis—Principles and applications. John Wiley & Sons, Inc., New York.

McKenzie, N.J., D.W. Jacquier, D.J. Maschmedt, E.A. Griffin, and D.M. Brough. 2005. ASRIS Australian Soil Resource Information System. Technical specifications. Version 1.5. October 2005. Australian Collaborative Evaluation Program (ACLEP), Canberra, Australian Capital Territory, Australia.

McKenzie, N.J., and P.J. Ryan. 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89:67–94.

McSweeny, K., B.K. Slater, R.D. Hammer, J.C. Bell, P.E. Gessler, and G.W. Peterson. 1994. Towards a new framework for modelling the soil–landscape continuum, p. 127–145. *In* Factors of soil formation: A 50th anniversary retrospective. SSSA Special Publication 33. SSSA, Madison, WI.

Mendonça-Santos, M.L., A.B. McBratney, and B. Minasny. 2006. Soil prediction with spatially decomposed environmental factors, p. 277–286. *In* P. Lagacherie, A.B. McBraney, and M. Voltz (eds.) Digital soil mapping: An introductory perspective. Elsevier, Amsterdam, the Netherlands.

Mendonça Santos, M.L., C. Guenat, C. Thevoz, F. Bureau, J.C. Vedy. 1997. Impacts of embanking on the soil-vegetation relationships in a floodplain ecosystem of a pre-alpine river. Global Ecol. Biogeogr. 6:339–348.

Milne, G. 1935. Some suggested units of classification and mapping particularly for East African soils. Soil Res. 4:183–198.

Minasny, B., and A.B. McBratney. 2002. The *neuro-m* method for fitting neural network parametric pedotransfer functions. Soil Sci. Soc. Am. J. 66:352–361.

Minasny, B., and A.B. McBratney. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32:1378–1388.

Minasny, B., and A.B. McBratney. 2007a. Spatial prediction of soil properties using EBLUP with a Matérn covariance function. Geoderma 140:324–336.

Minasny, B., and A.B. McBratney. 2007b. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. Geoderma 142:285–293.

Minasny, B., and A.B. McBratney. 2010. Methodologies for global soil mapping, p. 429–436. *In* J.L. Boettinger, D.W. Howell, A.C. Moore, A.E. Hartemink, and S. Kienast-Brown (eds.) Digital soil mapping: Bridging research, production, and environmental application. Springer, New York.

Minasny, B., A.B. McBratney, M.L. Mendonça-Santos, I.O.A. Odeh, and B. Guyon. 2006. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. Aust. J. Soil Res. 44:233–244.

Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Peterson. 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57:443–452.

Moore, I.D., R.B. Grayson, and A.R. Ladson. 1991b. Digital terrain modelling: A review of hydrological, geomorphological and biological applications. Hydrol. Processes 5:3–30.

Moore, D.M., B.G. Lees, and S.M. Davey. 1991a. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. Environ. Manag. 15:59–71.

Moran, C.J., and E. Bui. 2002. Spatial data mining for enhanced soil map modelling. Int. J. Geogr. Inf. Sci. 16:533–549.

Moran, M.S., Y. Inoue, and E.M. Barnes. 1997. Opportunities and limitations for image-based remote sensing in precision crop management. Remote Sens. Environ. 61:319–346.

Mücher, C.A., E.T. Steinnocher, F.P. Kressler, and C. Heunks. 2000. Land cover characterization and change detection for environmental monitoring of pan-Europe. Int. J. Remote Sens. 21:1159–1181.

Munyati, C. 2000. Wetland change detection on the Kafue Flats, Zambia, by classification of a multitemporal remote sensing image dataset. Int. J. Remote Sens. 21:1787–1806.

Murphy, P.N.C., J. Ogilvie, and P. Arp. 2009. Topographic modelling of soil moisture conditions: A comparison and verification of two models. Eur. J. Soil Sci. 60:94–109.

Nelson, A., H.I. Reuter, and P. Gessler. 2009. DEM production methods and sources, p. 65–85. *In* T. Hengl and H.I. Reuter (eds.) Geomorphometry: Concepts, software, applications. Developments in soil science. Vol. 33. Elsevier, Amsterdam, the Netherlands.

Nelson, M.A., and I.O.A. Odeh. 2009. Digital soil class mapping using legacy soil profile data: A comparison of a genetic algorithm and classification tree approach. Aust. J. Soil Res. 47:632.

New, M., M. Todd, M. Hulme, and P. Jones. 2001. Precipitation measurements and trends in the twentieth century. Int. J. Climatol. 21:1899–1922.

Nield, S.J., J.L. Boettinger, and R.D. Ramsey. 2007. Digitally mapping gypsic and natric soil areas using Landsat ETM Data. Soil Sci. Soc. Am. J. 71:245–252.

Noller, J.S. 2010. Applying geochronology in predictive digital mapping of soils, p. 43–53. *In* J.L. Boettinger, D.W. Howell, A.C. Moore, A.E. Hartemink, and S. Kienast-Brown (eds.) Digital soil mapping: Bridging research, production, and environmental application. Springer, New York.

Noy-Meir, I. 1974. Multivariate analysis of the semiarid vegetation in south-eastern Australia. II. Vegetation catena and environmental gradients. Aust. J. Bot. 22:115–140.

Nye, P.H., and D.J. Greenland. 1960. The soil under shifting cultivation. Commonwealth Bureau of Soils, Harpenden, U.K.

Obukhov, A.I., and D.S. Orlov. 1964. Spectral reflectivity of the major soil groups and possibility of using diffuse reflection in soil investigations. Soviet Soil Sci. 2:174–184.

Odeh, I.O.A., D.J. Chittleborough, and A.B. McBratney. 1991. Elucidation of soil–landform interrelationships by canonical ordination analysis. Geoderma 49:1–32.

Odeh, I.O.A., and A.B. McBratney. 2000. Using AVHRR images for spatial prediction of clay content in the lower Namoi Valley of eastern Australia. Geoderma 97:237–254.

Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1992. Soil pattern recognition with fuzzy-c-means: Application to classification and soil–landform interrelationships. Soil Sci. Soc. Am. J. 56:505–516.

Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63:197–214.

Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1995. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging. Geoderma 67:215–225.

Odeh, I.O.A., A.B. McBratney, and B.K. Slater. 1997. Predicting soil properties from ancillary information: Non-spatial models compared with geostatistical and combined methods, p. 1008–1019. *In* E.Y. Baafi and N.A. Schofield (eds.) Geostatistics Wollongong '96. Vol. 2. Kluwer Academic Publishers, Amsterdam, the Netherlands.

Oliver, M.A., R. Webster, and K. Slocum. 2000. Filtering SPOT imagery by kriging analysis. Int. J. Remote Sens. 21:735–752.

Opsomer, J.D., D. Ruppert, M.P. Wand, U. Holst, and O. Hossjer. 1999. Kriging with nonparametric variance function estimation. Biometrics 55:704–710.

Ørka, H.O., E. Næsset, and O.M. Bollandsås. 2009. Classifying species of individual trees by intensity and structure derived from airborne laser scanner data. Remote Sens. Environ. 113:1163–1174.

Pace, R.K., and R. Barry. 1997. Quick computations of regressions with a spatially autoregressive dependent variable. Geog. Anal. 29:232–247.

Pachepsky, Y.A., D.J. Timlin, and W.J. Rawls. 2001. Soil water retention as related to topographic variables. Soil Sci. Soc. Am. J. 65:1787–1795.

Pain, C.F., M.A. Craig, D.L. Gibson, and J.R. Wilford. 2001. Regolith–landform mapping: An Australian approach, p. 29–56. *In* P.T. Bobrowsky (ed.) Geoenvironmental mapping, method, theory and practice. A.A. Balkema, Swets and Zeitlinger Publishers, Rotterdam, the Netherlands.

Pal, S.K., S. Bandyopadhyay, and C.A. Murthy. 1998. Algorithms for generation of class boundaries. IEEE Trans. Syst. Man Cybern. 28:816–828.

Pal, S.K., S. Bandyopadhyay, and C.A. Murthy. 2001. Genetic classifiers for remotely sensed images: Comparison with standard methods. Int. J. Remote Sens. 22:2545–2569.

Palacios-Orueta, A., and S.L. Ustin. 1998. Remote sensing of soil properties in the Santa Monica mountains. I. Spectral analysis. Remote Sens. Environ. 65:170–183.

Park, S., K. McSweeney, and B. Lowery. 2001. Identification of the spatial distribution of soils using a process-based terrain characterisation. Geoderma 103:249–272.

Park, S.J., and L.G. Vlek. 2002. Prediction of three-dimensional soil spatial variability: A comparison of three environmental correlation techniques. Geoderma 109:117–140.

Pedrycz, W., and J. Waletzky. 1997. Fuzzy clustering with partial supervision. IEEE Trans. Syst. Man Cybern. B. 27:787–795.

Peng, W., D.B. Wheeler, J.C. Bell, and M.G. Krusemark. 2003. Delineating patterns of soil drainage class on bare soils using remote sensing analyses. Geoderma 115:261–279.

Pennock, D.J., B.J. Zebarth, and E. De Jong. 1987. Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada. Geoderma 40:297–315.

Peucker, T.K., and D.H. Douglas. 1975. Detection of surface specific points by local parallel processing of discrete terrain elevation data. Comput. Graphics Image Process. 4:375–387.

Phillips, J.D. 2001. The relative importance of intrinsic and extrinsic factors in pedodiversity. Ann. Assoc. Am. Geogr. 91:609–621.

Pickup, G., and A. Marks. 2000. Identifying large-scale erosion and deposition processes from airborne gamma radiometrics and digital elevation models in a weathered landscape. Earth Surf. Proc. Landf. 25:535–557.

Pickup, G., and A. Marks. 2001. Regional-scale sedimentation process models from airborne gamma ray remote sensing and digital elevation data. Earth Surf. Proc. Landf. 26:273–293.

Pike, R.J. 1988. The geometric signature: Quantifying landslide terrain types from digital elevation models. Math. Geol. 20:491–511.

Pike, R.J., I.S. Evans, and T. Hengl. 2009. Geomorphometry: A brief guide, p. 1–28. *In* T. Hengl and H.I. Reuter (eds.) Geomorphometry: Concepts, software, applications. Developments in soil science. Vol. 33. Elsevier, Amsterdam, the Netherlands.

Post, D.F., E.H. Horvath, W.M. Lucas, S.A. White, M.J. Ehasz, and A.K. Batchily. 1994. Relations between soil color and landsat reflectance on semiarid rangelands. Soil Sci. Soc. Am. J. 58:1809–1816.

Qi, F., A.-X. Zhu, M. Harrower, and J.E. Burt. 2006. Fuzzy soil mapping based on prototype category theory. Geoderma 136:774–787.

Qin, C.-Z., A.-X. Zhu, X. Shi, B.-L. Li, T. Pei, and C.-H. Zhou. In press. Quantification of spatial gradation of slope positions. Geomorphology 110:152–161.

Quinlan, J.R. 1992. Learning with continuous classes, p. 343–348. *In* Proc. 5th Australian Joint Conf. Artificial Intelligence. World Scientific, Singapore, Singapore.

Rawlins, B.G., B.P. Marchant, D. Smyth, C. Scheib, R.M. Lark, and C. Jordan. 2009. Airborne radiometric survey data and a DTM as covariates for regional scale mapping of soil organic carbon across Northern Ireland. Eur. J. Soil Sci. 60:44–54.

Rennó, C.D., A.D. Nobre, L.A. Cuartas, J.V. Soares, M.G. Hodnett, J. Tomasella, and M.J. Waterloo. 2008. HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. Remote Sens. Environ. 112:3469–3481.

Rodríguez, P.G., Ma.E.P. González, and A.G. Zaballos. 2007. Mapping of salt-affected soils using TM images. Int. J. Remote Sens. 28:2713–2722.

Romiæ, M., T. Hengl, D. Romiæ, and S. Husnjak. 2007. Representing soil pollution by heavy metals using continuous limitation scores. Comput. Geosci. 33:1316–1326.

Rosenberg, A. 2000. Philosophy of science: A contemporary introduction. Routledge, London, U.K.

Rouse, J.W., R.H. Haas, J.A. Schell, and D.W. Deering. 1973. Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symp. NASA Publication No. SP-351:309–317.

Ruhe, R. 1960. Elements of the soil landscape, p. 165–170. *In* Transactions of the 9th Congress of the International Society of Soil Science. Vol. 4. ISSS, Madison, WI.

RuleQuest Research. 2000. Cubist. RuleQuest Research Pty Ltd., Sydney, Australia.

Ryan, P.J., N.J. McKenzie, D. O'Connell, A.N. Loughhead, P.M. Leppert, D. Jacquier, and L. Ashton. 2000. Integrating forest soils information across scales: Spatial prediction of soil properties under Australian forests. Forest Ecol. Manag. 138:139–157.

Samra, J.S., W.A. Stahel, and H. Kunsch. 1991. Modeling tree growth sensitivity to soil sodicity with spatially correlated observations. Soil Sci. Soc. Am. J. 55:851–856.

Schaetzl, R.J., L.R. Barrett, and J.A. Winkler. 1994. Choosing models for soil chronofunctions and fitting them to data. Eur. J. Soil Sci. 45:219–232.

Schmidt, J., and R. Andrew. 2005. Multi-scale landform characterization. Area 37:341–350.

Schmidt, J., I.S. Evans, and J. Brinkmann. 2003. Comparison of polynomial models for land surface curvature calculation. Int. J. Geogr. Inf. Sci. 17:797–814.

Schmidt, J., and A. Hewitt. 2004. Fuzzy land element classification from DTMs based on geometry and terrain position. Geoderma 121:243–256.

Schmidt, J., P. Tonkin, and A. Hewitt. 2005. Quantitative soil–landscape models for the Haldon and Hurunui soil sets, New Zealand. Aust. J. Soil Res. 43:137–247.

Schulp, C.J.E., and A. Veldkamp. 2008. Long term landscape–land use interactions as explaining factor for soil organic matter variability in Dutch agricultural landscapes. Geoderma 146:457–465.

Scull, P., J. Franklin, and O.A. Chadwick. 2005. The application of decision tree analysis to soil type prediction in a desert landscape. Ecol. Model. 181:1–15.

Scull, P., J. Franklin, O.A. Chadwick, and D. McArthur. 2003. Predictive soil mapping: A review. Prog. Phys. Geogr. 27:171–197.

Seibert, J., J. Stendahl, and R. Sørensen. 2007. Topographical influences on soil properties in boreal forests. Geoderma 15:139–148.

Shary, P.A., L.S. Sharayab, and A.V. Mitusov. 2002. Fundamental quantitative methods of land surface analysis. Geoderma 107:1–32.

Sheng, J., L. Ma, P. Jiang, B. Li, F. Huang, and H. Wu. In press. Digital soil mapping to enable classification of the salt-affected soils in desert agro-ecological zones. Agric. Water Manag.

Shi, X., R. Long, R. DeKett, and J. Philippe. 2008. Integrating different types of knowledge for digital soil mapping. Soil Sci. Soc. Am. J. 73:1682–1692.

Shi, X., A.X. Zhu, J.E. Burt, F. Qi, and D. Simonson. 2004. A case-based reasoning approach to fuzzy soil mapping. Soil Sci. Soc. Am. J. 68:885–894.

Shi, X., A.X. Zhu, and R. Wang. 2005. Fuzzy representations special terrain feature using a similarity-based approach, p. 233–251. *In* F. Petry and V.B. Robinson (eds.) Fuzzy modeling with spatial information for geographic problems. Springer, New York.

Simonett, D.S. 1960. Soil genesis in basalt in North Queensland, p. 238–243. *In* Trans. 7th Int. Congr. Soil Sci. Madison, WI.

Sinha, A.K. 1990. Stratigraphic mapping of sedimentary formations in southern Ontario by ground electromagnetic methods. Geophysics 55:1148–1157.

Skidmore, A.K. 1989. An expert system classifies eucalypt forest types using Landsat Thematic Mapper and a digital terrain model. Photogramm. Eng. Remote Sens. 55:1449–1464.

Skidmore, A.K. 1990. Terrain position as mapped from gridded digital elevation data. Int. J. Geog. Inf. Syst. 4:33–49.

Skidmore, A.K., P.J. Ryan, W. Dawes, D. Short, and E. O'Loughlin. 1991. Use of an expert system to map forest soils from a geographical information system. Int. J. Geogr. Inf. Sci. 5:431–445.

Skidmore, A.K., C. Varekamp, L. Wilson, E. Knowles, and J. Delaney. 1997. Remote sensing of soils in a eucalypt forest environment. Int. J. Remote Sens. 18:39–56.

Skidmore, E.L. 2006. The role of vegetation for reducing wind erosion on military lands. American Society of Agronomy/Crop Science Society of America/Soil Science Society of America (ASA-CSSA-SSSA) Annual Meeting, November 12–16, 2006. Indianapolis, Indiana.

Smith, M.P., A.-X. Zhu, J.E. Burt, and C. Stiles. 2006. The effects of DEM resolution and neighborhood size on digital soil survey. Geoderma 137:58–69.

Soil Survey Staff. 1993. Soil survey manual. Handbook no. 18. USDA, Washington, DC.

Speight, J.G. 1990. Landform, p. 9–57. *In* R.C. McDonald, R.F. Isbell, J.R. Speight, J. Walker, and M.S. Hopkins (eds.) Australian soil and land survey-field handbook. 2nd Ed. Inkata Press, Melbourne, Australia.

Stafford, J.V., B. Ambler, R.M. Lark, and J. Catt. 1996. Mapping and interpreting the yield variation in cereal crops. Comput. Electron. Agric. 14:101–119.

Stockwell, D.R.B., and I.R. Noble. 1992. Induction of sets of rules from animal distribution data: A robust and informative method of data analysis. Math. Comput. Simul. 33:385–390.

Story, M., and R.G. Congalton. 1986. Accuracy assessment: A user's perspective. Photogramm. Eng. Remote Sens. 52:397–399.

Su, Z., P.A. Troch, and F.P. De Troch. 1997. Remote sensing of bare surface soil moisture using EMAC/ESAR data. Int. J. Remote Sens. 18:2105–2124.

Sudduth, K.A., S.T. Drummond, and N.R. Kitchen. 2001. Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. Comput. Electron. Agric. 31:239–264.

Susskind, J., P. Piraino, L. Rokke, L. Iredell, and A. Mehta. 1997. Characteristics of the TOVS pathfinder path. A data set. Bull. Am. Meteorol. Soc. 78:1449–1472.

Tabbagh, A., M. Dabas, A. Hesse, and C. Panissod. 2000. Soil resistivity: A non-invasive tool to map soil structure horizonation. Geoderma 97:393–404.

Thompson, J.A., J.C. Bell, and C.A. Butler. 1997. Quantitative soil–landscape modeling for estimating the areal extent of hydromorphic soils. Soil Sci. Soc. Am. J. 61:971–980.

Thompson, J.A., J.C. Bell, and C.A. Butler. 2001. Digital elevation model resolution: Effects on terrain attribute calculation and quantitative soil–landscape modeling. Geoderma 100:67–89.

Townsend, P.A. 2001. Mapping seasonal flooding in forested wetlands using multi-temporal Radarsat SAR. Photogramm. Eng. Remote Sens. 67:857–864.

Townsend, P.A., and S.J. Walsh. 1998. Modeling floodplain inundation using an integrated GIS with radar and optical remote sensing. Geomorphology 21:295–312.

Triantafilis, J., I.O.A. Odeh, B. Minasny, and A.B. McBratney. 2003. Elucidation of hydrogeological units using fuzzy k-means classification of EM34 data in the lower Namoi valley, Australia. Environ. Model. Softw. 18:667–680.

Triantafilis, J., W.T. Ward, I.O.A. Odeh, and A.B. McBratney. 2001. Creation and interpolation of continuous soil layer classes in the lower Namoi valley. Soil Sci. Soc. Am. J. 65:403–413.

Troeh, F.R. 1964. Landform parameters correlated to soil drainage. Soil Sci. Soc. Am. Proc. 28:808–812.

Twidale, C.R. 1985. Old land surfaces and their implications for models of landscape evolution. Rev. Géomorphol. Dyn. 34:131–147.

Van Niekerk, H.S., J. Gutzmer, N.J. Beukes, D. Phillips, and G.B. Kiviets. 1999. An $^{40}Ar/^{39}Ar$ age of supergene K-Mn oxyhydroxides in a post-Gondwana soil profile on the Highveld of South Africa. S. Afr. J. Sci. 95:450–454.

Vauclin, M., S.R. Vieira, G. Vachaud, and D.R. Nielsen. 1983. The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J. 47:175–184.

Venables, W.N., and B.D. Ripley. 1994. Modern applied statistics with S-PLUS. Springer-Verlag, New York.

Vold, A., T.A. Breland, and J.S. Soreng. 1999. Multiresponse estimation of parameter values in models of soil carbon and nitrogen dynamics. J. Agric. Biol. Environ. Sci. 4:290–309.

Voltz, M., and R. Webster. 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. J. Soil Sci. 41:473–490.

Wackernagel, H. 1987. Multivariate geostatistics. Springer, Berlin, Germany.

Walker, P.H., G.F. Hall, and R. Protz. 1968. Relation between landform parameters and soil properties. Soil Sci. Soc. Am. Proc. 32:101–104.

Walter, C., A.B. McBratney, A. Douaoui, and B. Minasny. 2001. Spatial prediction of topsoil salinity in the Chelif Valley, Algeria, using local ordinary kriging with local variograms versus whole-area variogram. Aust. J. Soil Res. 39:259–272.

Walvoort, D.J.J., and J.J. De Gruijter. 2001. Compositional kriging: A spatial interpolation method for compositional data. Math. Geol. 33:951–966.

Waring, R.H., N.C. Coops, W. Fan, and J.M. Nightingale. 2006. MODIS enhanced vegetation index predicts tree species richness across forested ecoregions in the contiguous U.S.A. Remote Sens. Environ. 103:218–226.

Webster, R. 1977. Canonical correlation in pedology: How useful? J. Soil Sci. 28:196–221.

Webster, R., and T.M. Burgess. 1980. Optimal interpolation and isarithmic mapping of soil properties. III. Changing drift and universal kriging. J. Soil Sci. 31:505–524.

Webster, R., and P.A. Burrough. 1974. Multiple discriminant analysis in soil survey. J. Soil Sci. 25:120–134.

Webster, R., T.R. Harrod, S.J. Staines, and D.V. Hogan. 1979. Grid sampling and computer mapping of the Ivybridge area, Devon. Technical monograph no. 12. Soil Survey of England and Wales, Harpenden, Herts, U.K.

Webster, R., and M.A. Oliver. 1990. Statistical methods in soil and land resource survey. Oxford University Press, Oxford, U.K.

Wen, R., and R. Sinding-Larsen. 1997. Image filtering by factorial kriging—Sensitivity analysis and application to GLORIA side-scan sonar images. Math. Geol. 21:433–468.

Wielemaker, W.G., S. de Bruin, G.F. Epema, and A. Veldkamp. 2001. Significance and application of the multi-hierarchical landsystem in soil mapping. Catena 43:15–34.

Wilford, J.R., P.N. Bierwirth, and M.A. Craig. 1997. Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. AGSO J. Aust. Geol. Geophys. 17:201–216.

Wilford, J.R., D.L. Dent, T. Dowling, and R. Braaten. 2001. Rapid mapping of soils and salt stores. Australian Geological Survey Organisation, Canberra, Australia. AGSO Res. Newslett. 34:33–40.

Wilson, J.P., and J.C. Gallant. 2000. Secondary topographic attributes, p. 87–132. *In* J.P. Wilson and J.C. Gallant (eds.) Terrain analysis—Principles and applications. John Wiley & Sons, Inc., New York.

Wright, C., and A. Gallant 2007. Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data. Remote Sens. Environ. 107:582–605.

Wood, E.F., D.S. Lin, and M. Mancini. 1993. Inter-comparisons between active and passive microwave remote sensing, and hydrological modelling for soil moisture. Adv. Space Res. 13:5167–5176.

Wood, J. 1996. The geomorphological characterisation of digital elevation models. Ph.D. Thesis. University of Leicester. Leicester, U.K.

Wrigley, N. 1978. Probability surface mapping, p. 39–49. *In* J.C. Davis and S. Levi de Lopez (eds.) Mapas por computadora para el analisis de los recursos naturales; memorias de la reunion internacional. Univ. Nac. Auton. Mex., Inst. Geogr. Mexico City & Univ. Kans., Kans. Geol. Surv., Lawrence, Kansas, Mexico City, Mexico.

Wu, W., Y. Fan, Z. Wang, and H. Liu. 2008. Assessing effects of digital elevation model resolutions on soil–landscape correlations in a hilly area. Agric. Ecosyst. Environ. 126:209–216.

Yaalon, D.H. 1975. Conceptual models in pedogenesis: Can soil-forming functions be solved? Geoderma 14:189–205.

Yokoyama, R., M. Shirasawa, and R. Pike. 2002. Visualizing topography by openness: A new application of image processing to digital elevation models. Photogramm. Eng. Remote Sens. 68:257–265.

Zadeh, L. 1965. Fuzzy sets. Inf. Control 8:338–353.

Zhang, W.H., and D.R. Montgomery. 1994. Digital elevation model grid size, landscape representation, and hydrologic simulations. Water Resour. Res. 30:1019–1028.

Zhu, A.X. 1997. A similarity model for representing soil spatial information. Geoderma 77:217–242.

Zhu, A.X. 2000. Mapping soil landscape as spatial continua: The neural network approach. Water Resour. Res. 36:663–677.

Zhu, A.X. 2006. Fuzzy logic models in soil science, p. 215–239. *In* S. Grunwald and M.E. Collins (eds.) Environmental soil–landscape modeling: Geographic information technologies and pedometrics. CRC/Taylor & Francis, New York.

Zhu, A.X. 2008. Rule based mapping, p. 273–291. *In* J.P. Wilson and A.S. Fotheringham (eds.) The handbook of geographic information science. Blackwell, Malden, MA.

Zhu, A.X., and L.E. Band. 1994. A knowledge-based approach to data integration for soil mapping. Can. J. Remote Sens. 20:408–418.

Zhu, A.X., L.E. Band, B. Dutton, and T.J. Nimlos. 1996. Automated soil inference under fuzzy logic. Ecol. Model. 90:123–145.

Zhu, A.X., L.E. Band, R. Vertessy, and B. Dutton. 1997. Derivation of soil properties using a soil land inference model (SoLIM). Soil Sci. Soc. Am. J. 61:523–533.

Zhu, C., and X. Yang. 1998. Study of remote sensing image texture analysis and classification using wavelet. Int. J. Remote Sens. 19:3197–3203.

38

Soil Change in the Anthropocene: Bridging Pedology, Land Use and Soil Management

Daniel deB. Richter, Jr.
Duke University

Arlene J. Tugel
United States Department of Agriculture

38.1 Overview and Objectives

Pedology is the science of soil change. Historically, this meant that pedology was mainly a basic science of natural soil processes and soil formations. Today, in the new geologic epoch that Crutzen (2002) has named the Anthropocene, pedology is fundamentally challenged to bring humanity entirely within the soil continuum (Dudal et al., 2002). During this period in which humanity exerts growing and predominant influence over Earth's systems, and especially Earth's soil (Yaalen 2000), pedology is becoming a much more interdisciplinary science with basic and applied objectives, one that requires the natural and social sciences but also the humanities (Richter, 2007).

In this chapter, discussions of "soil change" focus on the domestication of Earth's soil over historic time, the transformation of soil into a cultural-historic-natural system. This chapter describes human-altered soil processes and the forcing factors that are fundamentally changing soils across the planet, for like it or not, soils now serve as parent materials for humanity's accelerating influence. Humanity's impact on Earth's soils can now be called "global soil change" (Arnold et al., 1990).

The objectives of this chapter are to review the history of pedology as it responds to the challenges of the Anthropocene, consider soil change as it impacts soil function and ecosystem goods and services, evaluate the approaches to the study of soil change, enumerate on the factors and mechanisms affecting soil change that will improve prediction and management, and speculate on the future science, inventory, and management of soil change, specifically in the next few decades. The evolution of pedology is briefly described from its time as a basic, natural science that was focused narrowly on soil as a relatively slowly changing natural system, to its present and future as a broadly interdisciplinary environmental science that quantifies and predicts human-affected soil and soil-environment change on human timescales (years, decades, and centuries). More than anything, this chapter means to encourage the pedological community to redouble and expand the work of its pioneers such as Hilgard (1860), Darwin (1882), Dokuchaev (1883), Jenny (1941, 1961), Bidwell and Hole (1965), and Yaalon and Yaron (1966) to further integrate humanity as agents of soil formation, evaluate humanity's effects on soil function, and quantify and predict soil responses to human influence with smaller margins of error. We see this historical development as being entirely in parallel with the conceptual development of Earth's critical zone (National Research Council, 2001; Wilding and Lin, 2006), an integrative concept of the near-surface environment, first articulated by the landmark National Research Council (2001) report on the future of the earth sciences.

38.2 A Brief History of Human Influence on Soil

Soils have been used for agricultural and engineering purposes for at least 10,000 years. Human influence on soil was probably first affected by fire, but over the Holocene effects included a wide range of agricultural impacts; construction of villages, cities, and roads; leveling and terracing; irrigation and drainage; mining; and compaction and compaction and erosion. Human action is often goal oriented and is strongly affected by culture.

Human use, impact, and reliance on the soil are nothing short of staggering, and today with more than 6 billion persons on Earth, we look to the prospect in only a few decades of 10 billion persons (FAO-STAT, 2009). Over 2 billion Mg (metric tons) of cereals are harvested from soils per year, and over 1 billion Mg of vegetables plus roots and tubers; and from animals, over half billion Mg of milk, a quarter billion Mg of meat, and 50 million Mg of eggs; and from forests, more than 3 billion Mg of wood for fuel wood and industrial products. Currently, over 100 million Mg of N, P, and K are applied each year (increasing by three- to eightfold since 1960); about 250 million ha are irrigated, an area that has about doubled since 1970; and on the order of a million hectares of land is severely disturbed each year by mining. The growth of humanity's numbers and affluence means that these products of the soil will be doubled in the coming few decades and century. Both the extent and intensity of soil management are thus substantially increasing, and by any measure, we are pushing the Earth's soil very hard indeed.

More than half of the Earth's 13 billion ha of soil are today plowed, pastured, fertilized, limed, irrigated, drained, fumigated, bulldozed, puddled, compacted, eroded, leached, mined, reconstructed, or converted to new uses. Like it or not, humanity has become a major factor in soil formation and can prompt

TABLE 38.1 Soil Properties Grouped Qualitatively According to Rate of Change in Response to Human Forcings Such as Common Agricultural and Forestry Practices, Regional Air Pollution, Alterations in Hydrology, or Climate Change

Dynamic		Persistent
10 Years	$10–10^2$ Years	$>10^2$ Years
Acidity and salinity	Fragipans	Texture
pH-dependent charge	Fe/Al oxides	Rock volume
Bulk density and porosity	Occluded fractions of C, N, P	
Bioavailability of macro- and micronutrients and contaminants	Eluvial clay	
	Non-pH dependent charge	
	Fragipans, duripans, and plinthites	
Infiltration and hydraulic conductivity	Stabilized humic substances	
Rooting depth and volume		
Aggregates and structure		
Redoximorphic features		
Labile fractions of organic carbon		

change through single events such as tillage or prolonged phenomena such as global climate change. A number of soil properties are responsive to common agricultural or forestry practices, regional air pollution, and alterations in hydrology and landforms (Table 38.1). Two general patterns are evident: Most soil properties are *dynamic*, that is, significant changes can occur over timescales of centuries, decades or less, and only a few are relatively static or *persistent,* that is, properties that are little affected by human forcings. Table 38.1 is most striking perhaps for how it illustrates that so many important soil properties are susceptible to change on timescales of decades or less.

38.3 Pedology, Anthropedology, and Earth's Critical Zone

Of the widely recognized five soil-forming factors, pedology has historically treated organisms nearly always without humanity, except as an actor that disturbs, interrupts, manipulates, truncates, arrests, or subverts the natural process of soil formation. Yet, like it or not, human influence on soils worldwide has increased so extensively in the past few decades that there is now no alternative but to embrace humanity as part of the soil system, (Amund & Jenny 1991) even to recognize human influence as a sixth soil-forming factor (Amundson and Jenny, 1991; Dudal et al., 2002).

Viewed from the perspective of the Anthropocene (Crutzen, 2002), pedogenesis and morphology have for too long overemphasized the natural environment to the neglect of humanity as a soil-forming factor (Dudal et al., 2002; Ibáñez and Boixadera, 2002). Most soil surveys provide little information about the response of dynamic soil properties, such as organic matter, bulk density, and aggregate stability, to management and the effects of these changes on soil functioning. Despite this critique however, the historic pedological emphasis on the natural environment is understandable given how daunting the tasks for traditional pedology have proven (e.g., Richter and Babbar, 1991; Lal and Sanchez, 1992). Soil, the subject of pedology, has been documented to possess extreme spatial diversity from local to global scales, a diversity derived from high-order interactions of soil-forming factors. Soil is described as the most unparsimonious of all Earth's natural entities according to Johnson (2005), and the most complex of Earth's biomaterials according to Young and Crawford (2004).

One of pedology's most challenging tasks, however, is to understand how soils function and develop over human timescales. While soils can be destroyed and reinitiated in a moment's passing due to floods, mudflows, wind storms, volcanic ejecta, earthquakes, or a plow or a backhoe, soil formation also plays out over incredibly long sweeps of time. In the natural environment, stream terraces stable for even 1000s of years have soils that are youthful, whereas ancient soils are distinguished if they develop and survive for millions of years on biogeomorphically stable landforms. With soils ranging so widely across time, it is understandable why pedology has traditionally considered humanity to be more an

interruption than an integral part of soil formation. Yet in response to humanity's growing influence on Earth's systems, Bidwell and Hole (1965) and Yaalon and Yaron (1966) forcefully argued that pedology needs to embrace the vast activities of humanity. Bidwell and Hole (1965) considered not only hunting, gathering, and cultivation as integral to soil formation, but also watershed management and planning. And when Yaalon and Yaron (1966) used the term "metapedogenesis" that described human alteration of soil, pedology became a science of human-soil interactions.

Pedologists in the Anthropocene thus conceive of soil differently than did Hilgard (1860) and Dokuchaev (1883), whose pedological frontiers were focused explicitly on the natural formation of "virgin soil" according to Hilgard (1860). In the Anthropocene, pedology must increasingly focus on the science, inventory, and management of domesticated soils for purposes related to food production, human safety, and environmental quality. Because of the significance of this shift of focus, Yaalon and Yaron's (1966) ideas about anthropedogenesis can be recognized to be as fundamental to the development of pedology as the original model of soil as a natural body, attributed to Dokuchaev and Hilgard over a century ago.

No longer can soil be portrayed as it was in the past: as a relatively static component of ecosystems or as one formed by processes that do not include humanity. Recent pedological studies show how soil in the Anthropocene is a dynamic natural, historic, and cultural system subject to fundamental changes on human timescales. Boxell and Drohan (2009) found changes in soil morphology and physical properties with potential for increased runoff and erosion after *Bromus tectorum* L. (cheatgrass) invasion of a native sagebrush ecosystem in Western North America. After an evaluation of mature *Pinus palustris* Miller (longleaf pine) stands and intensively cultivated land use systems in the southeastern United States, Levi et al. (2010) discovered that near-surface soil properties were more similar by land use than by taxonomic-based soil map units.

Soil is an integral and interactive part of the rapidly changing ecosystem and wider environment and an important component of Earth's critical zone (National Research Council, 2001; Wilding and Lin, 2006), the solid surface of the planet where life is sustained by the interactions of physical, chemical, and biological processes in well integrated above- and belowground environments. Information about human impacts on these dynamic systems is needed to evaluate and predict the effects of land use and management on soil and its capacity to function. Recently, a new USDA-NRCS soil survey program was initiated to gather information about soil change in order to define potential and achievable values of dynamic soil properties for specific soils under a variety of management systems (Tugel et al., 2008).

38.4 Soil Change and Soil Function

Of great interest to anthropedologists are changes in soil functions in both natural and managed systems. Soil functions are described by what soil does and by how soil sustains, regulates, and controls the many biotic and abiotic processes in Earth's

pedosphere. In the Anthropocene, soil functions and ecosystem services are often parallel concepts for the evaluation and management of human impacts on ecosystems. Soil functions and ecosystem services have developed into critical discussions not only in the soil sciences literature but also in the literature of environmental change, ecosystem ecology, biogeochemistry, and land management. The general similarity of these concepts across disciplines is striking.

Soil functions relating to plant production, human and animal health, and the environment are widely discussed in the soil quality literature (Larson and Pierce, 1991; Parr et al., 1992; Blum and Santelises, 1994; Doran and Parkin, 1994; Karlen and Stott, 1994; Acton and Gregorich, 1995; Warkentin, 1995; Harris et al., 1996; Seybold et al., 1997; Carter et al., 2003; Andrews et al., 2004) and include (1) sustaining biological activity, diversity, and productivity; (2) regulating water and gaseous flows; (3) retaining and processing nutrients, organic compounds, and pollutants; (4) resisting structural and physical degradation; (5) partitioning and processing energy; and (6) supporting aesthetic and cultural attributes. For soil survey inventories, the emphasis is on fundamental functions such as nutrient and elemental cycling, hydrologic functions, and providing a medium for plant growth (Tugel et al., 2008).

The Millennium Ecosystem Assessment (Hassan et al., 2005) provides a broad definition that does not explicitly identify soil functions but alludes to them as supporting services. The report defines ecosystem services, some of which are soil functions, as follows: "Ecosystem services are the benefits people obtain from ecosystems." Soil functions are ecosystem services that include (1) provisioning services such as food and water (products); (2) the regulating of floods, drought, land degradation, and disease; (3) cultural services such as recreational, spiritual, religious, and other nonmaterial benefits; and (4) supporting services such as soil formation and nutrient cycling, which are essential for the production of provisioning, regulating, and cultural services. In a discussion of the economic valuation of services provided by natural ecosystems, Daily (1997) specifically describes services provided by soil, such as the physical support of plants, and places a value on that service using costs of a replacement practice, such as hydroponics.

In landscape assessments, Tongway and Ludwig (1997) use the term "functional" to describe how landscapes capture, retain, and use water and nutrients. In the procedures described in *Interpreting Indicators of Rangeland Health*, soil functional status is evaluated for biotic integrity, hydrologic function, and soil and site stability (Pyke et al., 2002; Pellant et al., 2005).

Across these definitions of soil function, the consequences of change are paramount. Arnold et al. (1990) suggest that the significance of soil change depends mainly on its reversibility. Irreparable change can compromise the ability of land managers to enhance or maintain soil functional capacity, produce goods and services, and minimize adverse effects on the environment. Minimizing the negative consequences of change so that future management goals and soil functions are not compromised is an overarching goal for long-term soil conservation and ecosystem

management and a primary concern with respect to soil degradation (Oldeman and van Lynden, 1997, Gerasimova et al., 2000). Inventorying and disseminating information about soil change and its consequences is a new challenge for soil survey.

38.5 Scientific Approaches to the Study of Soil Change

Because so many soil changes play out over years and decades (Table 38.1), special approaches are needed to quantify these dynamics. Although an understanding of soil change is vital for sustained land management, we cannot readily predict many soil changes with our current knowledge. Moreover, since soil changes result from high-order interactions of biology, chemistry, and physics, quantification of soil change requires direct observation of specific soils and their responses to management, at research sites over periods of decades. However, direct observation of changes over time in particular soils has been feasible for only a few soils. Five alternative and complimentary approaches are discussed: (1) long-term soil–ecosystem experiments (LTSEs), (2) short-term soil experiments, (3) space-for-time substitutions (SFTSs), (4) long-term resource monitoring (LTRM), and (5) computer models.

LTSEs are studies used to observe whole soil systems as they respond to controlled management regimes. To more fully understand the Earth's soil and its relations to the larger environment, there is no substitute for long-term observation, and repeated sampling, archiving, and analysis, in studies that pass from one generation of scientists to the next. Such foresighted research is challenging to establish and sustain. The first global inventory of such studies is being accumulated with descriptive meta-data on a real-time Web site (www.ltse.env.duke.edu).

Long-term experiments provide the base lines for understanding how and why soils change through time, for estimating rates, lags, and thresholds of change, and for documenting soil's resilience. However, long-running experiments are difficult to initiate and sustain; they require organization, data management, and unusual collaborations among scientists, even across generations. Long-term studies are susceptible to neglect or abandonment, and even productive LTSEs such as those described by Farina et al. (2000a, 2000b) in southern Africa can be summarily terminated simply due to an absence of interested scientists. Although LTSEs can be initiated with enthusiasm and the best of intentions, they may be terminated by lack of funding, shifts in research priorities, or societal instability.

Compared with complex experiments that require much labor and expense to maintain, long-term experiments that are labor efficient and straightforward in design are most likely to survive the tests of time. Even still, scientific returns from LTSEs may be few during their early years, and though challenging, the cycles of field work, samplings, archiving, and analyses must be balanced by scientific achievements. Long-term soil experiments face technical challenges as well (Steiner, 1995), as, for example, sampling soil with precision and accuracy is rarely easy and the statistical designs of many LTSEs can be questioned (Loughin,

2006). Some experimental treatments such as cultivation may eventually undermine the integrity of the experiment itself (Sibbesen, 1986). The representativeness of LTSEs is also not a trivial issue; for example, Debreczeni and Körschens (2003) estimated that >70% of the world's long-term field experiments are in Europe, and a recent web-based inventory of LTSEs (Richter et al., 2009) indicates that >80% of LTSEs test agricultural objectives and that >50% of these studies are located on relatively level Alfisols and Mollisols, many of which have remarkably high native fertility. Steiner (1995) suggested that the reason why so few long-term experiments test marginal soils was due to the unsustainability of such studies.

While LTSEs may be instrumental and necessary to the study of soil change they are by no means sufficient. Long-term studies need to be complemented by short-term soil–ecosystem experiments (STSEs) in the laboratory and field; by broadly focused SFTSs; by LTRM programs; and by computer modeling. Each is summarized in Tables 38.2 and 38.3 and discussed in detail in the following paragraphs.

STSEs in the lab and field include most studies of soils that are conducted up to a few years in duration. Nearly all of the sciences of soil physics, chemistry, and biology have been built from data arising from STSEs. Yet ironically, while STSEs often explore soil processes such as aggregation, adsorption, complexation, carbon dynamics, weathering, microbial activity, redox reactions, and soil fertility itself, extrapolating soil change over decadal timescales becomes a major challenge for STSEs. Small errors scaled across many years can readily bias long-term projections. Although STSEs can greatly enrich soil concepts and models, most are reductionist (isolating individual components

TABLE 38.2 Five Major Approaches in the Science and Management of Soil Change

Approach	Time Scale (Years)	Strengths	Limitations
STSEs	<1–10	Field or lab based, experimental control, versatile, short-term process	Extrapolation to larger scales of space and time, reductionist
LTSEs	>10	Field based, direct observation of whole soil system, experimental control, sample archive	Duration before useful data, vulnerable to loss, extrapolation to larger scales
SFTSs	>10–1000	Field based, highly time efficient	Temporal responses confounded entirely with spatial variation
LTRM	>10	Field based, direct observation, regional perspectives, sample archive	Complex planning and operational details, expense, no experimental control, duration before useful data
Computer models	<1 to >1000	Versatile, heuristic, and predictive, can support all other approaches	Dependent on quality of observational data

TABLE 38.3 Suitability of Measurement Approaches for Documenting Soil Change

Soil Change Attribute	Measurement Approach			
	STSE	LTSE	SFTS	LTRM
Variable at steady state (dynamic soil property)	R	S	M, R	S
Rate				
Short-term	F, R	F		F
Long-term		S	N	S
Fluctuation				
Short-term	F, R	F		F
Long-term		S		S
Trend, long-term		S		S
Pathway of change, feedback, thresholds, and hysteresis				
Short-term	F, R, S	F, S		F
Long-term		S	N	S
Resistance				
Short-term	R, S	S	M, R	S
Long-term		S	M, N	S
Resilience				
Short-term	R, S	S	M, R	S
Long-term		S	M, N	S

The three techniques, STSEs, LTSEs, and SFTSs, include experimental designs to facilitate comparisons among different treatments or kinds of land uses. LTRMs are not commonly designed to make comparisons among monitored sites or conditions but rather to simply track trends over time. Quantifying rates, pathways of change, and thresholds generally requires long-term studies or monitoring.

F, suitability depends on measurement frequency; M, suitability depends on management history at location sampled; N, suitability depends on the number of age classes sampled; R, suitability depends on characteristic response rate of the variable measured; S, suitability depends on time span over which repeat measurements are made. No entry indicates poorly suited.

and reactions), and do not examine the whole soil, complete with its high-order interactions and lag times that become apparent only with time in LTSEs. Even still, if STSEs are performed in conjunction with LTSEs, they can provide critical short-term process data important to the interpretation of longer-term soil change. Examples of short-term process data describe the seasonal dependence of nutrient mineralization, the depth-dependence of soil water chemistry, the rate at which annually produced plant-organic matter is harvested or returned to the soil as detritus, or management susceptibility of biologically mediated aggregate formation.

SFTSs, also called *chronosequence studies* or soil comparison studies, are used to efficiently examine temporal change in soil, ecosystems, and landscapes (Pickett, 1989; Hotchkiss et al., 2000). For questions of how soils change with time, SFTSs sidestep the great burden of LTSEs, the need for time itself to pass to directly acquire the information about soil change. One of the most well-known examples is that of Jenny's Mendocino Staircase in California, which is a set of marine terraces tectonically raised over the duration of the Pleistocene, with all soil-forming factors except time assumed to be similar (Jenny, 1980). The SFTSs are especially well suited for studies of change over geologic time and as such they are widely used to describe soil and ecosystem change over multimillennia (Jenny, 1980). SFTSs are, however, indirect in their inquiry, purposefully confound space and time, but thereby leave open the possibility for entirely faulty interpretation if misused. As an example, in a study of soil formation on mine reclaimed land, a poorly designed SFTS with improper site selection might over- or underestimate rates of soil formation on acidic mine spoils if recent soils are composed of parent materials with greater or lesser acidity than those used for older soils (Pickett, 1989). Many scientists are skeptical about the use of SFTSs (Gleason, 1927; Hotchkiss et al., 2000; Buol et al., 2003). On the other hand, only SFTSs can describe many soil changes that operate over many centuries and, if carefully designed to control for physical environment, are useful for documenting change in situations where repeating observations is not feasible (Ballantine and Schneider, 2009).

Carefully initiated comparison studies are being used in the soil survey programs of the Natural Resources Conservation Service to document land use and management effects on dynamic soil properties (Tugel et al., 2008). These studies apply the SFTS technique to make statements about change over time. They manage the limitations owing to variable historical conditions and inherent spatial variability of soil properties by restricting study sites to those with similar management history on closely similar soil and by replicating data collection across multiple spatial and temporal scales. They use state and transition models or other simple conceptual models of cause and effect to derive testable hypothesis and organize results. Selecting extensive soil map units for soil survey comparison studies as well as other SFTSs, LTSEs, and STSEs facilitates the extension of results across broad areas. LTRM programs are designed to document regional or national trends in soil resource condition over time. National monitoring programs generally use a random sampling procedure to select permanent monitoring sites where repeated observations of land use and resource condition are made at multiyear intervals. The tremendous cost of field sampling at all monitoring points is a major barrier to national scale soil monitoring programs. The Forest Inventory and Assessment reduced field sampling after initial efforts (Bechtold and Patterson, 2005) and the Natural Resources Inventory conducted field data collection for a few selected regions on an exploratory basis (Brejda et al., 2001). Vital Signs monitoring programs of the National Park Service were initiated in the early 2000s and to date have limited data to report (Fancy et al., 2009). Long-term monitoring is not limited to national scale monitoring programs. Some land managers monitor specific fields, pastures, timber stands, and other ecosystems as a part of their overall operations. Typical monitoring objectives will vary with the situation and include (1) evaluation and documentation of the progress toward management goals, (2) detection of changes that may be an early warning of future degradation, and (3) determination of trends for areas in desired condition, at risk, or with potential for recovery (Elzinga et al., 1998; Herrick et al., 2005; Madson et al., 2006).

Monitoring data produced by individual land managers are generally site specific in terms of objectives and properties measured and can be difficult to extrapolate. Compared with LTSEs, national resource monitoring programs quantify changes in soils not under experimental control but at randomly selected points across the landscape. Such LTRM programs thus have the potential to quantify soil change regionally as affected by shifts in regional land uses or other environmental conditions. But because management practices are not controlled, LTRM programs are challenged to interpret causes of observed dynamics. Causes of regional changes in soil carbon in two notable LTRM programs in England and Wales (Bellamy et al., 2005) and across Belgium (van Wesemael et al., 2004) have been controversial (Smith et al., 2007).

Computer models offer an approach to understanding soil change from the instantaneous to the multimillennial (Parton et al., 1987; Bouma and Hack-ten Broek, 1993; Pulleman et al., 2000; McBratney et al., 2002). The role and rationale of models are as much heuristic as they are predictive, as they represent refined hypotheses and depend on linkages with empirical studies. Models are instrumental for making progress in understanding soil change whatever the timescale, and are most convincing when simulation results are comparable with observational data. Whether observations are from LTSEs, STSEs, SFTSs, or LTMP's, empirical data are critical for gauging model competence and performance.

38.6 Forcing Factors, Process, Resistance, and Resilience

Understanding the causes of soil change and their effects on function requires interdisciplinary analysis at multiple spatial and temporal scales (Dent et al., 1996) as well as reductionistic basic research (Bouma, 1997). The challenge of studying open, dynamic systems is that there are few simple cause and effect relationships. Yet, motivating the study of soil change is the understanding that humanity is an integral part of the soil system, an agent affecting the flow through soils of energy, water, gases, and solid materials (Bidwell and Hole, 1965; Yaalon and Yaron, 1966), not simply an external player disturbing or perturbing the system (Goudie, 2005). Human activities impose a variety of forces on soil to affect the processes essential to the production of goods and services. This section provides a brief discussion on how forcing factors and processes involved in soil change are described by various disciplines. Also included are perspectives on the soil's ability to resist and recover from such forcings.

38.6.1 Forcing Factors

Forces driving ecosystem change are commonly described as forcing factors, stressors, disturbances, and perturbations. Although forcings often have negative connotations, their impacts can of course be positive or negative. Stressors and disturbances refer to factors that cause significant changes in ecological processes as well as the plant, soil, and animal components in ecological systems (Barrett et al., 1976; White, 1979). Stressors are forces that can be continuous, cyclic, or intermittent and include climate change, prolonged drought, agricultural impacts, grazing, fire suppression, and proliferation of invasive species. We use "disturbance" to represent relatively discrete events in time. Each occurrence of the following is an example: plowing, fertilization, irrigation, brush removal, fire, flooding, short-term drought, high-intensity storms, and wind storms. Stressors have a number of features in common. They can be either natural or anthropogenic and an integral feature of natural systems or a necessary operation in managed systems. Forcing factors can also produce direct impacts on small to large areas, which can lead to impacts on other areas. For example, climate change, the cumulative result of many forcing factors, will impact the soil system on large and small areas.

Disturbance regimes, such as recurring fire or flooding, or repeated tillage of agricultural fields are characteristic of many systems. A change in frequency or intensity of a forcing factor can place new stresses on a system. Five attributes are used to define disturbance regimes: disturbance type, spatial scale, intensity, frequency, and predictability (Herrick et al., 1999). A disturbance matrix can be constructed to help compare and understand the effect of disturbances on process (Table 38.4). In this example, fire, logging, and grazing remove aboveground vegetation that affects the supply of organic matter to the system. Logging, grazing, and vehicle traffic may compact the soil, thereby affecting runoff, water availability, soil aeration, and nutrient return to the system. In the Anthropocene, disturbance can also include an absence of fire, as fire has been removed from the landscape over many 100s of millions of hectares with major effects on soils and ecosystems.

The length of time required for a soil to change in response to outside forces is widely variable. It may take decades or longer for the cumulative effects of stressors to cause a functionally important change in soil. Discrete disturbance events, however, can precipitate almost immediate change such as the decrease in soil bulk density that occurs with the first tillage operation of the growing season (Pikul et al., 2006). Episodic events, such as hurricanes and drought, are events that may trigger a shift in systems that have experienced gradual change resulting from long-term management impacts (Scheffer et al., 2001). The combined effects of discrete events and longer-term stress can sometimes be devastating, such as those experienced in the Dust Bowl when continuous cultivation and prolonged drought were followed by intense wind storms.

38.6.2 Soil Processes

Predicting and managing the effects of human actions on soil function and ecosystem services require a detailed understanding of processes that operate within the pedosphere. However, there is tremendous complexity in the processes involved in soil change. Ross (1989) describes four increasingly complex and interlinked soil processes: (1) processes that form soil materials,

TABLE 38.4 Disturbance Matrix Illustrating a Classification of Disturbances and Their Effects on Soil Properties and Processes

Event	Pattern	Biomass Removal	Soil Compaction	Nutrient Return Form	Nutrient Return Distribution
Wildfire (hot, long-duration burn)	Continuous	All	Diffuse[a]	Mineral, black carbon	Concentrated and diffuse
Selective logging	Patchy	Woody	Linear and patches	Unprocessed and processed, organic	Depends on management of logging slash
Grazing (moderate)	Patchy	Herbaceous	Linear and patches	Mineral and processed organic	Discrete, concentrated (dung), diffuse (trampled vegetation)
Vehicle traffic (off-road)	Corridors	All within wheel tracks	Linear	Negligible within track	Negligible

Source: Modified from Herrick, J.E., and W.G. Whitford. 1999. Integrating soil processes into management: From microaggregates to macro-catchments, p. 91–95. *In* D. Eldridge and D. Freudenberger (eds.) People and rangelands: Building the future. Proc. Int. Rangeland Congr. Vol. 1. Townsville, Australia. July 19–23, 1999. International Rangeland Congress, Inc., Townsville, Australia.

[a] The effect here is indirect and maybe delayed: Fire may increase soil hydrophobicity, crusting, and compaction by removing protective vegetation and litter layers.
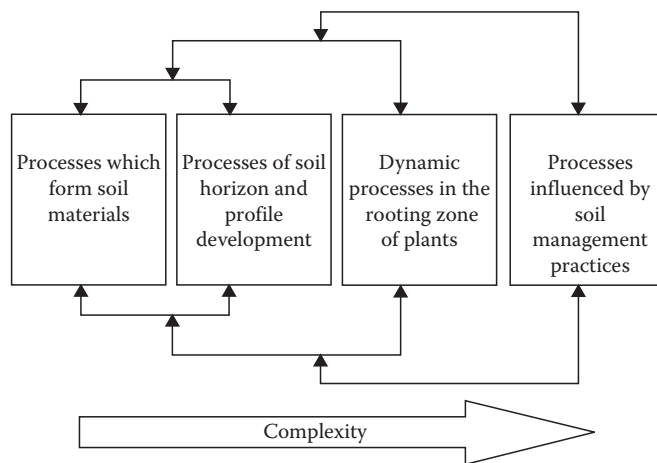
FIGURE 38.1 Processes involved in soil change. Soil systems function through many interacting processes. (Modified from Ross, S. 1989. Soil processes: A systematic approach. Routledge, London, U.K.)

(2) processes of soil horizon and profile development, (3) dynamic processes in the rooting zone of plants, and (4) processes influenced by management practices (Figure 38.1). Acknowledging that probably no soil process is uniquely biogeochemical, pedogenic, ecological, or geomorphic, it is important to draw upon many disciplines in our efforts to understand soil change. In the Anthropocene, it is important to join these traditions to enrich and broaden perspectives of soil and what has recently been called Earth's critical zone (National Research Council, 2001; Richter and Mobley, 2009). Pedology may focus on processes that alter soil material, geo- and biogeochemistry on chemical reactions, and ecology on system interactions, but the system is the same and the needs for understanding changes in the system justify a more integrated critical zone science (Wilding and Lin, 2006). The four primary processes of additions, losses, transformations, and translocations, originally formulated by the pedologist, Simonson (1959), are biogeochemical and ecological processes discussed in detail by Chadwick and Graham (2000)

and Richter and Markewitz (2001). Distinctions between disciplines become blurred when considering biologically mediated processes such as the physical-displacement forces applied to soil by growing roots (Brimhall et al., 1991; Richter et al., 2007) or porosity-dependent, acid-induced transformations and weathering of soil minerals affected by soil microbes and plant respiration (Richter and Markewitz, 1995).

Processes related to forcing factors and plant growth are a primary focus in disciplines such as agronomy and silviculture, as well as the sciences of desertification and land degradation. Land use conversions and agricultural practices can quickly alter existing soil processes and even introduce new processes to the soil system (Ross, 1989). The alteration of soil physical and chemical properties and processes through plowing, drainage, and additions of new materials alter soil physical, chemical, and biological processes and primarily affect (1) soil structure, (2) porosity, (3) soil moisture and temperature dynamics, (4) heat flux, (5) soil trafficability and workability, and (6) nutrient storage and release. Tillage may control weeds, but it may also break down soil aggregates and promote soil compaction, processes that have generally negative consequences to hydrologic soil function and nutrient cycling.

Studies of soil and land degradation and desertification are concerned with processes that substantially decrease an area's biological productivity or usefulness to humanity. Soil degradation itself is defined as "a process that describes human-induced phenomena, which lower the current and/or future capacity of the soil to support human life" (Oldeman and van Lynden, 1997). Water erosion, wind erosion, salinization, and chemical, physical, and biological deterioration are among the many soil degradation processes enumerated in the literature and of grave concern with respect to the production of goods and services. The set of processes mapped in the Global Assessment of the Status of Human-induced Soil Degradation, while commonly cited, is not comprehensive and is limited to those processes occurring across extensive areas of the world (Oldeman et al., 1991; Chen et al., 2002; Lal et al., 2004).

Synthesis of process information from a wide range of disciplines is essential to understand and predict human impacts on soil systems over time. Palm et al. (2007) advocate identifying relationships between land management, degradation processes, and impacts of those processes. Positive or negative feedbacks among soils, plants, and animals, the atmosphere and climate, and management can together lead to soil change (Lawrence et al., 2007) and impact the functional capacity of soil at a number of spatial scales (Herrick and Whitford, 1999). Furthermore, a specific land management practice can trigger more than one type of soil degradation process and one type of impact (Stocking, 1995).

38.6.3 Resistance and Resilience

The response of soils to forces and stressors is controlled by soil's ability to absorb, resist, and recover from the forces. The concepts resistance, resilience, and buffering help describe soils' response. Ecological discussion over concepts of resistance and resilience is extensive and vigorous (Holling, 1973; Pimm, 1984; Gunderson, 2000; Briske et al., 2008; Bestelmeyer et al., 2009a), a discussion that extends to engineering and pedology (Greenland and Szablocs, 1994; Holling, 1996). From perspective of the soil, resistance is the capacity to maintain functional capacity through disturbances, and resilience is the capacity to recover the functional and structural integrity lost after a disturbance or prolonged stress (Blum, 1997; Seybold et al., 1999). Some concepts of resilience incorporate resistance to change as well as the capacity to recover (Holling, 1973; Szabolcs, 1994; Gunderson, 2000; Tenywa et al., 2006). Quantitative estimates of soil resilience are based on pre- and postconditions and computed using process rates (Szabolcs, 1994), dynamic soil properties (Seybold et al., 1999), or a combination of properties and processes (Lal, 1994). Soil resilience differs somewhat from ecological resilience, as ecological resilience involves multiple subsystems of the ecosystem (e.g., plant community, soil fauna) (Walker et al., 2004). Soil resilience is therefore a component of ecological resilience.

One of the interests in soil resilience is to estimate the possibility of reversing change in soil and soil function. Arnold et al. (1990) offer an approach for estimating reversibility that uses a ratio of rates of change that gives insight to the time frame required for reversal. From a land management perspective, the length of time required to reverse a change can have a significant influence on the perception of the ease of reversibility and its costs. Reversibility is characterized using rate or velocity of change and is classified on the basis of the ratio of the primary change process to the reverse process. If both the primary change and reversal processes are either rapid or slow, the system is considered reversible. This approach also illustrates that the reversal process is often different from the primary change process thus exhibiting pronounced hysteresis.

While few soil changes may be truly irreversible, reverse changes that may prove intractable include severe erosion, salinization without a suitable outlet for drainage water or salts,

drainage and oxidation of sulfidic soil materials (iron pyrite, FeS_2), and heavy metal contamination.

38.7 Spatial and Temporal Patterns of Soil Change: Attributes for Prediction

The ability to predict how soils change over human timescales is critical to the success of sustainable soil management. Dynamic soil properties can be used to document change as well as quantify processes and soil functional capacity. Predicting change and its effect on processes and function requires observational data and models to describe and simulate temporal patterns exhibited by dynamic soil properties. Measuring temporal variation across space, however, as is required in LTSEs, STSEs, and SFTSs, is complicated by inherent and human-induced spatial variability of soil properties. Spatial and temporal variability of dynamic soil properties within systems are treated in this section.

38.7.1 Dynamic Soil Properties

When studying soil change within the Anthropocene, the variables of particular interest are dynamic soil properties (Tugel et al., 2005). Dynamic soil properties include use-dependent properties (Grossman et al., 2001). Many dynamic properties are indicative of soil function (see Chapter 26 of *Handbook of Soil Sciences: Resource Management and Environmental Impacts*) and recently soil survey document attainable levels and change in properties and function (Tugel et al., 2008).

Guidelines for selecting a minimum data set of dynamic soil properties for monitoring and soil survey comparison studies focus on functions of interest as well as ease of measurement, reproducibility, and cost (Larson and Pierce, 1991). Minimum soil-data sets typically will include bulk density, aggregate stability, infiltration, organic matter, pH, EC, exchangeable cations, extractable P, and potentially mineralizable N, which help characterize key soil functions and processes. Biological measures such as microbial biomass-C are also included where time of sampling can be controlled or seasonal fluctuations can be modeled.

In addition to land use and management, dynamic soil properties are affected by changes that are gradual, for example, a slowly warming climate or a prolonged drought, and that are oscillatory, for example, diurnal and seasonal regimes of temperature or moisture (Table 38.5). The dynamic soil properties' sensitivity to diurnal and seasonal environmental fluctuations may vary greatly and should be considered when selecting properties for inclusion in a study, especially SFTSs and LTRMs where seasonal measurements are not likely to be made. Repeated observations of properties sensitive to environmental fluctuations can be misinterpreted as management-related differences unless they are measured at the same time of day (diurnal sensitivity) or year (seasonal sensitivity). Environmental fluctuations

TABLE 38.5 Sensitivity of Soil Properties to Environmental Fluctuation in Temperature and Moisture

Sensitivity	
Diurnal	Seasonal
Soil temperature	Soil temperature
Composition of soil air	Moisture content
Redox potential	Bulk density, total porosity
	Aggregate stability
	Infiltration rate, hydraulic conductivity
	Bioavailability of macro- and micronutrients
	Microbial activity
	Electrical conductivity
	pH
	Redoximorphic features

Many soil properties respond to environmentally induced fluctuations in temperature and moisture. The environmental sensitivity should be considered when designing a study to ensure selection of suitable properties and the proper timing and frequency of measurement.

and anticipated changes in trends should be considered when designing a study to ensure proper selection of suitable properties, timing and frequency of measurement, and overall duration of study.

38.7.2 Spatial Patterns

The spatial variability of dynamic soil properties depends upon the type of soil, the plant community, the scale of disturbance, and the management history as well as the scale of measurement (Wilding et al., 1994). Legacies from past land use and management can persist in modern systems, such as tillage-related decreases in soil organic matter, contributing to spatial variability of dynamic soil properties (Foster et al., 2003; McLauchlan, 2006; Li et al., 2010). Understanding legacies and historic ranges of anthropogenic and nonanthropogenic variability is essential in interpreting modern changes in soil, although this knowledge does not include all of the information necessary for predicting future change (Millar and Woolfenden, 1999; Parsons et al., 1999).

The spatial variability of dynamic soil properties is often associated with spatial patterns of vegetation or local terrain, such as coppiced shrubs and concave intershrub spaces. Information about these patterns is important for the prediction of soil change and its effects on function. An example includes rows and furrows within a field and their effect on (1) salt redistribution within the soil after irrigation and (2) toxic salt damage to the crop (Wadleigh and Fireman, 1949; Miyamoto and Cruz, 1987; Ashraf and Saeed, 2006).

38.7.3 Temporal Patterns

Timescales over which dynamic soil properties change in response to land use and management vary from minutes to hundreds of years or more (Sparling, 2006; Richter, 2007).

Fluctuation, trend, rate, pathway of change, feedback, hysteresis, thresholds, and fluxes are used to describe temporal patterns of soil change. Study objectives and the suitability of the various measurement techniques for documenting these attributes should be considered when initiating a project to document change. Detailed descriptions of the following attributes of change are located in Arnold et al. (1990) and Tugel et al. (2008).

38.7.3.1 Fluctuations and Trends

Fluctuation is temporal variation in soil properties and includes nonsystematic, random variation and regular periodic, cyclic variation.

Trend is the general direction of change and can be increasing, decreasing, or steady-state equilibrium. The time it takes to approach or a dynamic equilibrium varies with the property, the kind of soil, the type of management, and any continuing forcing factors.

38.7.3.2 Rate, Pathways of Change, and Feedbacks

Rates of soil change are rarely constant. Temporal variation in dynamic soil properties can follow pathways that are logistic or exponential and are uncommonly linear, owing to fluctuations and feedbacks. Feedback mechanisms are involved when the rate changes. Positive feedbacks intensify process rates, and negative feedbacks diminish or limit rates. Where feedbacks accelerate process rates, such as increasing rates of surface runoff and erosion over time, the feedback is positive, though the result may have a negative effect on soil function. Plant–soil feedbacks have been suggested to be strongest for plants growing in extreme environments (Ehrenfeld et al., 2005).

38.7.3.3 Thresholds

Thresholds in soils are tipping points and have application to many processes (Chadwick and Chorover, 2001). In ecosystem management, an ecological threshold represents the conditions at a point in time after which future management options become limited and corresponds to the shift from one alternative state to another (Hobbs and Harris, 2001; Bestelmeyer, 2006; Bestelmeyer et al., 2009b). Recognition of a threshold implies that a functionally important change in process rate has occurred.

38.7.3.4 Fluxes

Fluxes in soils can be energetic or material and occur via the gas, liquid, or solid phases. The importance of fluxes is underscored by Simonson's (1959) classical view of the processes of pedogenesis in which three of four of Simonson's general processes involved material and energy fluxes: inputs, removals, and translocations. Fluxes are measured as the rate of movement of chemical elements, or materials such as organic matter, clay minerals, or colloids. Fluxes are key to the functioning of the internal soil system and to how and why soil interacts with the wider environment. Much about soil management, in the past and in the future, involves the management of energetic and material fluxes.

Examples of how inputs, removals, and translocations have altered soils are far too few. The best examples of research and management sites that link changes in soil systems with fluxes of materials into, through, and out of the soil over decadal timescales are long-term soil experiments (Richter et al., 2007).

38.7.4 Reorganization in Space and Time

Changes in process rates, feedbacks, and thresholds are involved in pattern reorganization within systems. Because dynamic soil properties reflect processes, property–process–pattern relationships can be established. Furthermore, spatial patterns of soil properties and plant communities can shift and reorganize through time in response to stressors, disturbances, and vegetation dynamics (Bestelmeyer et al., 2006; Ravi et al., 2010).

Awareness of property–process–pattern reorganization is essential for prediction of future changes. In one example (Figure 38.2), heavy continuous grazing followed by drought may produce positive feedbacks between vegetation and soil properties that intensify physical, chemical, and biological degradation. The feedbacks lead to the following: (a) a decrease in soil organic matter and an increase in size of bare spaces, (b) a decrease in soil aggregate stability and reduced resistance to erosion, (c) a loss of topsoil through erosion and a decrease in infiltration, and (d) an additional loss of grass and increase in shrubs, which cause the feedback loop to continue (Bird et al., 2001). Eventually, a system will reach and cross a threshold to a new state with altered properties, processes, and patterns.

Of concern to sustained land and environmental management are postthreshold conditions such as reduced acid buffering capacity, impaired hydrologic processes, displaced plant communities, or different crop suitabilities when compared to the prethreshold states (Groffman et al., 2006). The management actions used to restore a postthreshold state may be different and more costly than those used to keep a system from crossing a threshold (Stringham et al., 2003). Once a threshold has been crossed, the limited reversibility of the resulting conditions may restrict future management options.

38.8 The Science and Management of Soil Change: Status and Future

About half of the approximately 13 billion ha of Earth's soil are now managed for human use: cultivated for crops; managed for pastures and hayfields; logged for wood; disturbed by mining; developed for urban, suburban, transportation, industrial, and recreational projects; and used to process burgeoning streams of human and animal wastes (FAO-STAT, 2009). Important areas are contaminated by chemical compounds and large areas lie in wait of conversion for use or reuse in the coming decades. All of these soils are also being affected by changing climates and increasing concentrations of atmospheric CO_2 and other greenhouse gases. The age of pedogenesis has given way to the age of anthropedogenesis.

Sustainably managing this change so that future management and soil function are not compromised is a goal well-worth bequesting to future generations. Creating this legacy is entirely contingent upon our ability to predict and improve management of soil change on the human timescale. Land managers and policy analysts need new soil survey products that provide information about soil's resistance to stress and disturbance, soil's ability to recover readily (resilience), and any prethreshold
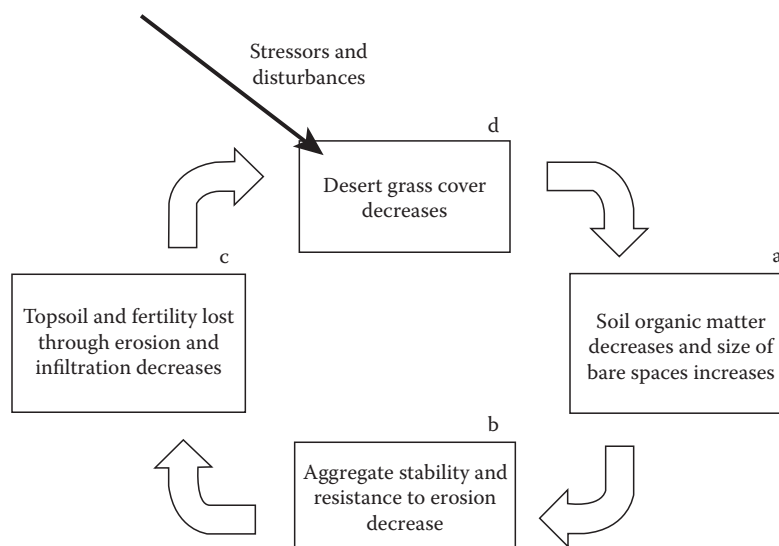

FIGURE 38.2 Effects of management on reorganization of soil property–process–pattern. Interactions between plant and soils contribute to the reorganization. Decreases in organic matter, aggregate stability, and resistance to erosion occur are associated with bare spaces that can continue to expand as degradation proceeds. (Modified from Bird, S.B., J.E. Herrick, and M.M. Wander. 2001. Exploiting heterogeneity of soil organic matter in rangelands: Benefits for carbon sequestration, p. 121–138. *In* R.J. Follett, J.M. Kimble, and R. Lal (eds.) The potential of U.S. grazing lands to sequester carbon and mitigate the greenhouse effect. CRC Press LLC, Boca Raton, FL.)

indices that warn about the loss of soil function. At present, our understanding of how soils are changing over decades' time and the potential prolonged effects of these changes on soil function can only be described as elementary (Richter and Markewitz, 2001; Tugel et al., 2005). Pedology and anthropedology must advance an interdisciplinary science of Earth's soil, that is, Earth's critical zone (National Research Council, 2001) in order to provide the foundation for long-term sustained soil and landscape management.

Pedologists in the Anthropocene (i.e., anthropedologists) use a science that is both basic and applied and has ever increasing interdisciplinary requirements to address sustainability and management challenges. Anthropedology must focus on management resilience of soils to help increase the production of goods and services without forcing systems across thresholds to undesired or difficult-to-reverse states (Bestelmeyer et al., 2009a). We suggest that improved management requires future soils research in five areas:

1. Pedogenic processes in human-altered systems
2. Soil changes and effects on functions and potentials for use
3. Mechanisms involved in soil change, degradation, and recovery
4. Prethreshold indices to predict difficult-to-reverse soil conditions
5. Restoration of highly degraded soil

The battery of approaches described in this chapter (LTSEs, STSEs, SFTSs, LTRM's, and models) can quantify soil change and improve sustainability of soil management. Many disciplines from the sciences and humanities will be needed for successful quantification and interpretation of humanity's effect on soil function.

Contemporary pedologists need to focus on the human element as a factor driving soil formation and transformation in the Anthropocene. It is humanity's transformation of Earth's soil that challenges scientists to develop a pedology with broad purview and decades' to centuries' timescales, and a pedology that supports the science, inventory, and management of the environment, ecosystems, and global change.

References

Acton, D.F., and L.J. Gregorich. 1995. Understanding soil health, p. 5–10. *In* D.F. Acton and L.J. Gregorich (eds.) The health of our soils: Toward sustainable agriculture in Canada. Centre for Land and Biological Resource Research, Research Branch, Agriculture and Agri-Food Canada, Ottawa, Canada.

Amundson, R., and H. Jenny. 1991. The place of humans in the state factor theory of ecosystems and their soils. Soil Sci. 151:99–109.

Andrews, S.S., D.L. Karlen, and C.A. Cambardella. 2004. The soil management assessment framework: A quantitative soil quality evaluation method. Soil Sci. Soc. Am. J. 68:1945–1962.

Arnold, R.W., I. Szabolcs, and V.O. Targulian. 1990. Global soil change. Report of an IIASA-ISSS-UNEP task force on the role of soil in global change. International Institute for Applied System Analysis, Laxenburg, Austria.

Ashraf, M., and M.M. Saeed. 2006. Effect of improved cultural practices on crop yield and soil salinity under relatively saline groundwater applications. Irrig. Drain. Syst. 20:111–124.

Ballantine, K., and R. Schneider. 2009. Fifty-five years of soil development in restored freshwater depressional wetlands. Ecol. Appl. 19:1467–1480.

Barrett, G.W., G.M. Van Dyne, and E.P. Odum. 1976. Stress ecology. Bioscience 26:192–194.

Bechtold, W.A., and P.L. Patterson. 2005. The enhanced forest inventory and analysis program—National sampling design and estimation procedures. General Technical Report SRS-80. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC.

Bellamy, P.A., P.J. Loveland, R.I. Bradley, R.M. Lark, and G.J.D. Kirk. 2005. Carbon losses from all soils across England and Wales, 1978–2003. Nature 437:245–248.

Bestelmeyer, B.T. 2006. Threshold concepts and their use in rangeland management and restoration: The good, the bad, and the insidious. Restor. Ecol. 14:325–329.

Bestelmeyer, B.T., K.M. Havstad, B. Damindsuren, G. Han, J.R. Brown, J.E. Herrick, C. Steele, and D.C. Peters. 2009a. Resilience theory in models of rangeland ecology and restoration: The evolution and application of a paradigm, p. 78–95. *In* J.J. Hobbs and K.N. Suding (eds.) New models for ecosystem dynamics and restoration. Island Press, Washington, DC.

Bestelmeyer, B.T., D.A. Trujillo, A.J. Tugel, and K.M. Havstad. 2006. A multi-scale classification of vegetation dynamics in arid lands: What is the right scale for models, monitoring, and restoration? J. Arid Environ. 65:296–318.

Bestelmeyer, B.T., A.J. Tugel, and G.L. Peacock et al. 2009b. State-and-transition models for heterogeneous landscapes: A strategy for development and application. Rangel. Ecol. Manag. 62:1–15.

Bidwell, O.W., and F.D. Hole. 1965. Man as a factor in soil formation. Soil Sci. 99:65–72.

Bird, S.B., J.E. Herrick, and M.M. Wander. 2001. Exploiting heterogeneity of soil organic matter in rangelands: Benefits for carbon sequestration, p. 121–138. *In* R.J. Follett, J.M. Kimble, and R. Lal (eds.) The potential of U.S. grazing lands to sequester carbon and mitigate the greenhouse effect. CRC Press LLC, Boca Raton, FL.

Blum, W.E.H. 1997. Basic concepts: Degradation, resilience, and rehabilitation, p. 1–30. *In* R. Lal, W.H. Blum, C. Valentine, and B.A. Stewart (eds.) Methods for assessment of soil degradation. Advances in soil science. CRC Press, Boca Raton, FL.

Blum, W.E.H., and A.A. Santelises. 1994. A concept of sustainability and resilience based on soil functions: The role of ISSS in promoting sustainable land use, p. 535–542. *In* D.J. Greenland and I. Szabolcs (eds.) Soil resilience and sustainable land use. CAB International, Wallingford, U.K.

Bouma, J. 1997. Soil environmental quality: A European perspective. J. Environ. Qual. 26:26–31.

Bouma, J., and M.J.D. Hack-ten Broek. 1993. Simulation model-ling as a method to study land qualities and crop productivity related to soil structure differences. Geoderma 57:51–67.

Boxell, J., and P.J. Drohan. 2009. Surface soil physical and hydro-logical characteristics in *Bromus tectorum* L. (cheatgrass) versus *Artemisia tridentata* Nutt. (big sagebrush) habitat. Geoderma 149:305–311.

Brejda, J.J., M.J. Mausbach, J.J. Goebel et al. 2001. Estimating surface soil organic carbon content at a regional scale using the national resource inventory. Soil Sci. Soc. Am. J. 65:842–848.

Brimhall, G.H., O.A. Chadwick, C.J. Lewis et al. 1991. Deformational mass transport and invasive processes in soil evolution. Science 255:695–702.

Briske, D.D., B.T. Bestelmeyer, T.K. Stringham, and P.L. Shaver. 2008. Recommendations for development of resilience-based state-and-transition models. Rangel. Ecol. Manag. 61:359–367.

Buol, S.A., R.J. Southard, R.C. Graham, and P.A. McDaniel. 2003. Soil genesis and classification. Blackwell Publications, Oxford, U.K.

Carter, M.R., S.S. Andrews, and L.E. Drinkwater. 2003. Systems approaches for improving soil quality, p. 261–281. *In* P. Schjonning, S. Elmholt, and B.T. Christensen (eds.) Managing soil quality: Challenges in modern agriculture. CAB International, Wallingford, U.K.

Chadwick, O.A., and J. Chorover. 2001. The chemistry of pedo-genic thresholds. Geoderma 100:321–353.

Chadwick, O.A., and R.C. Graham. 2000. Pedogenic processes, p. E41–E75. *In* M.E. Sumner (ed.) Handbook of soil science. CRC Press, Boca Raton, FL.

Chen, J., J.-Z. Chen, M.-Z. Tan, and Z.-T. Gong. 2002. Soil degra-dation: A global problem endangering sustainable develop-ment. J. Geogr. Sci. 12:243–252.

Crutzen, P.J. 2002. Geology of mankind. Nature 415:23.

Daily, G. (ed.) 1997. Nature's services: Societal dependence on natural ecosystems. Island Press, Washington, DC.

Darwin, C. 1882. The formation of vegetable mold, through the action of worms. D. Appleton, New York.

Debreczeni, K., and M. Körschens. 2003. Long-term field experi-ments of the world. Arch. Agron. Soil Sci. 49:465–483.

Dent, J.B., M.J. McGregor, and G. Edward-Jones. 1996. The inter-action between soil and social scientists in rural land use planning, p. 113–122. *In* R.J. Wagenet and J. Bouma (eds.) The role of soil science in interdisciplinary research. SSSA Special Publication 45. ASA and SSSA, Madison, WI.

Dokuchaev, V.V. 1883. Russian chernozem, p. 14–419. *In* Selected works of V.V. Dokuchaev. Vol. 1. Moscow, 1948. Israel Program for Scientific Translations Ltd. (for USDA-NSF), S. Monson, Jerusalem, Israel.

Doran, J.W., and T.B. Parkin. 1994. Defining and assessing soil quality, p. 3–21. *In* J.W. Doran, D.C. Coleman, D.F. Bezdicek, and B.A. Stewart (eds.) Defining soil quality for a sustain-able environment. SSSA Special Publication 35. ASA-SSSA, Madison, WI.

Dudal, R., F. Nachtergaele, and M.F. Purnell. 2002. The human factor of soil formation. *In* 17th World Congress of Soil Science, CD-ROM, Paper 93. International Union of Soil Sciences, Bangkok, Thailand.

Ehrenfeld, J.G., B. Ravit, and K. Elgersma. 2005. Feedback in the plant-soil-system. Annu. Rev. Resour. 30:75–115.

Elzinga, C.L., D.W. Salzer, and J.W. Willoughby. 1998. Measuring and monitoring plant populations. Technical Reference 1730–1. Available online at: http://www.blm.gov/nstc/library/techref.htm

Fancy, S.G., J.E. Gross, and S.L. Carter. 2009. Monitoring the con-dition of natural resources in U.S. national parks. Environ. Monit. Assess. 151:161–174.

FAO-STAT. Food and Agriculture Organization of the United Nations. 2009. FAOSTAT database. FAO, Rome, Italy. Available online at: http://faostat.fao.org/

Farina, M.P.W., P. Channon, and G.R. Thibaud. 2000a. A compar-ison of strategies for ameliorating subsoil acidity: I. Long-term growth effects. Soil Sci. Soc. Am. J. 64:646–651.

Farina, M.P.W., P. Channon, and G.R. Thibaud. 2000b. A compar-ison of strategies for ameliorating subsoil acidity. II. Long-term soil effects. Soil Sci. Soc. Am. J. 64:652–658.

Foster, D., F. Swanson, J. Aber et al. 2003. The importance of land-use legacies to ecology and conservation. Bioscience 53:77–88.

Gerasimova, M.I., N.A. Karavaeva, and V.O. Targulian. 2000. Soil degradation: Methodology and potentialities of mapping. Eurasian Soil Sci. 33:311–318.

Gleason, H.A. 1927. Further views on the succession concept. Ecology 8:299–326.

Goudie, A. 2005. The human impact on the natural environment, past, present, and future. Wiley-Blackwell Publishing, Chichester, U.K.

Greenland, D.J., and I. Szablocs. 1994. Soil resilience and sustain-able land use. CAB International, Wallingford, U.K.

Groffman, P.M., J.S. Baron, T. Blett et al. 2006. Ecological thresh-olds: The key to successful environmental management or an important concept with no practical application? Ecosystems 9:1–13.

Grossman, R.B., D.S. Harms, C.A. Seybold, and J.E. Herrick. 2001. Coupling use-dependent and use-invariant data for soil quality evaluation in the United States. J. Soil Water Conserv. 56:63–68.

Gunderson, L.H. 2000. Ecological resilience—In theory and application. Annu. Rev. Ecol. Syst. 31:425–439.

Harris, R.F., D.L. Karlen, and D.J. Mulla. 1996. A conceptual framework for assessment and management of soil qual-ity and health, p. 61–83. *In* J.W. Doran and A.J. Jones (eds.) Methods for assessing soil quality. SSSA Special Publication 49. ASA-SSSA, Madison, WI.

Hassan, R., R. Scholes, and N. Ash. 2005. Ecosystems and human well-being: Current state and trends: Findings of the condi-tion and trends working group. Island Press, Washington, DC. Available online at: http://www.millenniumassessment.org/en/Condition.aspx

Herrick, J.E., J.W. Van Zee, K.M. Havstad, L.M. Burkett, and W.G. Whitford. 2005. Monitoring manual for grassland, shrubland, and savanna ecosystems. University of Arizona Press, Tucson, AZ.

Herrick, J.E., M.A. Weltz, J.D. Reeder, G.E. Schuman, and J.R. Simanton. 1999. Rangeland soil erosion and soil quality: Role of soil resistance, resilience, and disturbance regime, p. 209–233. *In* R. Lal (ed.) Soil quality and erosion. CRC Press, Boca Raton, FL.

Herrick, J.E., and W.G. Whitford. 1999. Integrating soil processes into management: From microaggregates to macrocatchments, p. 91–95. *In* D. Eldridge and D. Freudenberger (eds.) People and rangelands: Building the future. Proc. Int. Rangeland Congr. Vol. 1. Townsville, Australia. July 19–23, 1999. International Rangeland Congress, Inc., Townsville, Australia.

Hilgard, E.W. 1860. Report on the geology and agriculture of the state of Mississippi. E. Barksdale State Printer, Jackson, MS.

Hobbs, R.J., and J.A. Harris. 2001. Restoration ecology: Repairing the earth's ecosystems in the new millennium. Restor. Ecol. 9:239–246.

Holling, C.S. 1973. Resilience and stability of ecological systems. Annu. Rev. Ecol. Syst. 4:1–23.

Holling, C.S. 1996. Engineering resilience versus ecological resilience, p. 31–43. *In* P.C. Schulze (ed.) Engineering within ecological constraints. National Academy Press, Washington, DC.

Hotchkiss, S., P.M. Vitousek, O.A. Chadwick, and J. Price. 2000. Climate cycles, geomorphological change, and the interpretation of soil and ecosystem development. Ecosystems 3:523–533.

Ibáñez, J.J., and J. Boixadera. 2002. The search for a new paradigm in pedology: A driving force for new approaches to soil classification, p. 93–110. *In* E. Micheli, F. Nachtergaele, R.J.A. Jones, and L. Montanarella (eds.) Soil classification 2001. European Research Report No. 7. Office for Official Publications of the European Communities, Luxembourg, U.K.

Jenny, H. 1941. Factors of soil formation. McGraw-Hill Book Co., New York.

Jenny, H. 1961. E.W. Hilgard and the birth of modern soil science. Collana Della Rivista Agrochimica, Pisa, Italy.

Jenny, H. 1980. The soil resource: Origin and behavior. Ecological Studies. 37. Springer-Verlag, New York.

Johnson, D.L. 2005. Reflections on the nature of soil and its biomantle. Ann. Assoc. Am. Geogr. 95:11–31.

Karlen, D.L., and D.E. Stott. 1994. A framework for evaluating physical and chemical indicators of soil quality, p. 53–72. *In* J.W. Doran, D.C. Coleman, D.F. Bezdicek, and B.A. Stewart (eds.) Defining soil quality for a sustainable environment. SSSA Special Publication 35. SSSA, Madison, WI.

Lal, R. 1994. Sustainable land use systems and soil resilience, p. 41–67. *In* D.J. Greenland and I. Szabolcs (eds.) Soil resilience and sustainable land use. CAB International, Wallingford, U.K.

Lal, R., and P.A. Sanchez (eds.). 1992. Myths and science of soils in the tropics. SSSA Special Publication 29. SSSA, Madison, WI.

Lal, R., T.M. Sobecki, T. Iivari, and J.M. Kimble. 2004. Soil degradation in the United States: Extent, severity, and trends. Lewis Publishers, CRC Press, Boca Raton, FL.

Larson, W.E., and F.J. Pierce. 1991. Conservation and enhancement of soil quality, p. 175–203. *In* Evaluation for sustainable land management in the developing world. International Board for Soil Research and Management (IBSRAM) Proceedings 12. Vol. 2. Bangkok, Thailand.

Lawrence, D., P. D'Odorico, L. Diekmann, M. DeLonge, R. Das, and J. Eaton. 2007. Ecological feedbacks following deforestation create the potential for a catastrophic ecosystem shift in tropical dry forest. Proc. Natl. Acad. Sci. U. S. A. 104:20696–20701.

Levi, M.R., J.N. Shaw, C.W. Wood, S.M. Hermann, E.A. Carter, and Y. Feng. 2010. Land management effects on near-surface soil properties of Southeastern U.S. Coastal Plain Kandiudults. Soil Sci. Soc. Am. J. 74:258–271.

Li, J.W., D.deB. Richter, A. Mendoza, and P. Heine. 2010. Land-use history effects on soil spatial heterogeneity in the Southern Piedmont USA. Geoderma 156:60–73.

Loughin, T.M. 2006. Improved experimental design and analysis for long-term experiments. Crop Sci. 46:2492–2502.

Madson, S.L., D. Markewitz, T. Hinckley, and G. Rachel. 2006. Ecosystem services scorecard (ESS): Users guide. Ver. 1.4. D.B. Warnell School of Forest Resources. The University of Georgia, Athens, GA.

McBratney, A.B., B. Minasny, S.R. Cattle, and R.W. Vervoort. 2002. From pedotransfer functions to soil inference systems. Geoderma 109:41–73.

McLauchlan, K. 2006. The nature and longevity of agricultural impacts on soil carbon and nutrients: A review. Ecosystems 9:1364–1382.

Millar, C.I., and W.B. Woolfenden. 1999. The role of climate change in interpreting historical variability. Ecol. Appl. 9:1207–1216.

Miyamoto, S., and I. Cruz. 1987. Spatial variability of soil salinity in furrow-irrigated Torrifluvents. Soil Sci. Soc. Am. J. 51:1019–1025.

National Research Council. 2001. Basic research opportunities in Earth science. National Academies Press, Washington, DC.

Oldeman, L.R., R.T.A. Hakkeling, and W.G. Sombroek. 1991. World map of the status of human-induced soil degradation: An explanatory note. 2nd revised Edn. ISRIC World Soil Information, Wageningen, the Netherlands.

Oldeman, L.R., and G.W.J. van Lynden. 1997. Revisiting the GLASOD methodology, p. 423–440. *In* R. Lal, W.H. Blum, C. Valentine, and B.A. Stewart (eds.) Methods for assessment of soil degradation. Advances in soil science. CRC Press, Boca Raton, FL.

Palm, C., P. Sanchez, S. Ahamed, and A. Awiti. 2007. Soils: A contemporary perspective. Annu. Rev. Environ. Resour. 32:99–129.

Parr, J.F., R.I. Papendick, S.B. Hornick, and R.E. Meyer. 1992. Soil quality: Attributes and relationship to alternative and sustainable agriculture. Am. J. Altern. Agric. 7:5–11.

Parsons, D.J., T.W. Swetman, and N.L. Christensen. 1999. Uses and limitations of historical variability concepts in managing ecosystems. Ecol. Appl. 9:1177–1178.

Parton, W.T., D.S. Schimel, C.V. Cole, and D.S. Ojima. 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. Soil Sci. Soc. Am. J. 51:1173–1179.

Pellant, M., P. Shaver, D.A. Pyke, and J.E. Herrick. 2005. Interpreting indicators of rangeland health. Ver. 4. Tech. Ref. 1734-6. USDI-BLM, Denver, CO. Available online at: http://usda-ars.nmsu.edu/monit_assess/PDF_files/IIRHv4.pdf

Pickett, S.T.A. 1989. Space-for-time substitution as an alternative to long-term studies, p. 110–135. *In* G.E. Likens (ed.) Long-term studies in ecology: Approaches and alternatives. Springer-Verlag, New York.

Pikul, J.L., Jr., R.C. Schwartz, J.G. Benjamin, R.L. Baumhardt, and S. Merrill. 2006. Cropping system influences on soil physical properties in the Great Plains. Renew. Agric. Food Syst. 21:15–25.

Pimm, S.L. 1984. The complexity and stability of ecosystems. Nature 307:321–326.

Pulleman, M.M., J. Bouma, E.A. van Essen, and E.W. Meijles. 2000. Soil organic matter content as a function of different land uses history. Soil Sci. Soc. Am. J. 64:689–693.

Pyke, D.A., J.E. Herrick, P. Shaver, and M. Pellant. 2002. Rangeland health attributes and indicators for qualitative assessment. J. Range Manag. 55:584–597.

Ravi, S., P. D'Odorico, T.E. Huxman, and S.L. Collins. 2010. Interactions between soil erosion processes and fires: Implications for the dynamics of fertility islands. Rangel. Ecol. Manag. 63:267–274.

Richter, D.deB. 2007. Humanity's transformation of Earth's soil: Pedology's new frontier. Soil Sci. 172:957–967.

Richter, D.D., and L.I. Babbar. 1991. Soil diversity in the tropics. Adv. Ecol. Res. 21:316–389.

Richter, D.deB., M. Hofmockel, D. Powlson, and P. Smith. 2009. LTSEs: The first global inventory. Available online at: http://ltse.env.duke.edu (verified July 2009). Duke University, Durham, NC.

Richter, D.D., and D. Markewitz. 1995. How deep is soil? Bioscience 45:600–609.

Richter, D.D., Jr., and D. Markewitz. 2001. Understanding soil change: Soil sustainability over millennia, centuries, and decades. Cambridge University Press, Cambridge, U.K.

Richter, D.deB., and M.L. Mobley. 2009 Monitoring Earth's critical zone. Science 326:1067–1068.

Richter, D.D., N.H. Oh, R. Fimmen, and J. Jackson. 2007. The rhizosphere and soil formation, p. 179–200. *In* Z. Cardon and J. Whitbeck (eds.) The rhizosphere: An ecological perspective. Elsevier, Inc., New York.

Ross, S. 1989. Soil processes: A systematic approach. Routledge, London, U.K.

Scheffer, M., S. Carpenter, J.A. Foley, C. Folke, and B. Walker. 2001. Catastrophic shifts in ecosystems. Nature 413:591–596.

Seybold, C.A., J.E. Herrick, and J.J. Brejda. 1999. Soil resilience: A fundamental component of soil quality. Soil Sci. 164:224–234.

Seybold, C.A., M.J. Mausbach, D.L. Karlen, and H.H. Rogers. 1997. Quantification of soil quality, p. 387–404. *In* R. Lal, J.M. Kimble, R.F. Follett, and B.A. Stewart (eds.) Soil processes and the carbon cycle. Advances in soil science. CRC Press, Washington, DC.

Sibbesen, E. 1986. Soil movement in long-term field experiments. Plant Soil 91:73–85.

Simonson, R.W. 1959. Outline of a generalized theory of soil genesis. Soil Sci. Soc. Am. Proc. 23:152–156.

Smith, P. et al. 2007. Climate change cannot be entirely responsible for soil carbon loss observed in England and Wales, 1978–2003. Global Change Biol. 13:2605–2609.

Sparling, G. 2006. Quality indicators, p. 1408–1411. *In* R. Lal (ed.) Encyclopedia of soil science. 2nd Edn. Taylor & Francis, New York. Available online at: http://prod.informaworld.com/10.1081/E-ESS-120042739 (accessed August 2007).

Steiner, R.A. 1995. Long-term experiments and their choice for the research study, p. 15–21. *In* V. Barnett et al. (eds.) Agricultural sustainability: Economic, environmental, and statistical considerations. John Wiley & Sons, Chichester, U.K.

Stocking, M.A. 1995. Soil erosion and land degradation, p. 223–242. *In* T. O'Riordan (ed.) Environmental science for environmental management. Pearson Educated Limited, Longman, U.K.

Stringham, T.K., W.C. Krueger, and P.L. Shaver. 2003. State-and-transition modeling: An ecological process approach. J. Range Manag. 56:106–113.

Szabolcs, I. 1994. The concept of soil resilience, p. 33–39. *In* D.J. Greenland and I. Szabolcs (eds.) Soil resilience and sustainable land use. CAB International, Wallingford, U.K.

Tenywa, M., J.G.M. Majaliwa, and A. Lufafa. 2006. Resilience, p. 1484–1487. *In* R. Lal (ed.) Encyclopedia of soil science. 2nd Edn. Taylor & Francis, New York.

Tongway, D.J., and J.A. Ludwig. 1997. The nature of landscape dysfunction in rangelands, p. 49–61. *In* J. Ludwig, D. Tongway, D. Freudenberger, J. Noble, and K. Hodgkinson (eds.) Landscape ecology, function and management: Principles from Australia's rangelands. CSIRO, Australia.

Tugel, A.J., J.E. Herrick, J.R. Brown, M.J. Mausbach, W. Puckett, and K. Hipple. 2005. Soil change, soil survey, and natural resources decision making: A blueprint for action. Soil Sci. Soc. Am. J. 69:738–747.

Tugel, A.J., S.A. Wills, and J.E. Herrick. 2008. Soil change guide: Procedures for soil survey and resource inventory, Version 1.1. USDA-NRCS, National Soil Survey Center, Lincoln, NE.

van Wesemael, B., S. Lettens, C. Roelandt, and J. Van Orshoven. 2004. Changes in soil carbon stocks from 1960 to 2000 in the main Belgian cropland areas. Biotechnol. Agron. Soc. Environ. 8:133–139.

Wadleigh, C.H., and M. Fireman. 1949. Salt distribution under furrow and basin irrigated cotton and its effect on water removal. Soil Sci. Soc. Am. J. 13:527–530.

Walker, B., C.S. Holling, S.R. Carpenter, and A. Kinzig. 2004. Resilience, adaptability and transformability in social–ecological systems. Ecol. Soc. 9:5.

Warkentin, B.P. 1995. The changing concept of soil quality. J. Soil Water Conserv. 50:226–228.

White, P.S. 1979. Pattern, process and natural disturbance in vegetation. Bot. Rev. 45:229–299.

Wilding, L.P., J. Bouma, and D.W. Goss. 1994. Impact of spatial variability on interpretive modeling, p. 61–75. *In* R.B. Bryant and R.W. Arnold (eds.) Quantitative modeling of soil forming processes. SSSA Special Publication No. 39. SSSA, Madison, WI.

Wilding, L.P., and H.S. Lin. 2006. Advancing the frontiers of soil science towards a geoscience. Geoderma 131:257–274.

Yaalon, D.H. 2000. Down to earth: Why soil—and soil science—matters. Nature 407:301.

Yaalon, D.H., and B. Yaron. 1966. Framework for man-made soil changes—An outline of metapedogenesis. Soil Sci. 102:272–278.

Young, I.M., and J.W. Crawford. 2004. Interactions and self-organization in the soil-microbe complex. Science 304:1634–1637.

39

Noninvasive Geophysical Methods Used in Soil Science

James A. Doolittle
*United States Department
of Agriculture*

39.1 Introduction

Geophysical methods measure changes in some physical property (i.e., density, seismic velocity, electrical conductivity, electrical resistivity [ER], magnetic susceptibility, dielectric permittivity) of the subsurface without direct access to the sampled volume (Daniels et al., 2003; Allred et al., 2008). Over the last three decades, the use of geophysical methods has significantly increased in soil investigations. The geophysical methods most commonly used in these pursuits include electromagnetic induction (EMI), ground-penetrating radar (GPR), transient EMI, galvanic resistivity, and magnetics. These methods are used to infer and better understand the spatial variability of soils and soil properties, and to guide observations and sampling. Selection of the most suitable geophysical method often requires an understanding of the soil properties that influence the method's response, and whether, and to what extent, a selected soil property affects the measured physical property (Allred et al., 2008). Whereas surface geophysical methods allow more continuous coverage than traditional approaches, they are limited in their capacity to resolve and characterize many pedologic features. Recent improvements in instrumentation and the integration of geophysical methods with other technologies (global positioning systems [GPS], data processing software, and surface mapping programs; e.g., geographic information systems [GIS], Geosoft, and Surfer) have fostered the expanded use of geophysical methods in soil investigations. The impetus has been the needs to improve quality control, provide more comprehensive coverage, and increase the efficiency of field operations. In agriculture, the three most commonly used geophysical methods are EMI, GPR, and Electrical resistivity (Allred et al., 2008). This chapter will focus on two of these geophysical methods: GPR and EMI. The initial uses, expansion, and present-day applications of GPR and EMI in agriculture have been summarized by Collins (2008) and Corwin (2008), respectively. As numbers of applications have become diverse and numerous, with a considerable amount of literature written on the uses of these geophysical methods in soil science, it is impossible for this brief discussion to cover all of these applications and papers.

39.2 Ground-Penetrating Radar

39.2.1 System Overview

39.2.1.1 Principles of Operation

Ground-penetrating radar is an impulse radar system designed for shallow (0–30 m) subsurface investigations. GPR systems transmit short pulses of very-high- and ultrahigh-frequency (from about 30 MHz to 1.2 GHz) electromagnetic energy into the ground from an antenna. Each pulse consists of a spectrum of frequencies distributed around the center frequency of the transmitting antenna. The transmitted pulses of energy are propagated downward into the soil. Whenever this wave of energy contacts an interface that separates layers with different permittivity (e.g., boundaries separating major soil, stratigraphic, and lithologic layers or features), a portion of this energy is reflected back to a receiving antenna. The more abrupt and contrasting the permittivity on opposing sides of an interface, the greater the amount of energy that is reflected back to the antenna. The receiving unit amplifies and samples the reflected energy and converts it into a similarly shaped waveform in a lower frequency range. The processed reflected waveforms are displayed on a video screen and can be stored on a hard disk for future playback and/or processing.

Ground-penetrating radar is a time scaled system. This system measures the time that it takes electromagnetic energy to travel from an antenna to a subsurface interface and back. To convert the travel time into a depth scale, the velocity of pulse propagation or the depth to a reflector must be known. The relationships among depth (*D*), two-way pulse travel time (*T*), and velocity of propagation (*v*) are described in the following equation (after Daniels, 2004):

$$v = 2\frac{D}{T}. \tag{39.1}$$

The velocity of propagation is principally affected by the relative permittivity (*E*$_r$) of the profiled material(s) according to the equation (after Daniels, 2004):

$$E_r = \left(\frac{C}{v}\right)^2. \tag{39.2}$$

In Equation 39.2, *C* is the velocity of light in a vacuum (0.3 m ns^{-1}). Velocity is often expressed in meters per nanosecond (m ns^{-1}).

Relative permittivity is a dimensionless constant that describes a material's capacity to store and release electromagnetic energy in the form of an electrical charge (Cassidy, 2009). Relative permittivity is expressed as (after Cassidy, 2009)

$$E_r = \frac{(E)}{(E_0)}. \tag{39.3}$$

where

E is the permittivity of the material
E$_0$ is the permittivity of free space (8.8542 × 10^{-12})

The relative permittivity varies with the amount of free and bound water (Cassidy, 2009). Relative permittivity ranges from 1 for air to 78–88 for water (Cassidy, 2009). Small increments in soil moisture can result in substantial increases in the relative permittivity of soils (Daniels, 2004). Using a 100 MHz antenna, Daniels (2004) observed that the relative permittivity of most dry mineral soil materials is between 2 and 10, while for most wet mineral soil materials, it is between 10 and 30. The relative permittivity of saturated peat deposits can range from 43 to 69 (Ulriksen, 1982). The relative dielectric permittivity of permafrost ranges from 2 to 8 (Cassidy, 2009). It should be noted that these reported values are simply approximations that were derived largely from laboratory experiments rather than real-world experiences (Cassidy, 2009). There is a tendency for permittivity to decrease with increasing frequency and conductivity to increase with increasing frequency (Annan, 2001).

39.2.1.2 Equipment

A typical GPR system consists of a radar control unit with transmitting and receiving antennas. The control unit consists of a screen, microprocessor, and mass storage medium. A microcomputer is used to control measurement processes, store data, and serve as a user interface. Modern GPR systems are lightweight and highly portable (Figure 39.1). Principal manufacturers of GPR include Geophysical Survey Systems Inc. (GSSI; Salem, New Hampshire); MALÅ (Stockholm, Sweden); and Software & Sensors (Mississauga, Ontario).

The most common mode of GPR data acquisition is the reflection profiling mode (Figure 39.1) in which radar waves are transmitted, received, and recorded as the antenna is moved along the soil surface. Transillumination methods are less commonly applied in soil investigations and require boreholes. In this method, the transmitter and receiver are placed and moved through two boreholes placed on opposite sides of the measured medium.

Based on GPR systems design and user preference, data are collected in either a continuous or step mode. The continuous mode involves the unidirectional movement of both transmitting and

(a)   (b)

FIGURE 39.1 Radar traverses being conducted in the continuous mode with a GSSI SIR-3000 system and a 200 MHz antenna (a) and a Software & Sensors Noggin system and a 250 MHz antenna (b).

receiving antennas (usually housed in one container or bound together) along a traverse or grid line (Figure 39.1). This mode is well suited to the rapid collection of data over relatively large areas. In the step mode, transmitting and receiving antennas are placed on the soil surface at a set distance apart (constant offset), a scan or multiple scans are obtained, and the antennas are then moved a short measured distance along the traverse line to repeat the process. This method is rather slow and cumbersome. However, averaging multiple scans collected at each measurement point results in a reduction of background noise.

Multi-offset methods are used with some GPR systems. These methods involve varying the distance between the transmitting and receiving antennas. Multi-offset methods are used to obtain estimates of the radar's velocity of propagation versus depth in the ground (Annan, 2009). Two commonly used multi-offset methods are common midpoint (CMP) and wide-angle reflection and refraction (WARR). In CMP, the distance between the antennas is increased stepwise, while the same midpoint position is maintained between the two antennas. In WARR acquisition, the distance between the two antennas is increased stepwise while the transmitter is maintained at a fixed location.

39.2.1.3 Antennas

Antennas are used to transmit energy, receive energy, or both. The resolution and penetration depth of GPR are determined by the antenna frequency and the electrical properties of earthen materials (Olhoeft, 1998; Daniels, 2004). A major constraint of GPR is the relationships between antenna frequency, penetration depth and resolution; resolution will decrease with decreasing antenna center frequency and with increasing depth of penetration (Buynevich et al., 2009). Higher-frequency (>400 MHz) antennas provide higher resolution, but have less penetration depth than lower frequency antennas (<300 MHz). The antennas most commonly used in soil investigations have center frequencies between 100 and 500 MHz. Lower frequency antennas (100–300 MHz) are commonly used for water table, lithologic, and stratigraphic studies, and for investigations in areas of more conductive and attenuating soil materials. Higher-frequency antennas (400–500 MHz) provide better results in relatively dry, electrically resistive soils. Antennas with frequencies of 900 MHz–1.5 GHz have been used for shallow investigations in some coarser-textured soils. For organic soils, where greater depths of penetration are often desired to profile the organic/mineral soil interface, lower frequency (70–200 MHz) antennas are commonly used.

39.2.1.4 Field Methods

For most GPR soil investigations, a traverse line or small grid is established across the study site. Generally, more information is obtained by using a network of traverse lines that capture the extent and variability of GPR reflections (Buynevich et al., 2009). Typically, reference points are located at uniform intervals along traverse or grid lines. The spacing of these reference points varies with the anticipated complexity of soils and soil properties, and can range from several centimeters to several tens of meters.

The number of reference points depends on the feature of interest and the length of the GPR traverse line.

Typically, for soil survey investigations, an appropriate antenna is moved along the ground surface in the continuous mode (Figure 39.1a). On relatively smooth surfaces that are free of debris or other hindrances, a survey wheel can be attached or the antenna placed in a cart with an odometer wheel (Figure 39.1b). Trigger mechanisms on the survey wheel or cart control the start and stop of data recording and provide reliable distance marks (Buynevich et al., 2009). At uniform distances along the traverse line, the operator or the trigger mechanism impresses a mark on the radar record indicating a reference or distance point.

After reviewing the radar record in the field, soils are cored, described, and sampled at few selected reference points along the traverse line to verify GPR depth measurements and interpretations. Based on these observations, subsurface horizons, contrasting layers, and/or features are identified on radar records. For soil survey investigations, the presence and depth to subsurface horizons or features are used to identify soils on the radar record. Later, during the post-surveying analysis of the radar records, the depths to a soil interface of interest are interpreted and manually or semiautomatically entered into a spreadsheet. Basic statistics and descriptive summaries for each traverse line are developed and presented as survey results.

39.2.1.5 GPS Option

During the 1990s, it was realized that, in order to be more useful, GPR data needed to be integrated with available digital soil data and maps. This has led to the integration of GPR with GPS. As noted by Rial et al. (2005), the accurate positioning of radar data with GPS and its importation into GIS is a goal, which will shortly become a perquisite in GPR investigations. Modules have been developed in newer GPR systems and processing software that permits the georeferencing of GPR data. These modules not only visually georeference the GPR data (GSSI, 2008), but have widened the scope of GPR surveys (Gustafsson, 2007).

With GPS, as the radar is moved across a soil polygon, its position is continuously tracked. The number of potential reference points is determined by the speed of advance and the sampling rate set on the radar unit. During post-surveying analysis of radar data, the position of each radar scan is proportionally adjusted according to the time stamp of the two nearest positions recorded with the GPS receiver.

39.2.1.6 Soil Properties That Influence the Effectiveness and Penetration Depth of GPR

Compared with other geophysical techniques, GPR provides the highest resolution of subsurface features. However, GPR is not appropriate for use on all soils (Doolittle, 1987). There are a number of factors that may increase dissipation of radar energy passing through soils including electrical conductivity, water content, soluble salts, carbonate minerals, and gypsum (Campbell, 1990; Olhoeft, 1998; Daniels, 2004). Soils having high electrical conductivity rapidly attenuate radar energy, restrict penetration depths, and severely limit the effectiveness of GPR.

Radar energy dissipation increases with increasing soil water content (SWC) (Campbell, 1990; Daniels, 2004). When an alternating electrical field is applied, water molecules align their permanent dipole moments parallel to the direction of the applied electrical field (Daniels, 2004). The displacement of bound water molecules results in the loss of energy as heat (Neal, 2004) and storage of electrical field energy. At frequencies above 500 MHz, the absorption of energy by water is the principal mechanism for radar energy loss in soils (Daniels, 2004). Even under very dry conditions, the amount of bound water is sufficient to affect radar energy loss in soils (Daniels, 2004).

Electrical conductivity and energy loss are also affected by the concentration of salts in the soil solution (Curtis, 2001). All soil solutions contain some salts, which act as electrolytes and increase ionic conduction. In semiarid and arid regions, however, accumulations of soluble and sparingly soluble salts of K, Na, Ca, and Mg in soils increase the electrical conductivity of the soil solution and the attenuation of electromagnetic energy (Doolittle and Collins, 1995). Because of their high electrical conductivity, saline (electrical conductivity >4 dS m^{-1}) and sodic (sodium absorption ratio ≥13) soils are considered unsuited to most GPR applications. In saline soils, depending on moisture contents, penetration depths typically range from a few to less than 25 cm (Daniels, 2004; Ben-Dor et al., 2009) and GPR is considered an inappropriate tool.

Mukherjee et al. (2010) observed that the dielectric permittivity and electrical conductivity of karstified limestone increased with increasing porosity. Because of increased signal attenuation in saturated carbonate terrains, GPR penetration depths are restricted (Mukherjee et al., 2010). At similar moisture and clay contents, soils with high CaCO$_3$ equivalent have been reported to reduce the depth of GPR penetration in soils (Grant and Schultz, 1994; Lebron et al., 2004).

Because of their high adsorptive capacity for water and exchangeable cations, the penetration depth of GPR is inversely related to clay content. Olhoeft (1986), using a 100 MHz antenna, observed a penetration depth of about 30 m in clay-free sands. However, the addition of only 5% (by weight) smectitic clays reduced the penetration depth by a factor of 20 (Olhoeft, 1986). Doolittle and Collins (1998) noted that, depending on antenna frequency and the specific conductance of the soil solution, GPR penetration depths range from 5 to 30 m in dry, sandy soils, but average less than 50 cm in wet, clayey soils.

Electrical conductivity and energy loss increase with increasing cation exchange capacity (CEC) of the clay fraction (Saarenketo, 1998). Cations adsorbed to the clay particles provide an alternative pathway for electrical conduction and, therefore, contribute to electromagnetic energy losses (Saarenketo, 1998). Soils with clay fractions dominated by high-CEC clays (e.g., smectite and vermiculite) are more attenuating to GPR than soils with an equivalent percentage of low-CEC clays (e.g., kaolinite, gibbsite, and halloysite). For soils with comparable clay and moisture contents, greater depths of penetration can be achieved in highly weathered soils of tropical and subtropical regions than in soils of temperate regions. Additionally, the contribution of clay and surface conduction to electrical conductivity and electromagnetic energy loss is more noticeable in soils that have low concentrations of soluble salts (Klein and Santamarina, 2003).

39.3 GPR Soil Suitability Maps

Knowledge of anticipated depths of penetration is important for the effective use of GPR. Most radar users have limited knowledge of soils and are unable to predict the likely penetration depths or the relative suitability of soils for GPR within study sites. In many soils, high rates of signal attenuation severely restrict penetration depths and limit the suitability of GPR for a large number of applications. In saline and alkali (or sodic) soils, where penetration depths are typically less than 25 cm (Daniels, 2004), GPR is unsuited for most applications. In wet clays, where penetration depths are typically less than 50–100 cm (Doolittle et al., 2003), GPR has very low potential for most applications. However, GPR is highly suited to most applications in dry sands, where penetration depths can exceed 50 m with low frequency antennas (Smith and Jol, 1995).

A map of effective ground conductivity (Fine, 1954) provided broad guidance on anticipated rates of signal attenuation, GPR penetration depths, and the relative suitability of soils to GPR applications within the United States. Because this map was prepared at a small scale (1:2,500,000) and from a limited sample population, resolution was poor and data were generalized, which resulted in many discrepancies between predicted and actual usefulness of GPR at specific sites.

Collins (1992) developed a GPR suitability rating and created GPR suitability maps based on soil classification and soil properties within the upper 2 m. Doolittle et al. (1998) used the USDA–Natural Resources Conservation Services (NRCS) State Soil Geographic (STATSGO) database to create maps that summarized the distribution of several soil properties affecting GPR suitability. Subsequently, these general soil spatial and attribute data were used to develop and later revise the GPR Soil Suitability Map of the Conterminous United States (GSSM-USA) (Figure 39.2; Doolittle et al., 2002b, 2003, 2007), which was based on a larger sample population and offered greater detail than the effective ground conductivity map (Fine, 1954).

Soil attributes used to determine the GPR suitability indices of soil polygons shown on the GSSM-USA include taxonomic criteria, clay content and mineralogy, electrical conductivity, sodium absorption ratio, and CaCO$_3$ and CaSO$_4$ contents. Each soil attribute for each horizon is rated and assigned an index value ranging from 1 to 6. Lower index values are associated with lower rates of signal attenuation and greater penetration depths. For each soil attribute, the most limiting index value within 2 m for each soil in the map unit is selected, and these limiting indices are summed for each soil (the component index value). For each soil map unit, the relative area of soils with the same index values is summed, and the GPR suitability index (GSI) for the soil map unit is the most aerially extensive index value in the map unit.

GPR suitability index and
application potential

High potential
■ 1
■ 2
Moderate potential
□ 3
Low potential
■ 4
■ 5
Unsuited
▨ 6
Water
■
No data
▨

FIGURE 39.2 **(See color insert.)** The GSSM-USA.

Soil attribute index values and relative soil suitability indices are based on observed responses from antennas with center frequencies between 100 and 200 MHz. For mineral soils, the inferred suitability indices are based on unsaturated conditions and the absence of unpredicted, contrasting soil properties within 2 m. Within any map delineation, the actual performance of GPR will depend on variations in soil properties, the type of application, and the characteristics of the subsurface target. Penetration depths and the relative suitability of soils will be less under saturated conditions, and if shallow groundwater is present, GPR penetration will be reduced and contingent on the ions present in solution.

According to GSSM-USA, only 22% of the soils in the conterminous United States are considered well suited to GPR. Areas of moderate and low potential soils for GPR applications are more extensive and occupy an estimated 33% and 36% of the acreage, respectively. Because of saline and alkaline conditions, 7% of the soils in the conterminous United States are considered unsuited to GPR. As evident from this data, a majority of the soils in the conterminous United States are not well suited for GPR soil investigations, and most successful applications of GPR in the United States has been restricted to regions that have extensive areas of soils that are well suited to GPR.

Because of the small compilation scale (1:250,000) of GSSM-USA, the minimum polygon size is about 625 ha. Larger scale GPR soil suitability maps have been prepared on a state basis

using the USDA–NRCS's Soil Survey Geographic (SSURGO) database (Doolittle et al., 2006b), which is directly derived from soil surveys completed at scales ranging from 1:12,000 to 1:63,360 (minimum delineation size ranging from about 0.6 to 16.2 ha, respectively) (Soil Survey Division Staff, 1993). The same soil properties, attribute index values, and aggregation methods used to prepare GSSM-USA were used to derive these larger scale maps.

An example of these large scale maps, the GPR Soil Suitability Map of Nebraska (GPRSSM-NE), is shown in Figure 39.3. Compared with GSSM-USA (Figure 39.2), soil information contained on GPRSSM-NE (Figure 39.3) is less generalized, soil patterns are more intricate, and polygons are shown with greater resolution. Broad spatial patterns, which correspond to major soil and physiographic units within Nebraska, are evident on both maps (Figures 39.2 and 39.3). However, GPRSSM-NE provides a more detailed overview of the spatial distribution of soil properties that influence the depth of penetration and effectiveness of GPR.

The spatial information contained on GPR soil suitability maps can aid evaluations of the relative appropriateness of using GPR, selection of the most suitable antennas and survey procedures, and the need and level of data processing. GPR soil suitability maps are available for most states and can be accessed at http://soils.usda.gov/survey/geography/maps/GPR/index.html. These maps are periodically updated as additional soil information is collected and certified.
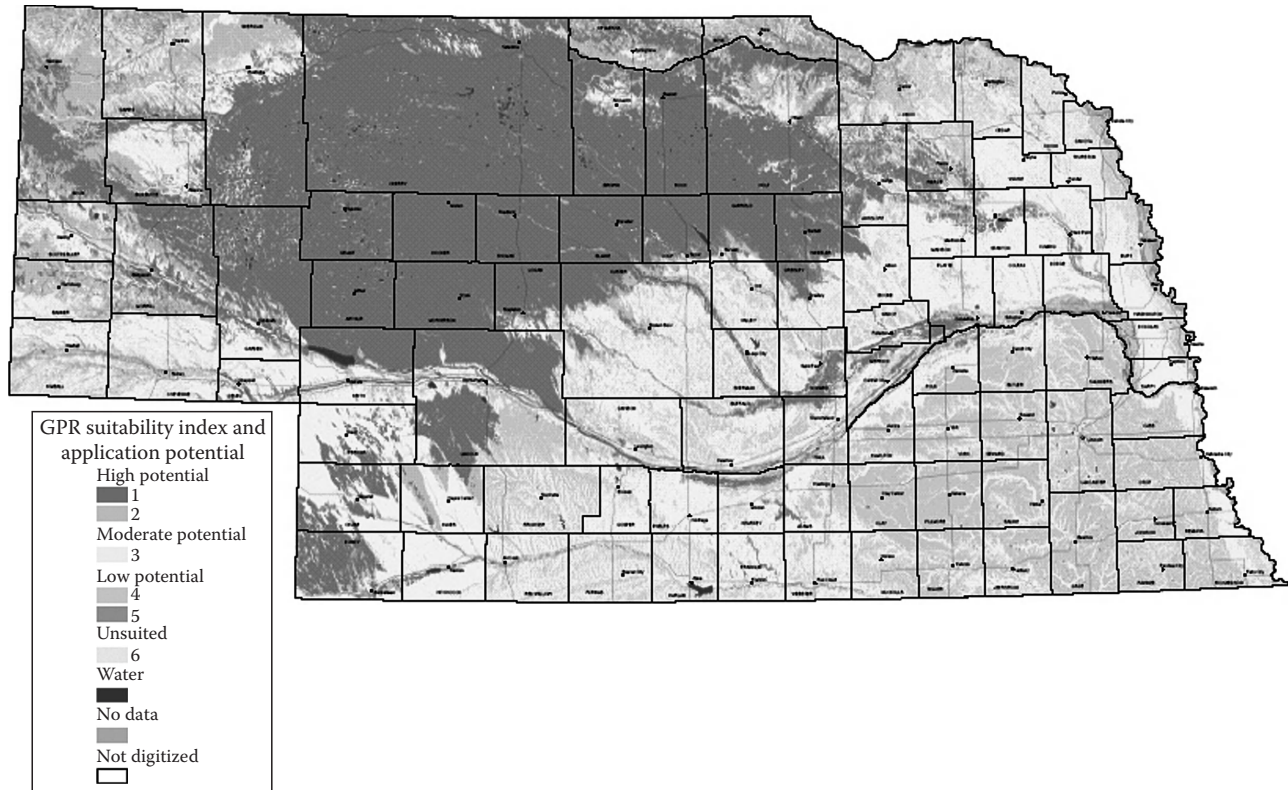
FIGURE 39.3 **(See color insert.)** The State GPRSSM-NE.

39.3.1 Factors Influencing Feature Detection

39.3.1.1 Reflection Coefficient

The amount of energy reflected back to an antenna is a function of the dielectric gradient that exists across a soil interface or boundary. The greater and more abrupt the contrast in the dielectric properties of adjoining soil materials, the greater the amount of energy reflected back to the antenna, and the more intense and conspicuous the amplitude of the reflected signal appearing on radar records. Soil horizons, layers, and features that have similar relative permittivity are poor reflectors of electromagnetic energy and are difficult to identify on radar records. The reflection coefficient, R, is a measure used to express the difference in relative permittivity that exists between two adjoining materials. The reflection coefficient is proportional to reflection strength and is expressed as (after Neal, 2004)

$$R = \frac{\sqrt{E_{r2}} - \sqrt{E_{r1}}}{\sqrt{E_{r2}} + \sqrt{E_{r1}}}, \qquad (39.4)$$

where E_{r1} and E_{r2} are the relative permittivity of adjoining materials 1 and 2. As evident in Equation 39.4, R is dependent on the difference in the relative permittivity between the two adjoining materials.

The E_r values of most dry and wet, mineral soil materials range from 2 to 10 and from 10 to 30, respectively (Daniels, 2004). The E_r of soil materials is strongly dependent upon moisture content,

and the amount of energy reflected back to the radar's antenna is greatly influenced by the abruptness and difference in moisture content between soil horizons, layers, or features.

Typically, strong radar reflections (high-amplitude reflections) are produced by abrupt soil interfaces that separate contrasting soil materials and often correspond to boundaries between soil horizons. Differences between horizons that result in high-amplitude reflections commonly are associated with differences in moisture contents, physical (grain size, texture, bulk density), and/or chemical (organic carbon, calcium carbonate, sesquioxides) properties of the horizons. GPR has been used to determine the depth to bedrock, contrasting master (E, B, C, and R) horizons, buried genetic horizons, frozen soil layers, illuvial accumulations of clays or organic matter, and cemented or indurated horizons. GPR does not detect subtle changes in soil properties (e.g., color, mottles, structure, porosity, slight changes in texture), transitional horizons (e.g., AB, AC, BC), or most vertical divisions in master horizons.

GPR does not directly measure the water table depth, but responds to near-saturated conditions within or near the top of the capillary fringe associated with a water table (Smith et al., 1992; Bentley and Trenholm, 2002). As the width of the capillary fringe above a water table increases, reflections from the water table have increasingly lower amplitudes, more dispersed characteristics and are, therefore, less distinguishable on radar records (Annan et al., 1991). In coarse-textured soils, the capillary fringe is narrow, the difference in permittivity

between the unsaturated and saturated zones is abrupt and contrasting, and the water table is often distinguishable on radar records. The more gradual change in water content over greater distance in clayey soils results in less distinct radar reflections.

39.3.1.2 Resolution

Resolution is the ability to detect two closely spaced interfaces or distinguish a subsurface feature. In order to be resolved, the thickness of a layer must be at least the same dimension as the wavelength, or the travel time through the layer must be equal to or greater than the pulse duration (Annan, 2001). Horizontal or lateral resolution depends on the velocity of propagation, the antenna's bandwidth, and the distance from the antenna. As radar waves propagate through the soil, they expand to form a conical "footprint" area. Consequently, the more distant a layer or feature is from the antenna, the wider the footprint area and the lower the horizontal resolution (Annan, 2001). Higher-frequency antennas have larger bandwidths and therefore provide greater horizontal resolution than lower frequency antennas (Annan, 2009).

The vertical resolution is dependent on the propagated wavelength (λ), which is determined by dividing the propagation velocity (v) by the antenna frequency (f) (after Daniels, 2004):

$$\lambda = \frac{v}{f}. \qquad (39.5)$$

In general, vertical resolution is considered to be about one-fourth the wavelength of the radar signal. If two features are separated in time by less than this amount, they will be indistinguishable on radar records and interpreted as one. However, as noted by Moorman and Michel (1997), the detection limit is considerably lower than the theoretical vertical resolution. Annan (2009) observed that two features must be separated in time by at least one-half of the wavelength to be identified. As can be seen in Equation 39.5, lower frequency antennas have longer wavelengths and therefore afford lower vertical resolution than higher-frequency antennas.

39.3.1.3 Scattering losses

Scattering losses are frequency dependent and become significant at high radar frequencies (Annan, 2009). Scattering is necessary for the detection of major interfaces or features. However, excessive scattering, caused by small-scale heterogeneities or scattering bodies (e.g., rock fragments, tree roots, animal burrows, and cultural features or debris) in the soil, results in unwanted background noise that impairs interpretations. Large numbers of scattering bodies confound interpretations by producing numerous, undesired, subsurface reflections, which clutter radar records and mask or obscure the signal from soil features of interest. Additionally, scattering losses may be the most significant source of signal attenuation restricting penetration depths (Annan, 2001).

39.3.2 Display and Interpretation

39.3.2.1 Two-Dimensional Display

Radar records are similar in appearance to sonar or seismic-reflection profiles. By moving an antenna along the soil surface, a large number of radar scans (or traces) are recorded at a fixed rate or intervals (time or distance) producing a 2D radar record. On 2D radar records, the horizontal scale is distance, which is based on the rate of antenna advance or the distance traveled. The vertical scale is based on the round-trip travel time of the electromagnetic wave. Radar records are displayed as either line-scan or wiggle-trace displays (Figure 39.4b and a, respectively). Wiggle traces are preferred by most engineers and geologists who are familiar with oscilloscope displays. Line-scan displays assign colors or color intensity to the amplitude range on the traces. Line-scan displays are most frequently used for GPR data because of the high data volume (number of traces) (Daniels et al., 2008).

Figure 39.4 contains two displays of the same 2D radar record, which was collected in an area of Bridgehampton soil (coarse-silty, mixed, active, mesic Typic Dystrudepts) in Rhode Island. Bridgehampton soils formed in thick, silty aeolian deposits that are underlain by stratified sands and gravels. In Figure 39.4, the horizontal scale is expressed in meters and is based on measured distances along the traverse line. The vertical scale is based on a velocity of propagation that was determined by comparing the measured depth of a feature with the two-way travel time to the subsurface interface (e.g., the contact of the silty aeolian mantle with the underlying coarse-textured outwash) and Equation 39.1. In the renditions shown in Figure 39.4, the aeolian mantle is virtually free of reflectors. The uppermost, high-amplitude subsurface reflector represents the aeolian mantle/outwash interface. The underlying outwash is characterized by linear reflectors of varying amplitudes. In the two renditions of this 2D radar record, the aeolian mantle/outwash interface is punctuated by an incised feature (between the 6 and 8 m distance marks) whose form and expression suggests an ice-wedge cast.

39.3.2.2 Three-Dimensional Display

The effective visualization of radar data is the key to modern GPR interpretations. An emerging approach in GPR is the analysis of subsurface structures and geometries from a 3D perspective. Three-dimensional GPR allows the visualization of data volumes from different perspectives and cross sections (Beres et al., 1999), which can assist identification, characterize subsurface soil, stratigraphic and lithologic structures and geometries, and improve interpretations of subsurface features. In areas of electrically resistive materials, 3D GPR can provide unrivaled resolution and detail of subsurface features as compared with 2D GPR (Grasmueck and Green, 1996). The use of 3D GPR has allowed improved definition of subsurface features and resulted in more complete and less ambiguous interpretations than traditional 2D GPR (Beres et al., 1999).

The acquisition of data for 3D GPR requires greater expenditures of time and labor than 2D GPR. However, the additional expenditures of resources needed to collect, process, and visualize
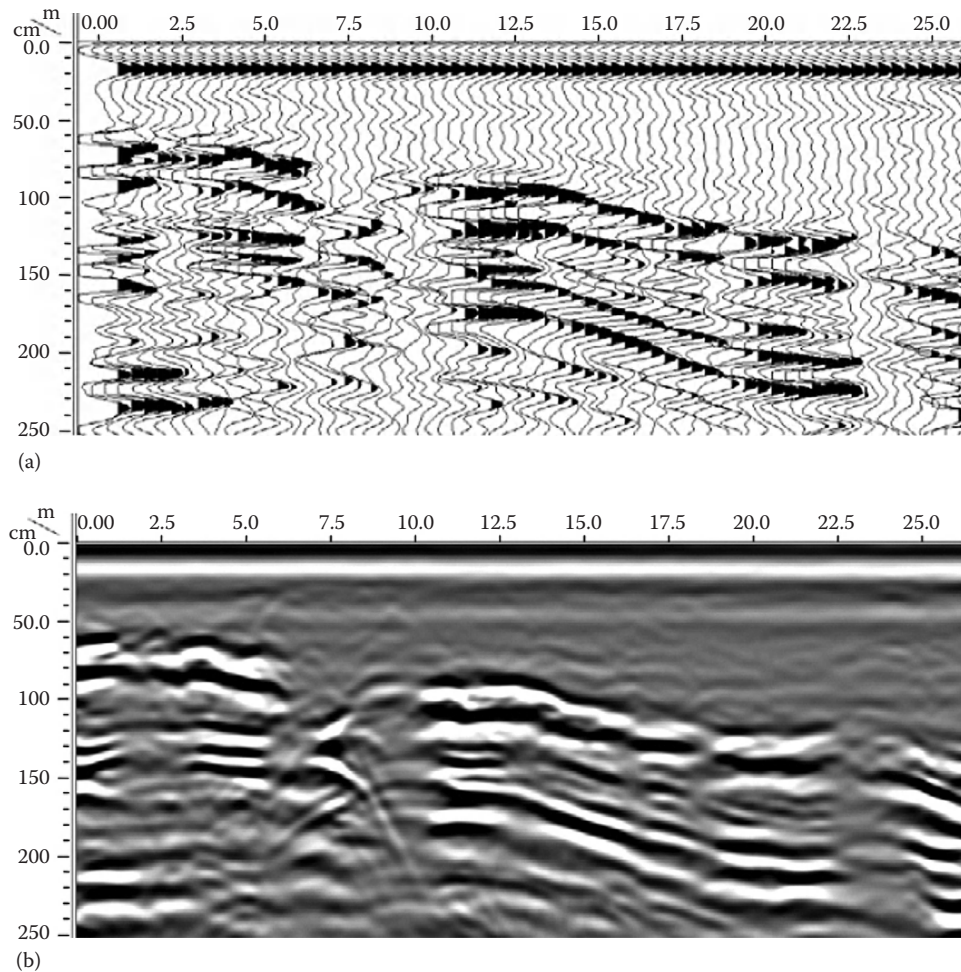
(a)

(b)

FIGURE 39.4 Two examples of the same 2D radar record, which was collected with a 200 MHz antenna in an area of Bridgehampton soils. The radar record is displayed as a (a) wiggle-trace and (b) a line-scan display.

3D GPR data are often compensated by the more comprehensive spatial coverage and greater resolution of subsurface features (Grasmueck and Green, 1996). Important considerations for collecting and interpreting 3D GPR surveys are discussed by Lehmann and Green (1999). The basic process to construct a 3D GPR pseudo-images of the subsurface is that a relatively small area (typically, 1–2500 m²) is intensively surveyed with multiple, closely spaced (typically, 0.1–1.0 m), parallel GPR traverse lines. This relatively dense set of grid lines is necessary to resolve the geometries and sizes of different subsurface features and to minimize spatially aliasing the data (Grasmueck and Green, 1996). Appropriate computer software allows the 3D GPR pseudo-image to be viewed from nearly any perspective (Junck and Jol, 2000) and arbitrary cross sections, insets, and time slices can be extracted from the 3D data set. In addition, animated imaging capability allows users to interactively examine the entire data volume (Grasmueck, 1996).

Three-dimensional GPR has been used to identify the presence and map the geometries of subsurface features in both consolidated and unconsolidated materials. Recently, 3D GPR was used to investigate the subsurface configuration of ice-wedge polygons near Barrow, Alaska (Munroe et al., 2007), and sediment filled wedges and buried polygonal ground in mid-latitude United States (Doolittle and Nelson, 2009).

Figure 39.5 contains a 2D radar record and two 3D GPR pseudo-images from a small (30 × 30 m) grid site located in an area of Delton soils (loamy, mixed, active, mesic Arenic Hapludalfs) in Wisconsin. Delton soils formed in a thin (50–100 cm) mantle of aeolian sands overlying clayey lacustrine deposits. On the 2D radar record (Figure 39.5a), the topography of the contact between the aeolian sands and underlying lacustrine deposits (highlighted with a segmented, line) appears wavy with two conspicuous concavities. On the two 3D GPR pseudo-images, a solid cube (Figure 39.5b) and a cube with a 23 × 29 m inset removed to a depth of 100 cm (Figure 39.5c) are shown. Along the sidewalls of the solid 3D GPR pseudo-image, the interface separating the sandy aeolian mantle from the underlying clayey lacustrine deposits is clearly expressed. The 3D GPR pseudo-image with the inset cube removed shows a series of linear and intersecting, high-amplitude (colored white) reflections on the base of the cutout cube. The geometry of these reflectors suggests buried polygonal ground created in a former periglacial environment. Closed segments in the lineations resulted from finer
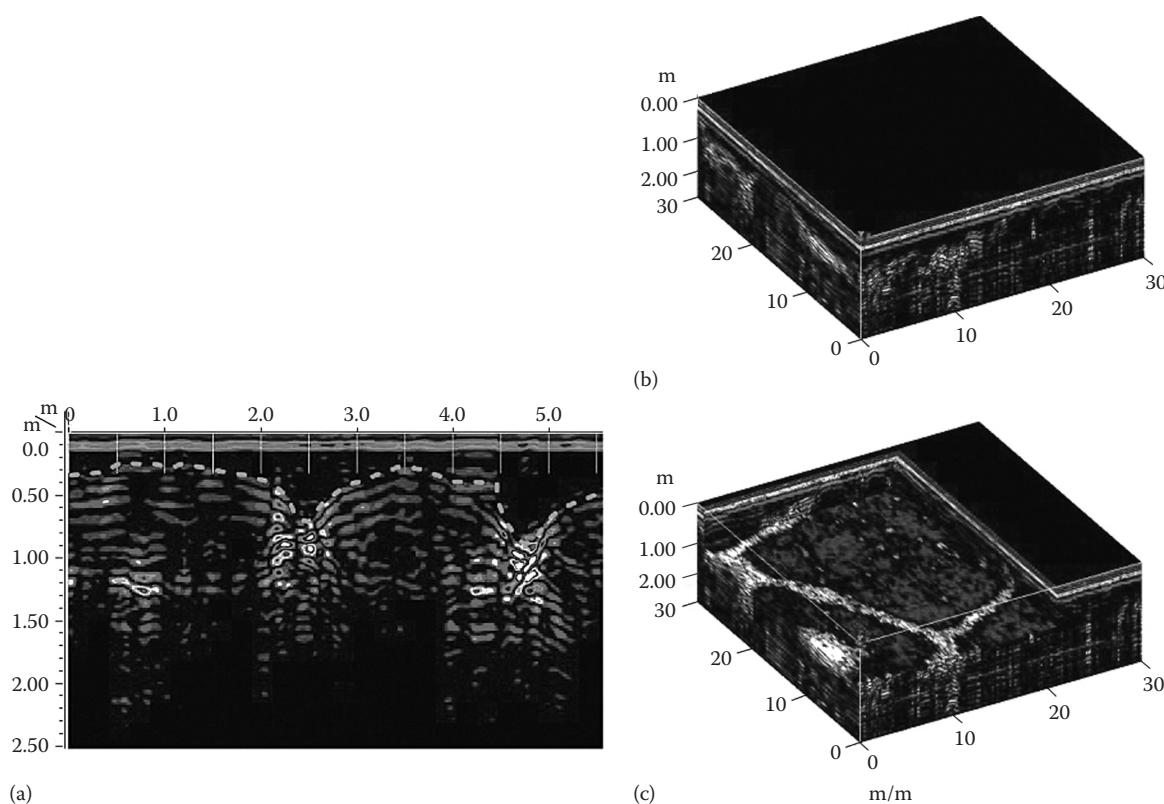
FIGURE 39.5 A 2D radar record and 3D GPR pseudo-images from a 30 × 30 m grid site in Wisconsin. (a) The segmented line in the 2D radar record shows the contact separating a sand mantle from the underlying clayey lacustrine deposits. (b and c). The 3D pseudo-images show the geometry of the ice-wedge pseudomorphs. (From Doolittle, J.A., and F.E. Nelson. 2009. Characterizing relict cryogenic macrostructures in mid-latitude areas of the USA with three-dimensional ground-penetrating radar. Permafrost Periglac. Process. 20:257–268.)

textured, more radar opaque material filling the ice-wedge voids. These closed segments were interpreted to have formed as portions of the finer-textured, ice-rich lacustrine deposits, which were susceptible to consolidation and deformation, closed during thawing and prevented infilling with superjacent sands.

39.3.3 Applications

39.3.3.1 Use of GPR in Soil Survey

Ground-penetrating radar provides data on the presence, depth, lateral extent, and variability of many diagnostic subsurface horizons that are used to classify soils (Collins et al., 1986; Doolittle, 1987; Schellentrager et al., 1988; Puckett et al., 1990). In the United States, GPR is principally used as a quality control tool to verify the taxonomic composition of soil map units, document the presence and depth to soil horizons and features, and assess spatial and temporal variations in soil properties.

GPR was first used in Florida in 1979 to identify and determine depths to diagnostic subsurface horizons used to classify and map soils (Benson and Glaccum, 1979; Johnson et al., 1979; Collins, 2008). Since that time, GPR has been used extensively for soil survey quality control in Florida because of the ubiquity of sandy soils with favorable physiochemical properties for electromagnetic wave propagation and the presence of contrasting soil horizons (Schellentrager et al., 1988).

Johnson et al. (1979) working in sandy soils with well-expressed horizons, observed that radar interpreted depths of selected horizon boundaries were within ±2.5–5.0 cm of the measured depths. Asmussen et al. (1986) observed an average difference of 19.2 cm between the radar interpreted and measured depths to argillic (Bt) horizons, which ranged in depth from about 20 to 450 cm. Rebertus et al. (1989) observed that the difference between the interpreted and measured depths to a discontinuity, which ranged in depth from 0 to about 230 cm, was less than 15 cm in 94% of the observations. Collins et al. (1989) observed an average difference of 6 cm between the interpreted and measured depths to bedrock, which ranged in depth from about 80 to 240 cm. Simeoni et al. (2009) reported an accuracy of ±10 cm between the interpreted and measured depths to B horizon, which ranged in depth from about 25 to 100 cm. For organic soils, Rosa et al. (2008) reported a mean maximum difference of 32 cm between measured and GPR interpreted depths of peat, which ranged in thickness from 0 to 8 m. Differences were attributed to surface and subsurface irregularities, and spatial variations in peat moisture contents and bulk densities (Rosa et al., 2008).

In a GPR evaluation of Fuquay soils (loamy, kaolinitic, thermic Arenic Plinthic Kandiudults) in northern Florida, the sandy mantle overlying loamy marine deposits is virtually free of reflectors (Figure 39.6). The underlying, loamy sediments, though relatively transparent to GPR, contain a large number
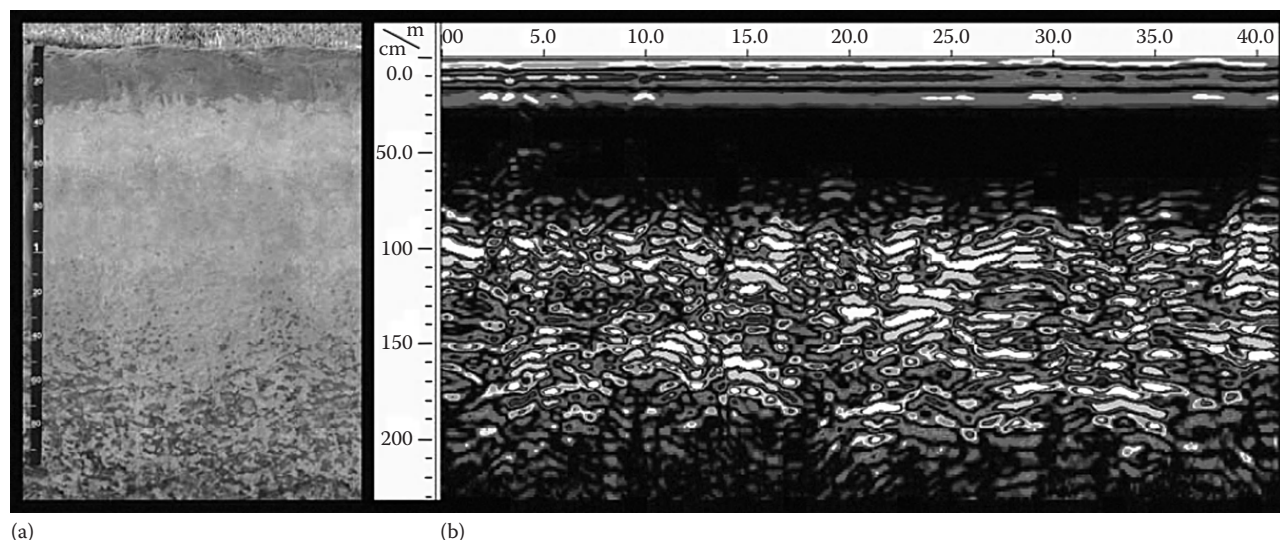
(a)   (b)

FIGURE 39.6 Radar record (b) showing contact of sandy deposits with the underlying, more heterogeneous, loamy marine sediments from an area of Fuquay soils (a) in Florida.

of reflectors, which aid identification of the contact between the two materials. The large number of reflectors of varying sizes, shapes, orientations, and reflected signal amplitudes suggests heterogeneous soil materials, which may be alternating layers of loamy and sandy materials with pockets of ironstone nodules that are evident in the lower part of the Fuquay photograph (Figure 39.6a). As over 1 m of penetration was achieved with the 200 MHz antenna through these loamy materials, the clay fraction is assumed to be dominated by low-activity clay minerals (e.g., kaolinite, gibbsite, and goethite) as implied in the taxonomic classifications of Fuquay soils (Kandiudults).

Ground-penetrating radar has been used to estimate the depth to argillic horizons (Asmussen et al., 1986; Collins and Doolittle, 1987; Doolittle, 1987; Truman et al., 1988; Doolittle and Asmussen, 1992), spodic horizons (Collins and Doolittle, 1987; Doolittle, 1987; Burgoa et al., 1991), placic horizons (Lapen et al., 1996), and buried palaeosols (Chapman et al., 2009). These diagnostic horizons generally have well-defined upper boundaries that have abrupt increases in bulk density, clay (argillic horizon), organic C complexed with Fe and Al (spodic horizon), or cemented Fe, Mn, or Fe-humus complexes (placic horizon). GPR has also been used to determine the thickness of albic horizons (E horizons) and evaluate the depth, lateral extent, and continuity of duripans, petrocalcic, and petroferric horizons (Doolittle et al., 2005), fragipans (Olson and Doolittle, 1985; Lyons et al., 1988; Doolittle et al., 2000), ortstein layers (Mokma et al., 1990a), and traffic pans (Raper et al., 1990). In these reports, the diagnostic horizons evaluated have distinct properties that enhanced their detection with GPR including cementation and induration (duripan, ortstein, petrocalcic, and petroferric horizons) and/or high bulk density (fragipans and traffic pans).

Ground-penetrating radar has been used to infer marked changes in soil color associated with abrupt and contrasting changes in organic carbon contents (Collins and Doolittle, 1987). GPR has also been used to infer the concentration of clay lamellae (Farrish et al., 1990; Mokma et al., 1990b; Tomer et al., 1996) and plinthite (Doolittle et al., 2005) in soils. Steelman and Endres (2009) used GPR to monitor soil freezing and the seasonal development of surficial, seasonally frozen soil layers. In areas of permafrost, GPR has been used to estimate the thickness of the active layer (Doolittle et al., 1990b) and to identify ice wedges (Hinkel et al., 2001).

In evaluations of a landscape in Pennsylvania containing Erie (fine-loamy, mixed, active, mesic Aeric Fragiaquepts) soils that have a fragipan and Fremont (fine-loamy, mixed, semiactive, acid, mesic Aeric Endoaquepts) soils that lack a fragipan, the radar record (collected with a 400 MHz antenna) has a seemingly continuous interface that grades from well to very poorly expressed (identified by a continuous line) at depths ranging from about 25 to 53 cm below the soil surface (Figure 39.7). Ground truth observations indicated that in areas where this interface had higher-amplitude reflections (lighter colors), the soil had a fragipan (Erie). In contrast, where the interface was denoted by lower amplitude reflections (black or darker colors) the soil had fragic properties but lacked a true fragipan and was considered to be Fremont. Multiple radar traverses across the landscape indicated that Fremont was the more abundant soil in the area.

Radar records of deep, sandy, excessively drained Valentine soils (mixed, mesic Typic Ustipsamments) in the Nebraska Sand Hills indicates thin, distinct, discontinuous, inclined to wavy bands throughout the soil (Figure 39.8b). These bands correspond to thin (<7.5 cm thick) layers of oriented clay on or bridging grains (lamellae) that are common in these sandy soils (Soil Survey Division Staff, 1999; Figure 39.8a). The radar record in Figure 39.8b has been surface normalized to correct the radar record for changes in elevation to produce graphic displays that
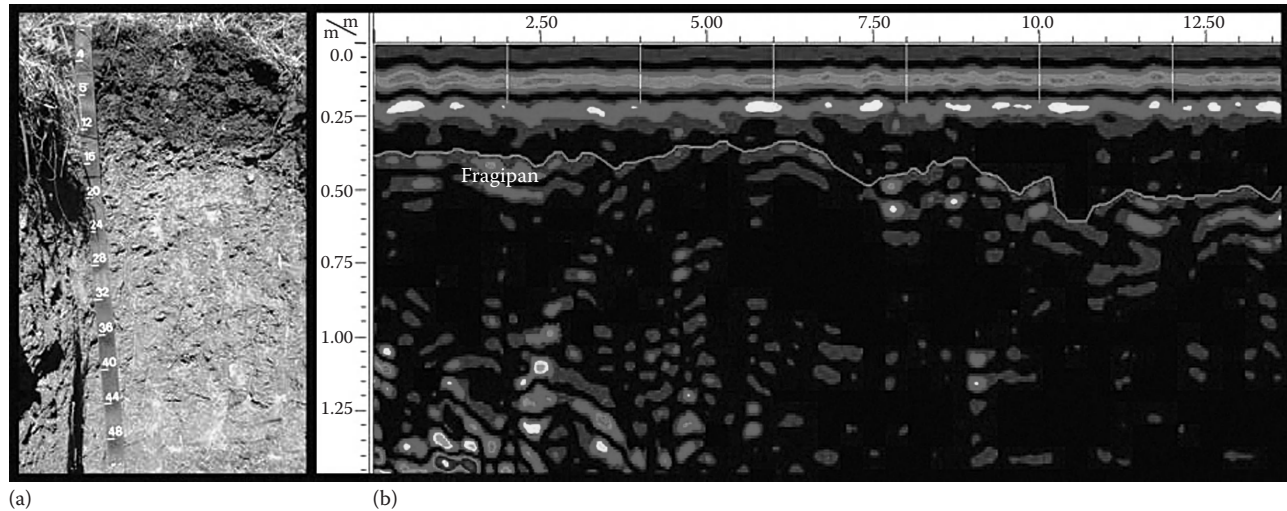
FIGURE 39.7   Radar record (b) showing a continuous subsurface interface with intermittent fragic properties in an area of Erie soils (a) in western Pennsylvania.
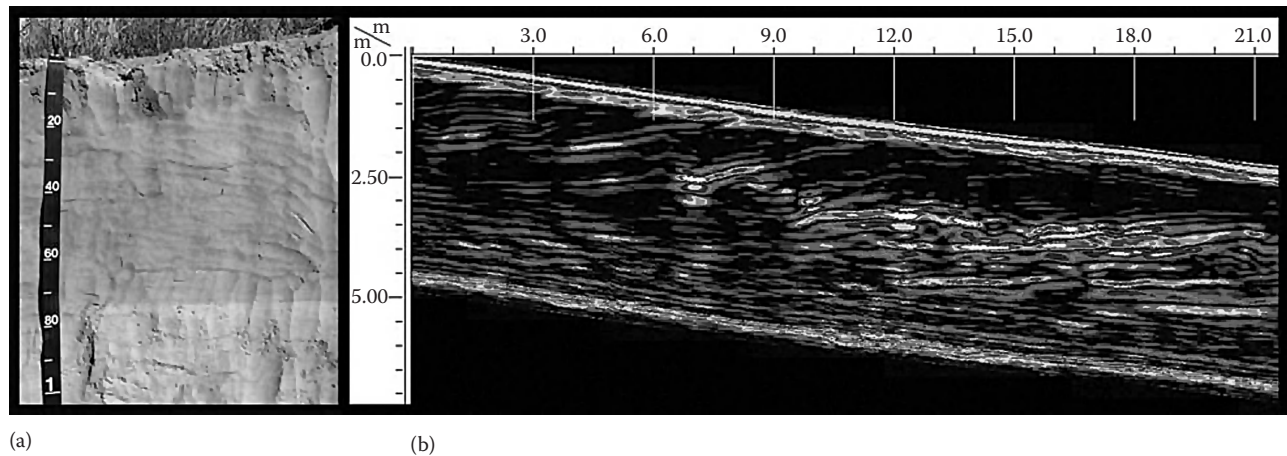

FIGURE 39.8   Surface normalized radar record (b) from an area of Valentine soils (a) in western Nebraska with discontinuous, inclined-to-wavy, thin bands of lamellae.

more closely resemble the topography of the landform and the soil horizons or features.

In areas of Valentine soils, the penetration depth of GPR typically ranges from about 2 to 5 m, which is less than expected for a very deep, excessively drained sandy soil. The presence of small amounts of clay in the lamellae, however, increases signal attenuation and reduces penetration depths (Harari, 1996; Olhoeft, 1998). In some soils, clay contents of only 5%–10% have been reported to reduce penetration depths to less than 1 m (Walther et al., 1986). Additionally, soils, such as Valentine, with an abundance of smectite or vermiculite, and associated relatively high CEC are more attenuating to GPR than soils with low-activity clay minerals.

In many upland areas, rock fragments and irregular or weathered bedrock surfaces limit the effectiveness of conventional methods to examine soil profiles and determine the depth to bedrock. In these areas, GPR is more reliable and effective than traditional soil surveying tools for determining the depth to bedrock and the composition of soil map units based on soil-depth criteria (Collins et al., 1989; Schellentrager and Doolittle, 1991). As seen in radar record shown in Figure 39.9b, the soil/bedrock interface can provide an abrupt, well-expressed, and easily identifiable reflector.

In some soils, however, coarse fragments in the overlying soil, irregular bedrock surfaces, fracturing, and the presence of saprolite make the identification of the soil/bedrock interface on radar records more ambiguous. Even with these limitations, GPR can be an effective tool for evaluating bedrock depths (Collins et al., 1989; Davis and Annan, 1989; Gerber et al., 2010) and changes in rock type (Davis and Annan, 1989), characterizing internal bedding, cleavage and fracture planes (Holloway and Mugford, 1990; Stevens et al., 1995; Toshioka et al., 1995;
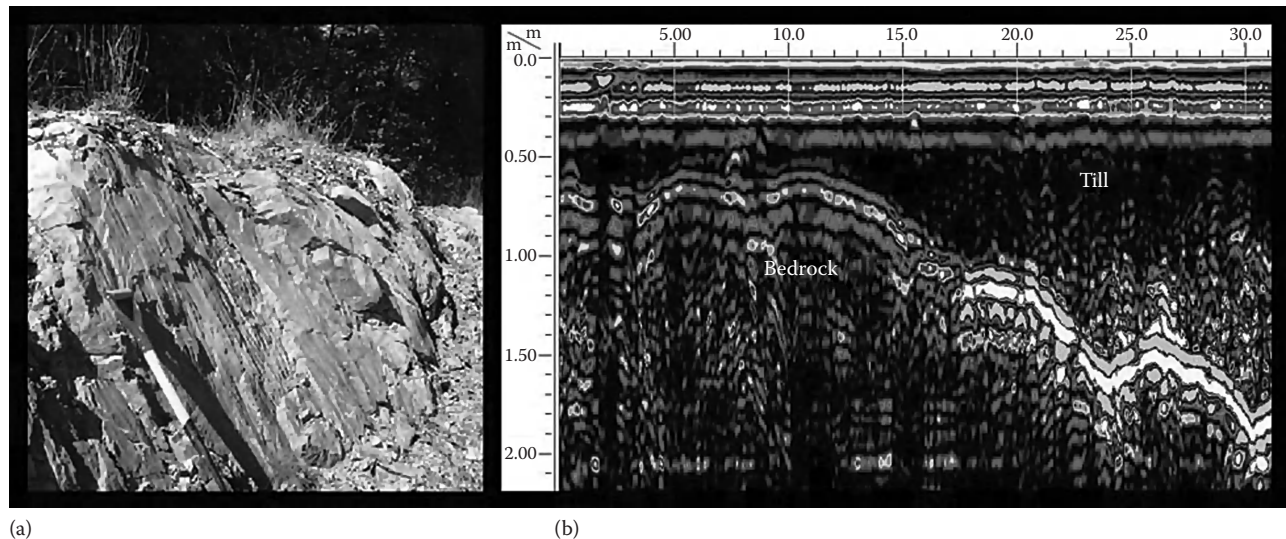
FIGURE 39.9 Radar record (b) collected with a 400 MHz antenna showing soil/bedrock interface in area of Monson soils (a) in central Maine.

Lane et al., 2000; Grasmueck et al., 2004; Nascimento da Silva et al., 2004; Porsani et al., 2005), and identifying cavities, sinkholes, and fractures in limestone (Barr, 1993; Pipan et al., 2000; Al-fares et al., 2002).

One example of the use of GPR over bedrock is a radar record from an area of Monson soils (loamy, isotic, frigid Lithic Haplorthods) in central Maine (Figure 39.9). This shallow, somewhat excessively drained soil formed in a thin mantle of glacial till over slate, phyllite, or schist. The phyllite shown in the photograph contains a large number of narrow, closely spaced, vertical fractures, which cannot be effectively resolved with the 200 MHz antenna used to collect the 2D radar record. Because of scattering losses, attenuation, wavelength scale heterogeneities, and geometric constraints, the number of fractures interpreted from radar data is an order of magnitude less than the number observed in outcrops (Lane et al., 2000). In addition, fractures with large dip angles reflect very little radar energy back to the antenna and are not accurately imaged because of spatial aliasing distortion (Lane et al., 2000). Fractures filled with water or saturated materials produce higher-amplitude reflections than air-filled or unsaturated fractures (Lane et al., 2000).

Ground-penetrating radar also has been used to improve soil-landscape models and soil map unit design on glacial-scoured uplands (Doolittle et al., 1988), former periglacial environments (Doolittle and Nelson, 2009; Gerber et al., 2010), wetland catenas (Lapen et al., 1996), and coastal plain sediments (Rebertus et al., 1989; Puckett et al., 1990). The use of GPR to improve soil-landscape models, however, has been limited.

Ground-penetrating radar has been extensively used in the investigation of peatlands. Compared with traditional coring methods, GPR provides a more rapid and efficient means for estimating the thickness and characterizing the subsurface topography of organic deposits (Ulriksen, 1980, 1982). GPR has been used to estimate the thickness and volume of organic deposits (Ulriksen, 1982; Shih and Doolittle, 1984; Tolonen et al., 1984; Collins et al., 1986; Worsfold et al., 1986; Welsby, 1988; Doolittle et al., 1990a; Pelletier et al., 1991; Hanninen, 1992; Turenne et al., 2006), to distinguish layers having differences in degree of humification and volumetric water content (Ulriksen, 1982; Tolonen et al., 1984; Worsfold et al., 1986; Chernetsov et al., 1988; Theimer et al., 1994; Lapen et al., 1996), and to classify organic soils (Collins et al., 1986). Lowe (1985) used GPR to estimate the number of logs and stumps buried in peatlands. More recently, GPR has been used in peatlands to study subsurface piping (Holden et al., 2002), permafrost (Moorman et al., 2003), and subsurface deposits and hydrostratigraphy (Comas et al., 2004, 2005; Kettridge et al., 2008; Lowry et al., 2009). Peatlands often display considerable anisotropy in composition, moisture content, and bulk density (Warner et al., 1990), and such differences have allowed separation of organic layers that differ in degree of humification, bulk density, and dielectric permittivity (Tolonen et al., 1982; Chernetsov et al., 1988; Hanninen, 1992; Nobes and Warner, 1992; Theimer et al., 1994; Mellett, 1995; Comas et al., 2005; Lowry et al., 2009). The successful identification of interfaces resulting from differences in water content and degree of humification, however, has not been universal (Wastiaux et al., 2000; Sass et al., 2010).

Figure 39.10 is a radar record obtained with a 70 MHz antenna in a ponded area of Carlisle (euic, mesic Typic Haplosaprists) soils in northwestern Rhode Island. On this radar record, the upper 80–90 cm is plagued by parallel bands of background noise caused by reflected signals from the air/ice and ice/water interfaces. Changes in reflective patterns and signal amplitudes are evident at a depth of about 90 cm, which represents the water/organic matter interface. On this radar record, the organic/mineral soil interface is easily identified (a segmented
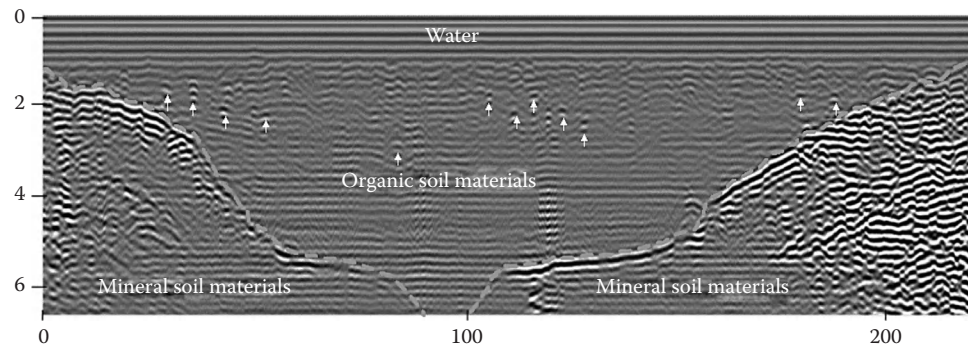
FIGURE 39.10 Radar record collected with a 70 MHz antenna over an inundated peatland in Rhode Island. A segmented line is used to highlight the organic/mineral soil interface. Arrows indicate locations of some buried logs, larger roots, or stumps. Scales are in meter.

line denotes the approximate location of this interface). This interface can be traced with reasonable confidence to a depth of about 675 cm (the maximum depth scanned). Buried logs and stumps produce hyperbolic reflections (indicted by arrows) within the organic materials.

Profiling depths of greater than 8 m have been reported in peatlands with low electrical conductivity (Ulriksen, 1980; Lowe, 1985; Worsfold et al., 1986; Theimer et al., 1994; Comas et al., 2005; Lowry et al., 2009), but GPR does not provide similar penetration depths in all organic soils. Penetration depths are largely affected by the specific conductance of the pore water (Theimer et al., 1994; Comas et al., 2005; Lowry et al., 2009). GPR is ineffective in coastal peatlands that are tidally influenced, contain sulfidic materials, and/or have high electrical conductivities. In noncoastal areas, GPR is generally more effective in acidic than alkaline peatlands. Lower frequency (<200 MHz) antennas are typically more effective for evaluating peatlands. In higher latitudes, peatlands are often surveyed during winter months, when surface layers are frozen and snow covered and can be easily traversed on foot or with snowmobiles or tracked vehicles. In lower latitudes, grass and reed covered peatlands have been successfully surveyed with airboats. Pelletier et al. (1991) described the use of helicopters to survey extensive peatlands in remote areas of Ontario.

39.3.3.2 GPR Evaluations of Groundwater, Soil Moisture, and Preferential Flow

Ground-penetrating radar has been used extensively for hydropedological investigations. In areas of coarse-textured soils, GPR has been used to chart water table depths among monitoring wells and into nearby areas (Sellmann et al., 1983; Davis et al., 1984; Wright et al., 1984; Shih et al., 1986; Johnson, 1987; Truman et al., 1988; Bohling et al., 1989; Smith et al., 1992; Iivari and Doolittle, 1994; Doolittle et al., 2006a). In addition, GPR has been used to design hydrologic models (Violette, 1987; Taylor and Baker, 1988), define recharge and discharge areas (Johnson, 1987; Bohling et al., 1989), predict groundwater flow patterns (Steenhuis et al., 1990; Iivari and Doolittle, 1994; van Overmeeren, 2004; Doolittle et al., 2006a), and delineate near-surface hydrologic conditions (Beres and Haeni, 1991; van Overmeeren, 1998). Collins et al. (1994) used GPR to identify and size subsurface cavities that influence preferential flow in a karst landscape; this information was used to improve contaminant transport models.

Radar records from an area of sandy Oakville (mixed, mesic Typic Udipsamments) soils formed in aeolian dunes in northwestern Indiana indicated depth to the water table ranging from 1.1 to 6.4 m (Figure 39.11; surface elevation has been normalized). Bedding planes within the dune and contrasting strata can
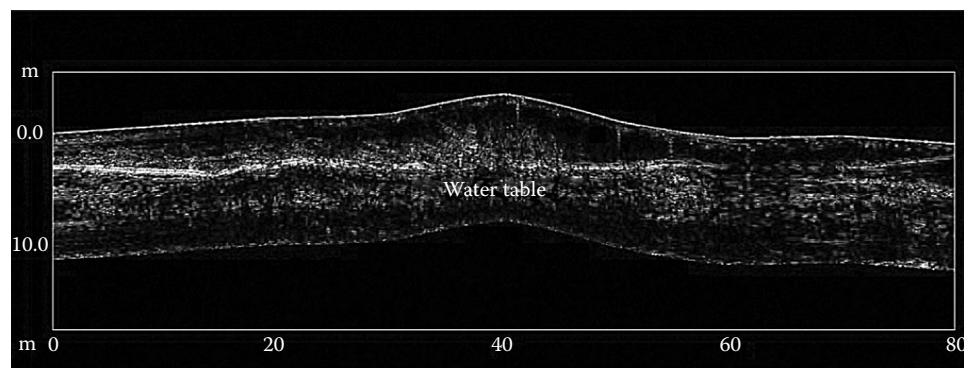

FIGURE 39.11 Surface-normalized radar record obtained with a 200 MHz antenna across a low aeolian dune in northwestern Indiana showing a water table as high-amplitude linear reflections, which stretch across the image.

be observed on the radar record. Reflections from the interior of dunes are largely produced by differences in moisture contents and differences in density and grain size between strata (Schenk et al., 1993; Harari, 1996).

At intermediate (field or catchment) scales, GPR has been used to map spatiotemporal variations in SWC (Lesmes et al., 1999; Huisman et al., 2001, 2002, 2003; Hubbard et al., 2002; Galagedara et al., 2003; Grote et al., 2003). A GPR method known as ground wave analysis has been used to measure the SWC of surface layers (Huisman et al., 2001). The ground wave is the portion of the energy that travels directly from a transmitting to a receiving antenna through the upper part of the soil. The velocity of the ground wave is strongly dependent on the water content of the upper few centimeters of the soil (Huisman et al., 2002). The volume of soil evaluated in ground wave analysis is determined by the antenna separation, width of the antenna, and depth of ground wave influence (Huisman et al., 2002). The depth of ground wave influence is typically 10–50 cm but is not well defined and decreases with increasing SWC (Huisman et al., 2002).

Ground-penetrating radar has been used to study preferential pathways in soils (Vellidis et al., 1990; Kung and Donohue, 1991; Kung and Lu, 1993; Boll et al., 1996; Tomer et al., 1996; Steenhuis et al., 1998). These studies have been conducted on sandy soils and under conditions favorable to GPR evaluations. However, Gish et al. (2002) used GPR data to locate and characterize soil layers that control subsurface flow at field scales in medium- and fine-textured soils. Using high-frequency antennas (450 or 500 MHz), these workers used GPR to identify depth to restrictive soil layers and the locations of discontinuities or breaks in these layers.

39.4 Electromagnetic Induction

39.4.1 Overview of Electromagnetic Induction

39.4.1.1 Principles of Operation

Electromagnetic induction methods involve the use of instruments known as ground conductivity meters (GCM), which consist of a transmitter coil and either one or two receiver coil(s) spaced at a set distance(s) apart. Ground conductivity meters operate by generating an alternating electrical current, which is passed through the transmitter coil. This alternating electrical current generates a time-varying electromagnetic field above the surface; the primary electromagnetic field. The primary electromagnetic field induces eddy currents that flow through the soil and generate a secondary electromagnetic field, which propagates through the ground. The secondary field is proportional to the ground current and is used to calculate the soil's "apparent" or "bulk" electrical conductivity. The amplitude and phase of these two electromagnetic fields are measured by the receiver coil(s).

Apparent conductivity (EC_a) is a weighted, average measurement for a column of earthen materials to a specific depth (Greenhouse and Slaine, 1983) and is a measure of the soil's ability to conduct electrical current. Variations in EC_a result from changes in the electrical conductivity of earthen materials, which is influenced by the type and concentration of ions in solution, the amount and type of clay in the soil matrix, the volumetric water content, and the temperature and phase of the soil water (McNeill, 1980). The EC_a of soils will increase with increases in soluble salts, water content, and/or clay contents (Rhoades et al., 1976; Kachanoski et al., 1988).

Even though interpretations are complicated by the interaction of multiple soil properties, EC_a has been used to infer and map variations in soils and soil properties at different spatial scales and levels of resolution. Variations in EC_a have been associated with changes in particle-size distribution, clay mineralogy, bulk density, CEC, salinity, plant nutrients, organic C concentration, and water content. Presently, EC_a mapping is recognized as one of the most valuable methods in agriculture for measuring the spatial variability of soils and soil properties at field and landscape scales (Corwin, 2008; Lück et al., 2009).

Electromagnetic induction is not suitable for use in all soil investigations. Generally, EMI has been most successful in areas where subsurface properties are reasonably homogeneous, the effects of one property (e.g., clay, water, or salt content) dominates over other properties, and variations in EC_a can be directly associated with changes in the dominant property affecting EC_a (Cook and Walker, et al., 1992). A weakness of this interpretative process is "equivalence" when several interacting properties can produce similar EMI responses. As an example, the simultaneous interaction of soil depth, clay, and moisture contents can lower the relationship of EC_a with any one of these properties alone. An increase in clay content will produce higher EC_a. Spatially, however, the increase in EC_a associated with an increase in clay can be offset by reduced soil depth over more electrically resistive materials and/or lower water content, both of which lower the measured EC_a. These three interacting and spatially varying soil properties can result in EC_a measurements that are difficult to associate with one soil property (e.g., clay content) alone. Because of equivalence, tacit knowledge, onsite soil sampling, and data analysis are required to decipher the exact site-specific causes for the variability in EC_a (Sommer et al., 2003).

Relationships between EC_a and soil properties can vary over short distances and foster inconsistent or ambiguous results (Carroll and Oliver, 2005). The effectiveness of EMI as a soil-mapping tool will depend upon the degree to which differences in physiochemical soil properties that affect EC_a correspond to differences in soil types. If a strong and meaningful relationship is established between soils and EC_a, field-scale EC_a mapping can be used to identify distinct zones, each with relatively homogenous soils and soil properties (Johnson et al., 2001).

39.4.1.2 Depth of Penetration

The depth of penetration of the electromagnetic current in the soil is dependent on the geometry (spacing and orientation) of the coils, frequency of the induced current, height above the soil surface, and the conductivity of the profiled material(s). The orientation of the transmitter and receiver coil axis (with respect to

the ground surface) affects the response from materials at different depths (McNeill, 1980). In the horizontal dipole orientation (HDO) or perpendicular geometry (PRP), GCMs are more sensitive to surface materials. In the vertical dipole orientation (VDO) or horizontal coplanar geometry (HCP), GCMs are more sensitive to deeper materials.

In most studies, a single GCM is used. The choice of GCM will depend on the objectives of the survey, soil conditions, and the desired depth of penetration (Lück et al., 2009). In the United States, the most widely used GCM in agriculture is the EM38 meter (Geonics Limited, Mississauga, Ontario) and the Dualem-1S meter (Dualem, Inc., Milton, Ontario), which have maximum theoretical penetration depths of 1.5 m. For these GCMs, the depth of penetration is considered "geometry limited" rather than "skin depth limited" because penetration depth is a function of the distance between the transmitter and receiver coils (McNeill, 1996). For geometry limited GCMs, larger intercoil spacings and lower operating frequency result in greater penetration depths. For multifrequency GCMs developed by Geophex Limited (Raleigh, NC) and GSSI (Salem, NH), the depth of penetration is dependent upon the operating frequency and the EC_a of the profiled material(s). With these GCMs, the depth of penetration is considered "skin depth limited" rather than "geometry limited" (Won, 1980, 1983; Won et al., 1996). Skin depth represents the maximum depth of penetration for a GCM operating at a specified frequency and sounding a medium of known conductivity. The skin depth (D) can be estimated using the following equation (McNeill, 1996):

$$D = \frac{500}{(sf)^{-2}}, \qquad (39.6)$$

where
    s is the ground conductivity (mS m^{-1})
    f is the frequency (kHz)

According to Equation 39.6, skin depth is inversely related to the operating frequency and the conductivity of the earthen materials. Low-frequency signals have longer periods of oscillation, lose energy less rapidly and, therefore, travel farther than high-frequency signals. Multifrequency sounding allows multiple depths to be profiled with one pass of the sensor (Won et al., 1996).

For all GCMs, surface and shallow layers contribute more to the overall response than deeper layers. de Jong et al. (1979) and Slavich (1990) reported that the actual depth of penetration is not fixed, but will vary depending on the EC_a of the profiled material(s). In general, GCMs will see deeper in electrically resistive than in conductive soils (Won et al., 1996).

Each instrument has characteristic depth sensitivity response curves. The theoretical depth of penetration has been arbitrarily defined as the depth at which 70% of a GCM's cumulative response is obtained in the overlying soil volume (Saey et al., 2008). Although contributions to the measured response come

from all depths within the theoretical depth of penetration, the largest contribution to the response comes from the "depth of observation," which is the depth that has the greatest contribution to the measured EC_a (Roy and Apparao, 1971). As noted by Roy and Apparao (1971), for all GCM, the depth of observation is often considerably shallower than the theoretical penetration depth specified by manufacturers. In saline soils, the high conductivity of surface and near-surface soil horizons contributes greatly to the depth-weighted response of GCMs and limits the depth of observation. In areas of highly conductive soils, manufacturer's specifications or skin depth estimations may provide fairly accurate estimates of the theoretical depth of penetration, but greatly overestimate the depth of observation.

The response of GCMs (EC_a) is directly proportional to the electrical conductivity of the soil under conditions of low induction numbers. These conditions (low induction numbers) are met when the intercoil spacing is less than the skin depth. Conditions of low induction numbers are typically satisfied in soils having relatively low conductivities (<100 mS m^{-1}) (McNeill, 1980). In saline soils, however, the low induction numbers condition is not fulfilled, and EC_a is not directly proportional to the electrical conductivity of the soil.

39.4.1.3 Temperature Effects

Soil electrical conductivity is an electrolytic process that is a function of the concentration and mobility of the dissolved ions present in the soil solution and on the surfaces of soil particles, and thus, is dependent on the temperature and phase of the soil water (Allred et al., 2005). Air temperature does not directly affect soil electrical conductivity, but does contribute to the instrument "drift" reported in some studies (Allred et al., 2005). Variations in soil temperature, however, do affect EC_a. As soil temperatures increase, the viscosity of water decreases, the mobility of dissolved ion increases, and, as a consequence, EC_a increases (McNeill, 1980; Allred et al., 2005). Seasonal fluctuations in soil temperature are greatest in the upper part of soil profiles and have the greatest affects on EC_a. Brevik et al. (2004) observed that daily changes in soil temperatures within the upper 10 cm do not significantly influence EC_a. Results from studies conducted by Allred et al. (2005) suggest a soil temperature threshold of 8°C, below which there is a substantial reduction in measured EC_a, especially when the soil is frozen. Apparent conductivity measurements are unreliable when frozen layers are present in soils (McKenzie et al., 1989).

Electrical conductivity increases at a rate of approximately 1.9% per °C increase. McKenzie et al. (1989), working on saline soils, stressed the importance of correcting the soil temperature to a standard temperature for more accurate conversions of EC_a to saturated paste electrical conductivity (EC_e). When comparing results from EMI surveys of the same area that are completed at different times of the year, the effect of temperature on EC_a is important and customarily expressed at a reference temperature of 25°C. The EC_a can be adjusted to a reference EC at 25°C (EC_{25})

using the following equation (United States Salinity Laboratory Staff, 1954):

$$EC_{25} = f_t EC_t, \quad (39.7)$$

where

f_t is a temperature conversion factor (see Table 15 in United States Salinity Laboratory Staff, 1954)

EC_t is the EC_a measured at a particular temperature

Approximations for the temperature conversion factor are also available in polynomial form (Stogryn, 1971; Rhoades et al., 1999; Wraith and Or, 1999) or other temperature conversion equations such as the one modified from Sheets and Hendrickx (1995) by Corwin and Lesch (2005):

$$EC_{25} = 0.4470 = 1.4034 \exp\left(-\frac{T}{26.8150}\right). \quad (39.8)$$

In most studies, only a single soil temperature is measured. A common concern is the depth at which the measurement should be made. Lück et al. (2009) suggested that because the maximum sensitivity of the EM38 meter operated in the vertical dipole orientation is between depths of 30–50 cm, temperature measurements should be made within this depth interval.

Changes in air temperature, humidity, and atmospheric electricity (spherics) affect the stability or drift of GCMs (Sudduth et al., 2003). Sudduth et al. (2001) reported instrument drifts as great as 3 mS m^{-1} for the EM38 meter. The causes for this drift are not entirely due to variations in air temperature but are a function of instrument instability over time (Sudduth et al., 2001). Instrument drift is more noticeable and has a greater impact on interpretations in areas of low EC_a. Suggested methods to minimize instrument drift include the use of a calibration traverse line (Sudduth et al., 2001), ample instrument warm-up time, shading the meter, and conducting surveys when temperatures are below 40°C (Robinson et al., 2004).

39.4.1.4 Instruments

An increasing number of commercially available GCMs are available (see Figure 39.12). Ground conductivity meters commonly used in soil investigations include the Dualem-1S and Dualem-2 meters (Dualem, Inc.); the EM31, EM38, EM38DD, and EM38-MK2 meters (Geonics Limited); GEM-2 sensor (Geophex Limited); and Profiler EMP-400 (Geophysical Survey

FIGURE 39.12 Selected EMI equipment. Pictured are the Veris Soil EC 3100 mapping system (a); the Profiler EMP-400 multifrequency sensor (b); the EM38 meter (c); and the Dualem-2 meter (d).

Systems, Inc.). The dual-geometry configuration of Dualem meters, the dual dipole orientations of the EM38DD meter, and the dual receiver–transmitter spacings of the EM38-MK2 meter allow simultaneous measurement of EC_a over two separate depth intervals. The aforementioned GCMs all have standard connectors for GPS communication and support data loggers and proprietary software, which are used to record and integrate EC_a and GPS data. Some GCMs, such as the multifrequency Profiler EMP-400, come with internal GPS.

Another sensor that is commonly used in precision agriculture and soil investigations is the towed-array resistivity [ER] unit (Figure 39.12a). This ER sensor measures the potential gradient resulting from the injection of electrical current into the soil through coulter electrodes. Soundings and profiles are made with an arrangement of current and potential electrodes known as arrays. The depth of penetration is "geometry limited" and dependent upon the spacing and the arrangement of the electrodes. Arrays are connected to a field computer or data logger and integrated with GPS. Compared with GCMs, towed-array resistivity units do not experience instrument drift, and are less susceptible to spherics and engine noise interference. However, towed-array resistivity units are invasive and require good ground contact. As a consequence, the use of towed-array resistivity units is restricted in cultivated fields and over rocky or frozen grounds.

In the United States, several towed-array resistivity units are manufactured by Veris Technologies (Salina, KS). Towed-array resistivity units used in agriculture include the Veris Soil EC 2000, 3100, and 3500 mapping systems. These mapping systems utilize direct-contact, coulter electrodes that are arranged in a modified Wenner configuration (Lund et al., 2000).

Each of the aforementioned GCM and ER instruments has distinct operational advantages and disadvantages (Sudduth et al., 2003). Comparative studies have generally revealed close similarities between EC_a data collected with different instruments (Sudduth et al., 1999, 2003; Doolittle et al., 2001, 2002a; Saey et al., 2008). However, differences in sensor calibration, depth sensitivity curves, and volume of soil material measured will affect measurements and result in slightly different EC_a values. In comparative studies using different GCM and ER units, the highest correlations were obtained with sensors having similar depth sensitivity curves (Sudduth et al., 1999, 2003). In general, differences in EC_a maps produced from data collected with different GCMs or ER units have been more noticeable over soils with highly contrasting layers (Sudduth et al., 2003).

39.4.2 Field Methodology

The speed, economy, and simplicity of EMI make this geophysical method attractive in agriculture. In the past, grids were laid out across survey areas and EC_a measurements were obtained at discrete stations, typically represented by the grid intersections. Presently, the synergism of EMI, GPS, and field data recording systems has enabled the rapid collection, storage, and processing of larger data sets and the more comprehensive coverage of sites. The speed and ease at which data are georeferenced and recorded greatly reduces survey time and makes practical the surveying of large areas. Kitchen et al. (2003, 2005) discuss the integration of EMI and GPS technologies to improve interpretations.

Integrated systems have been used at various spatial scales to map soil EC_a using either point or continuous sampling methods. Point sampling methods measure EC_a at specific stations, which are often located within a predetermined grid or at random points within a survey area. With this method, establishment of a grid and the positioning of measurement points are more time-consuming than the actual collection of EC_a data. As a consequence, compared with continuous sampling methods, the number of observation is often less for surveys conducted using point sampling methods. However, because of the greater time spent at each measurement point, a greater number of EC_a and position measurements can be collected and averaged. This procedure is known as "stacking" and is used to reduce unwanted background noise and improve data quality, but its use increases the time needed to complete the EMI survey. With the point sampling method, the operator typically moves sequentially along grid or traverse lines and measures EC_a in two dipole orientations or geometries. This method has been used most often for EMI surveys of small research areas (typically less than 40 ha) (Hendrickx et al., 1992; Khakural et al., 1998; Lesch et al., 1998; James et al., 2003; Jung et al., 2005).

Electromagnetic induction surveys of very large areas (300–10,000 km^2) have required larger grid intervals with more widely spaced stations. For surveys of very large areas, the spacing between measurement points has ranged from 0.5 (Williams and Arunin, 1990) to 5 km (Williams and Baker, 1982). Apparent conductivity data will be more generalized in surveys of very large areas as the short-range variability in soils and soil properties is masked by the coarse grid scale.

Continuous, simultaneous recording of both GPS and EC_a data in field computers as the GCM or ER is moved along traverse lines (continuous sampling) produces a greater number of measurements, provides more complete coverage of sites, and is less time-consuming than point sampling methods. Continuous sampling methods, however, may result in reduced positional accuracy, and GCMs must remain stable and in a fixed orientation when operated in the continuous mode. Consequently, either two separate surveys are required to collect EC_a data in both dipole orientations, or a GCM with a dual-geometry configuration, dual receiver–transmitter spacings, or multiple frequencies must be used to simultaneously collect measurements over different depth intervals.

Depending on the size of the survey area and the resources available, either a mobile or pedestrian survey is used to collect EC_a data. For large, open areas or whenever the total number of observations exceeds 1600 data points, Freeland et al. (2002) recommended the use of mobile over pedestrian surveys. In large open fields, mobile surveys provide more comprehensive

coverage, greater acquisition efficiency, and less operator fatigue than pedestrian surveys (Cannon et al., 1994; Chen et al., 2000; Johnson et al., 2001; Corwin et al., 2003; Greve and Greve, 2004; Corwin et al., 2006). Cannon et al. (1994) reported a fivefold increase in productivity when mobile survey methods were used instead of pedestrian survey methods. However, in rough terrain and under unfavorable field conditions, mobile EMI surveys are impractical and pedestrian surveys must be carried out.

For continuous sampling methods, the inline (along traverse lines) and lateral (distance between traverse lines) sampling intensities are more critical to map quality (precision and accuracy of boundaries) than the applied interpolation technique (e.g., kriging, inverse distance) (James et al., 2003; Farahani and Flynn, 2007). In general, slower speeds of advance, smaller traverse line spacing, and faster data acquisition rates lead to higher sampling intensities and map quality (Farahani and Flynn, 2007). These workers found that map quality decreased only slightly as the interval between traverse lines was increased from 2.5 to 50 m. As the interval between traverse lines was increased to greater than 80 m, however, map quality declined very rapidly as map delineations became highly distorted and incorrectly classified. King et al. (2001) considered a traverse line spacing of about 24 m reasonable for larger fields. Using these examples as guidelines, the final decision on the most suitable traverse spacing should be tempered by tacit knowledge of the soils and expert site assessments.

39.4.3 Applications

Apparent conductivity maps have been used to identify soil and management zones and to predict differences in soils, soil properties, and crop yields. Differences in EC_a have been associated with spatial changes in one or more specific soil physiochemical properties. Electromagnetic induction has been used to assess and map soil salinity, assist site-specific management, characterize variability in soil physiochemical properties, and direct soil sampling.

39.4.3.1 Soil Salinity Assessment

Methods commonly used to appraise soil salinity include: (1) visual observation of crop appearance (Soil Survey Division Staff, 1993), (2) sampling and measurement of the saturated paste extract (EC_e) (United States Salinity Laboratory Staff, 1954), (3) Wenner array or four-electrode sensor (Halverson et al., 1977; Nadler, 1981), (4) EMI (Corwin and Rhoades, 1982, 1984, 1990; Williams and Baker, 1982; Wollenhaupt et al., 1986; McKenzie et al., 1989; Rhoades et al., 1989a, 1989b), and (5) multi- and hyperspectral imagery (Corwin, 2008).

In soil surveys, visual crop observations and saturated paste extract method are the most commonly used methods to distinguish and map phases of soil salinity (United States Salinity Laboratory Staff, 1954; Soil Survey Division Staff, 1993). While visual observations are adequate for salinity mapping, these interpretations provide only a qualitative measure of soil salinity and are dependent on the presence of plant cover and surface salts. The saturated paste extract method is commonly accepted as the most accurate measure of soil salinity (McNeill, 1985), but it requires considerable resources for field sampling and laboratory analysis. Thus, only a limited number of soil samples can be collected and analyzed. Because of the limited sample numbers and the high spatial variability of soil salinity, analysis of saturated paste extracts is of limited value for the assessment of soil salinity across large management units or for soil surveys (Corwin, 2008).

In the early 1970s, the USDA-ARS Salinity Laboratory began to use bulk soil electrical conductivity as a field measure of soil salinity (Rhoades and Ingvalson, 1971), and techniques to evaluate salinity with EMI were subsequently developed (de Jong et al., 1979; Rhoades and Corwin, 1981;Williams and Baker, 1982). This early research was directed principally toward the vertical profiling of salinity through the root zone (Corwin, 2008). Several empirical relationships were developed but were site specific and could not be extrapolated and used at other sites without calibration (Corwin, 2008).

A major challenge in using EMI to map soil salinity has been the conversion of apparent conductivity (EC_a) into the conductivity of the saturated paste extract (EC_e); the most commonly used measure of soil salinity. A number of models have been developed that relate EC_a to EC_e (Wollenhaupt et al., 1986; McKenzie et al., 1989; Rhoades et al., 1989a, 1989b; Corwin and Rhoades, 1990; Slavich, 1990; Cook and Walker, 1992; Lesch et al., 1995a, 1995b; Johnston et al., 1997). Depth-weighted calibration models (Wollenhaupt et al., 1986; McKenzie et al., 1989) are useful in mapping relative spatial differences in soil salinity within units of management, but do not assess salinity variation with depth in the profile (Cassel et al., 2009). Rhoades et al. (1989a, 1989b) developed empirical equations and coefficients for determining soil salinity at different depths across units of management using a dual pathway parallel conductance model. This deterministic model relies on estimates of water content, texture, bulk density, and soil temperature to estimate soil salinity from EC_a data. Multilinear regression models for determining soil salinity at different soil depth intervals from EC_a data were later developed (Lesch et al., 1992, 1995a, 1995b). These stochastic models are independent of the salinity profile shape, but require the collection and analysis of soil samples for calibration (Cassel et al., 2009). Unfortunately, models relating EC_a to EC_e are both time dependent and site specific (Lesch et al., 1998). As a consequence, calibration equations and modeled results usually cannot be extrapolated to other sites (Cassel et al., 2009). To potentially reduce the number of samples needed for model calibration, directed soil sampling models have developed to characterize spatial variations in soil salinity across units of management (Lesch et al., 1992).

In areas of saline soils, of the properties that influence EC_a, the concentration of soluble salts typically have the greatest effect on EC_a (van der Lelij, 1983; Johnston et al., 1997). Williams and Baker (1982) estimated that 65%–70% of the variation in EC_a in saline soils can be explained by changes in the concentration of soluble salts alone, and EC_a can provide reasonably accurate estimates of soil salinity (Williams and Baker, 1982; van der Lelij, 1983; Diaz and Herrero, 1992). Although the estimates of salinity levels are reasonably accurate, they may not meet accuracy standards for specific objectives (Rhoades et al., 1989a; Johnston et al., 1997).

39.4.3.2 Soil Property Prediction, Site-Specific Management, and High-Intensity Soil Surveys

Apparent conductivity is a helpful surrogate for characterizing the variability of soil properties that are too difficult, time-consuming, or expensive to measure by other means (Jaynes, 1995; Stafford, 2000; Corwin et al., 2008). As apparent conductivity is a function of several interacting soil properties, the use of EC_a to characterize the variability of a soil physiochemical property is not always straightforward and uncomplicated. Effective use of EMI to predict soil properties requires sampling and analysis of a limited number of ground-truth observations for site-specific calibration of EC_a to the property of interest.

In nonsaline soils, EC_a is largely affected by water content, clay content, mineralogy, CEC, and organic matter content, with reported correlations ranging from 0.4 to 0.8 (Farahani et al., 2005). Electromagnetic induction has been used extensively to assess depths to argillic horizons, claypans, fragipans (Stroh et al., 1993; Sudduth and Kitchen, 1993; Doolittle et al., 1994; Sudduth et al., 1995; Mueller et al., 2003), and bedrock (Zalasiewicz et al., 1985; Palacky and Stephens, 1990; Bork et al., 1998). Apparent conductivity has been used as a surrogate measure of SWC (Kachanoski et al., 1988, 1990; Sheets and Hendrickx, 1995; Khakural et al., 1998; Waine et al., 2000; Mueller et al., 2003; Hezarjaribi and Sourell, 2007; Huth and Poulton, 2007; Tromp-van Meerveld and McDonnell, 2009), clay content (Williams and Hoey, 1987; Mueller et al., 2003; Sommer et al., 2003; King et al., 2005; Weller et al., 2007; Cockx et al., 2009; Harvey and Morgan, 2009), CEC (Triantafilis et al., 2009), exchangeable Ca and Mg (McBride et al., 1990), soil pH (Bianchini and Mallarino, 2002), soil organic carbon (Jaynes, 1996; Johnson et al., 2001), field-scale leaching rates of solutes (Slavich and Yang, 1990), herbicide partition coefficients (Jaynes et al., 1994), and available N (Eigenberg et al., 2002). In these reported studies, EC_a was either directly related to the soil property under investigation or the property (such as soil organic carbon) was associated with a property (moisture and clay content) that the sensor measures. The strength and significance of the relationships among EC_a and other properties, however, can vary both within and among individual

fields (Bekele et al., 2005; Carroll and Oliver, 2005), and often EC_a has a complex relationship with different, interacting soil properties (Farahani et al., 2005).

Site-specific management adjusts farming practices to measured variations in soil properties by dividing farmlands into management zones that have different seeding and chemical requirements (Mulla, 1993). Electromagnetic induction and ER have been used as accurate, fast, and inexpensive means of mapping soils and soil properties at a level of resolution that meets the requirements of site-specific management (Jaynes et al., 1995; Sudduth et al., 1995; Lund et al., 2000; Bianchini and Mallarino, 2002; Corwin and Lesch, 2003; Godwin and Miller, 2003; Adamchuk et al., 2004; Hedley et al., 2004; Corwin et al., 2008). As large, high-resolution data sets can be collected from mobile EMI platforms equipped with GPS, EMI is well suited to surveying of large units of management (Freeland et al., 2002; Adamchuk et al., 2004).

Spatial EC_a data identifies areas with reasonably homogenous soils (Doolittle et al., 1996; Frogbrook and Oliver, 2007) and often corresponds well with soil patterns across the landscape. These data can provide information on the distribution of soils at different scales of resolution; from field and landscape scales (Khakural et al., 1998; King et al., 2005; Weller et al., 2007) to regional inventories (Brevik and Fenton, 2003; Harvey and Morgan, 2009). Soil surveys incorporating EC_a data may have greater detail than those prepared with conventional methods due to the spatial intensity and abundance of data (Jaynes, 1995; Hedley et al., 2004; Doolittle et al., 2008) and may more accurately identify small included areas of dissimilar soils (Fenton and Lauterbach, 1999).

Kravchenko (2008) and Kravchenko et al. (2002) used EC_a to distinguish soils belonging to different soil drainage classes. Several studies have associated changes in EC_a with changes in the clay content of soils (King et al., 2005; Weller et al., 2007; Harvey and Morgan, 2009). In Coastal Plain landscapes, Anderson-Cook et al. (2002) used EC_a to classify soil series with an accuracy of greater than 85%. Doolittle et al. (2008) used EC_a to identify and delineate soils to improve the quality of a high-intensity soil survey in Illinois, but independent observations showed only slight improvements in the taxonomic purity of the map units when EC_a data were used. Inclusion of EC_a data in the survey effort, however, resulted in refinement of map units and increased confidence in mapping decisions.

Maps of EC_a are frequently compared with soil maps to better understand and interpret the distribution and extent of soils within delineations, adjust the placement of soil boundary lines, and refine soil maps. An example of this qualitative methodology is an EMI survey that was conducted in south-central Illinois to determine the distribution of sodium-affected soils; Natrudalfs and Natraqualfs. In this landscape, Natrudalfs and Natraqualfs are intermingled with non-sodium-affected Argiudolls, Argiaquolls, and Endoaqualfs. Areas of the sodium-affected soils are variable in size, distribution, and shape;
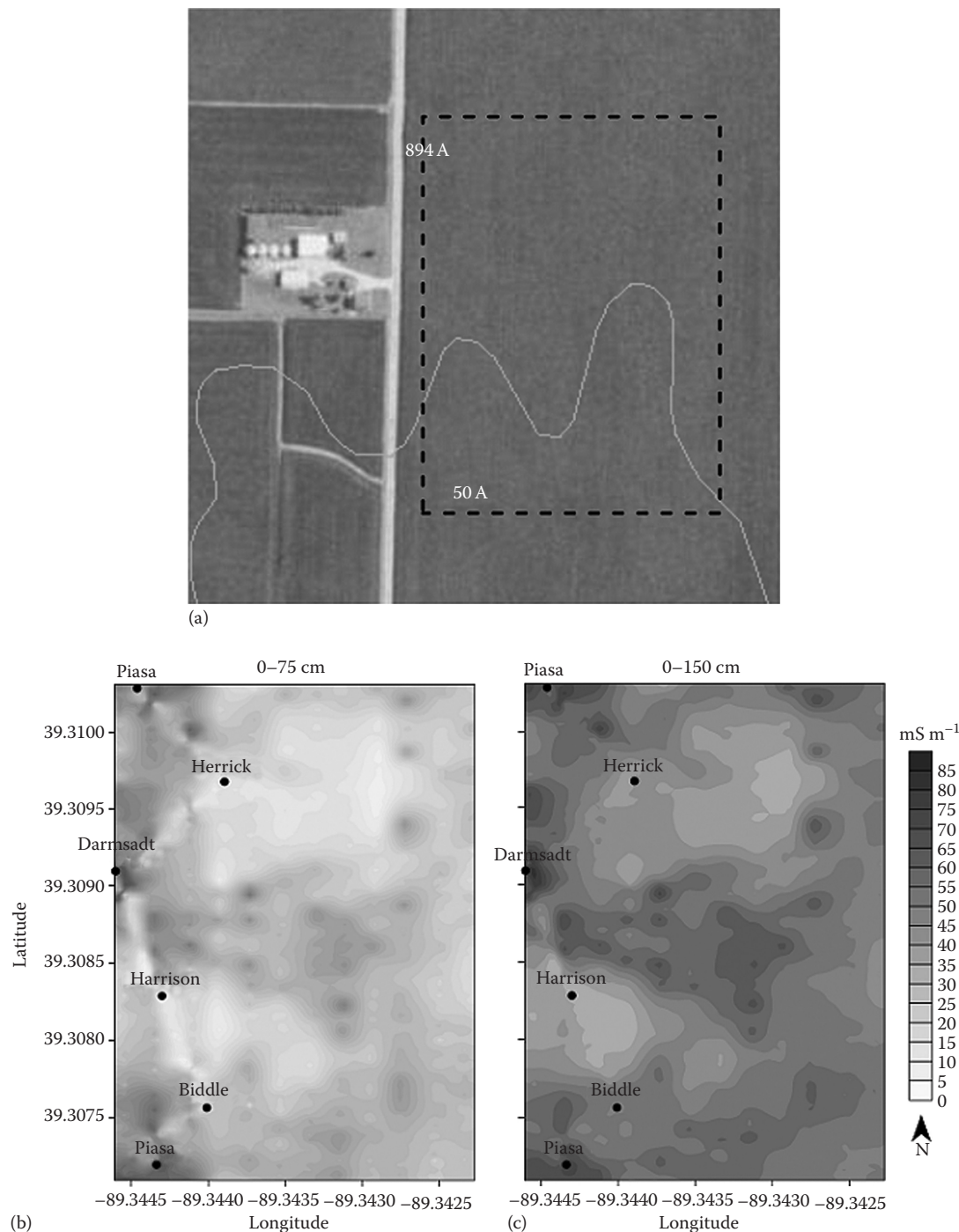
FIGURE 39.13 EMI survey of a site in south-central Illinois. The location is outlined on the soil map (Soil Survey Staff, 2009) as shown in (a). Spatial distributions of EC_a collected with the EM38-MK2-2 meter operated in the VDO are shown in the two plots: one for the shallower sensing 50 cm (b) and one for the deeper sensing 100 cm (c) intercoil spacings.

and their extent in individual soil polygons is unknown. The study site (enclosed by rectangle on the soil map in Figure 39.13a) contains two relatively large delineations of Herrick–Biddle–Piasa silt loams, 0%–2% slopes (894A), and Virden silty clay loam, 0%–2% slopes (50A). Of the named soils, only Piasa is a sodium-affected soil (Table 39.1).

The EMI survey of the area (Figure 39.13b) indicated that EC_a ranged from 0.6 to 84.0 mS m^{-1} and increased with depth and proximity to the water table. From this survey, sites for soil description were located from visual examination of prominent EC_a zones, and the soil identified at each observation point (Figure 39.13b). The soils identified at the three observation sites in areas

TABLE 39.1 Taxonomic Classifications of the Soil Series Identified at the Central Illinois Site

Soil Series	Taxonomic Classification
Biddle	Fine, smectic, mesic Aquic Argiudolls
Darmstadt	Fine-silty, mixed, superactive, mesic Aquic Natrudalfs
Harrison	Fine-silty, mixed, superactive, mesic Oxyaquic Argiudolls
Herrick	Fine, smectitic, mesic Aquic Argiudolls
Oconee	Fine, smectitic, mesic Udollic Endoaqualfs
Piasa	Fine, smectitic, mesic Mollic Natraqualfs
Virden	Fine, smectitic, mesic Vertic Argiaquolls

of $EC_a > 60\,mS\,m^{-1}$ (in VDO) were identified as sodium-affected soils (Darmstadt and Piasa). In contrast, the sites in areas with $EC_a < 55\,mS\,m^{-1}$ (in VDO) were identified as the non-sodium-affected soils (Harrison, Biddle, and Herrick) (Table 39.1).

39.4.3.3 Directed Sampling

Software has been developed by the USDA-ARS Salinity Laboratory (Riverside, California) to select optimal soil sampling locations at field scales based on EC_a data (Lesch et al., 1995a, 1995b, 2000; Lesch, 2005). The ESAP (EC_e Sampling, Assessment, and Prediction) software though designed to predict EC_e from EC_a data can be used to predict other physiochemical properties as well. The program employs prediction-based sampling and modeling as a cost-effective alternative to geostatistical techniques, which are more sample intensive (Eigenberg et al., 2008). The ESAP methodology relies on high-density EC_a data to direct low-density soil sampling to calibrate suitable predictive equations. One of ESAP's components is the response surface sampling design (RSSD), which employs a directed-sampling approach to locate calibration sample points from observed locations and magnitudes of EC_a data (Lesch, 2005; Corwin et al., 2006; Eigenberg et al., 2008).

The ESAP-RSSD was used in a salinity assessment of a 16 ha field in Grand Forks County, North Dakota (Figure 39.14) to identify 20 optimally located sample points to characterize variability and to predict soil salinity across this unit of management. Because EC_a of measurements in the EM38DD HDO mode (shallow mode, ~0–75 cm) were less than those in the VDO mode (deep mode, ~0–150 cm), soils in this area were persumed to have a normal salt profile, that is, salt concentration increases with depth. The locations of the 20 RSSD-selected sampling sites and their numerical identifier are shown in Figure 39.14a.

After laboratory data were available for soil layers sampled at each of the 20 sample points, the ESAP calibrate program was used to convert EC_a data into EC_e data using a stochastic calibration model (Lesch et al., 1992, 1995a, 1995b, 2000). This stochastic method uses multilinear regression models to predict EC_e from EC_a data and provide a spatial representation of soil EC_e across the area (Figure 39.15).
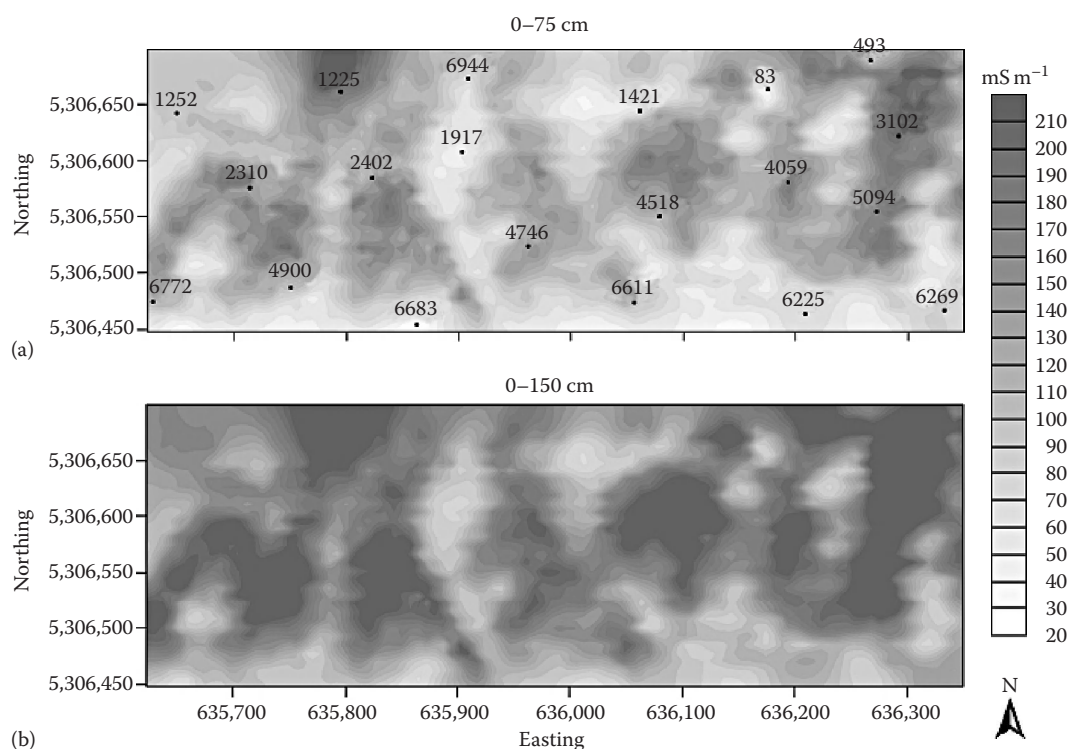
FIGURE 39.14 EC_a data from an area of saline Bearden soils in North Dakota. Data were collected with an EM38DD meter operated in the shallower sensing HDO (a) and the deeper sensing VDO (b). Numbers on the upper plot identify the locations of 20 sampling sites determined by the RSSD program of the ESAP software suite.
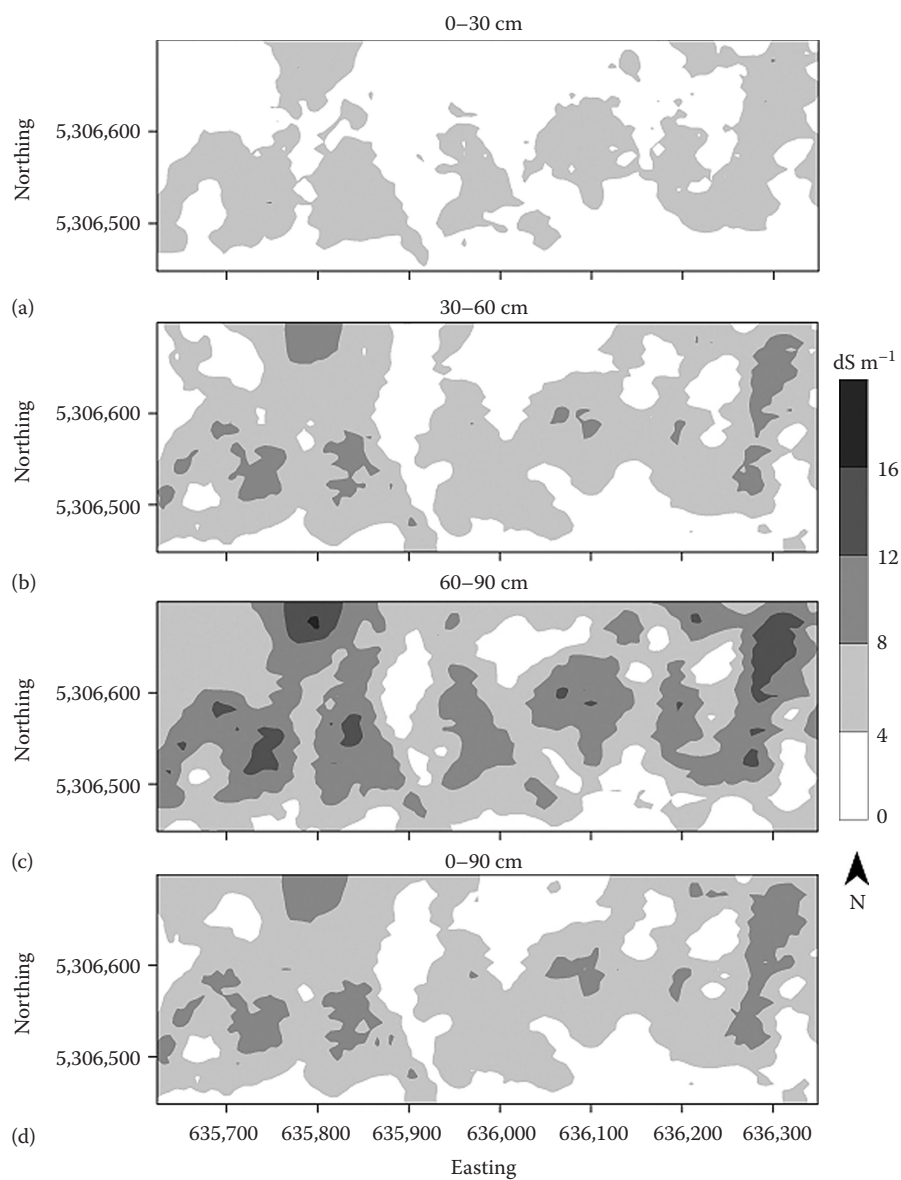
FIGURE 39.15   Distribution of estimated EC_e for different depth intervals ((a) 0–30 cm, (b) 30–60 cm, (c) 60–90 cm, and (d) 0–90 cm) within the Bearden saline site (Figure 39.14). Estimates from the stochastic model in the ESAP calibrate program.

Within the area evaluated, mean predicted EC_e ranged from 1.6 to 8.78 dS m^{-1} in the 0–30 and the 60–90 cm depth intervals, respectively (Table 39.2).

TABLE 39.2   Basic statistics for a field of saline Bearden soils in Grand Forks County, North Dakota. Data predicted for different soil depth intervals using the ESAP stochastic model. All EC_e values are expressed in dS m^{-1}

	0–30 cm	30–60 cm	60–90 cm	0–90 cm
Number	7040	7040	7040	7040
Minimum	0.00	0.00	1.08	0.38
25th percentile	2.39	3.28	4.60	1.36
75th percentile	4.76	6.47	9.28	6.83
Maximum	9.34	11.67	16.73	12.58
Mean	1.60	2.13	8.78	2.20

References

Adamchuk, V.I., J.W. Hummel, M.T. Morgan, and S.K. Upadhyaya. 2004. On-the-go soil sensors for precision agriculture. Comput. Electron. Agric. 44:71–91.

Al-fares, W., M. Bakalowicz, R. Guerin, and M. Dukhan. 2002. Analysis of the karst aquifer structure of the Lamalou area (Herault, France) with ground-penetrating radar. J. Appl. Geophys. 51:97–106.

Allred, B.J., M.R. Ehsani, and J.J. Daniels. 2008. General considerations for geophysical methods applied to agriculture, p. 3–16. *In* B.J. Allred, J.J. Daniels, and M.R. Ehsani (eds.) Handbook of agricultural geophysics. CRC Press, Taylor & Francis, Boca Raton, FL.

Allred, B.J., M.R. Ehsani, and M.R. Saraswat. 2005. The impact of temperature and shallow hydrologic conditions on the magnitude and spatial pattern consistency of electromagnetic induction measured soil electrical conductivity. Trans. Am. Soc. Agric. Eng. 48:2123–2135.

Anderson-Cook, C.M., M.M. Alley, J.K.F. Roygard, R. Khosla, R.B. Nobl, and J.A. Doolittle. 2002. Differentiating soil types using electromagnetic conductivity and crop yield maps. Soil Sci. Soc. Am. J. 66:1562–1570.

Annan, A.P. 2001. Ground penetrating radar workshop notes. Sensors and Software Inc., Mississauga, ON, Canada.

Annan, A.P. 2009. Electromagnetic principles of ground penetrating radar, p. 3–40. *In* H.M. Jol (ed.) Ground penetrating radar: Theory and applications. Elsevier Science, Amsterdam, the Netherlands.

Annan, A.P., S.W. Cosway, and J.D. Redman. 1991. Water table detection with ground penetrating radar, p. 494–495. *In* Expanded Abstracts 61st Annual Meeting. Society of Exploration Geophysicists. Vol. 1. Tulsa, OK.

Asmussen, L.E., H.F. Perkins, and H.D. Allison. 1986. Subsurface descriptions by ground-penetrating radar for watershed delineation. GA Agric. Exp. Sta. Res. Bull. 340.

Barr, G.L. 1993. Application of ground-penetrating radar methods in determining hydrogeologic conditions in a karst area, West-Central Florida. U.S. Geol. Sur. Water Resour. Invest. Rep. 92-4141.

Bekele, A., W.H. Hudnall, J.J. Daigle, J.A. Prudente, and M. Wolcott. 2005. Scale dependent variability of soil electrical conductivity by indirect measures of soil properties. J. Terramech. 42:339–351.

Ben-Dor, E., G. Metternicht, N. Goldshleger, E. Mor, V. Mirlas, and U. Bason. 2009. Review of remote sensing based methods to assess soil salinity, p. 39–60. *In* G. Metternicht and J.A. Zinck (eds.) Remote sensing of soil salinization. CRC Press, Boca Raton, FL.

Benson, R., and R. Glaccum. 1979. Test report: The application of ground-penetrating radar to soil surveying for National Aeronautical and Space Administration (NASA). Technos Inc., Miami, FL.

Bentley, L.R., and N.M. Trenholm. 2002. The accuracy of water table elevation estimates determined from ground penetrating radar data. J. Eng. Environ. Geophys. 7:37–53.

Beres, M., and F.P. Haeni. 1991. Application of ground-penetrating radar methods in hydrogeologic studies. Ground Water 29:375–386.

Beres, M., P. Huggenberger, A.G. Green, and H. Horstmeyer. 1999. Using two- and three-dimensional georadar methods to characterize glaciofluvial architecture. Sediment. Geol. 129:1–24.

Bianchini, A.A., and A.P. Mallarino. 2002. Soil-sampling alternatives and variable-rate liming for a soybean-corn rotation. Agron. J. 94:1355–1366.

Bohling, G.C., M.P. Anderson, and C.R. Bentley. 1989. Use of ground penetrating radar to define recharge areas in the Central Sand Plain. Technical Completion Report G1458-03. Geology and Geophysics Department. University of Wisconsin-Madison, Madison, WI.

Boll, J., R.P.G. van Rijn, K.W. Weiler, J.A. Ewen, J. Daliparthy, S.J. Herbert, and T.S. Steenhuis. 1996. Using ground-penetrating radar to detect layers in a sandy field soil. Geoderma 70:117–132.

Bork, E.W., N.E. West, J.A. Doolittle, and J.L. Boettinger. 1998. Soil depth assessment of sagebrush grazing treatments using electromagnetic induction. J. Range Manag. 51:469–474.

Brevik, E.C., and T.E. Fenton. 2003. Use of the Geonics EM-38 to delineate soil in a loess over till landscape, southwestern Iowa. Soil Survey Horiz. 44:16–24.

Brevik, E.C., T.E. Fenton, and R. Horton. 2004. Effect of daily soil temperature fluctuations on soil electrical conductivity as measured with the Geonics® EM-38. Precis. Agric. 5:145–152.

Burgoa, B., R.S. Mansell, G.J. Sawka, P. Nkedi-Kizza, J. Capece, and K. Campbell. 1991. Spatial variability of depth to Bh horizon in Florida Haplaquods using ground-penetrating radar. Soil Crop Sci. Soc. Fla. Proc. 50:125–130.

Buynevich, I.V., H.M. Jol, and D.M. FitzGerald. 2009. Coastal environments, p. 299–322. *In* H.M. Jol (ed.) Ground penetrating radar: Theory and applications. Elsevier, Amsterdam, the Netherlands.

Campbell, J.E. 1990. Dielectric properties and influence of conductivity in soils at one to fifty megahertz. Soil Sci. Soc. Am. J. 54:332–341.

Cannon, M.E., R.C. McKenzie, and G. Lachapelle. 1994. Soil salinity mapping with electromagnetic induction and satellite-based navigation methods. Can. J. Soil Sci. 74:335–343.

Carroll, Z.L., and M.A. Oliver. 2005. Exploring the spatial relations between soil physical properties and apparent electrical conductivity. Geoderma 128:354–374.

Cassel, F., D. Goorahoo, D. Zoldoske, and D. Adhikari. 2009. Mapping soil salinity using ground-based electromagnetic induction, p. 199–233. *In* G. Metternicht and J.A. Zinck (eds.) Remote sensing of soil salinization. CRC Press, Boca Raton, FL.

Cassidy, N.J. 2009. Electrical and magnetic properties of rocks, soils, and fluids, p. 41–72. *In* H.M. Jol (ed.) Ground penetrating radar: Theory and applications. Elsevier Science, Amsterdam, the Netherlands.

Chapman, H., J. Adcock, and J. Gater. 2009. An approach to mapping buried prehistoric palaeosols of the Atlantic seaboard in Northwest Europe using GPR, geoarchaeology and GIS and the implications for heritage management. J. Archaeol. Sci. 36:2308–2313.

Chen, F.D., E. Kissel, and D. Adkins. 2000. Mapping soil hardpans with penetrometer and electrical conductivity. *In* 2000 ASAE Annual International Meeting, Technical Papers: Engineering solutions for a new Century. 2037–2048. American Society of Agric. Engineers, St. Joseph, MI.

Chernetsov, E.A., N.A. Beletsky, and M.Y. Baev. 1988. Radar profiling of peat and gyttja deposits, p. 15–21. *In* Proc. 8th Int. Peat Congr. Leningrad, USSR. International Peat Society, Jyska, Finland.

Cockx, L., M. Van Meirvenne, U.W.A. Vitharana, L.P.C. Verbeke, D. Simpson, T. Saey, and F.M.B. Van Coille. 2009. Extracting topsoil information from EM38DD sensor data using neural network approach. Soil Sci. Soc. Am. J. 73:1–8.

Collins, M.E. 1992. Soil taxonomy: A useful guide for the application of ground penetrating radar, p. 125–132. *In* P. Hanninen and S. Autio (eds.) Proc. Fourth Int. Conf. Ground-Penetrating Radar. Rovaniemi, Finland, 8–13 June 1992. Geological Survey of Finland, Special Paper 16.

Collins, M.E. 2008. History of ground-penetrating radar applications in agriculture, p. 45–55. *In* B.J. Allred, J.J. Daniels, and M.R. Ehsani (eds.) Handbook of agricultural geophysics. CRC Press, Boca Raton, FL.

Collins, M.S., M. Crum, and P. Hanninen. 1994. Using ground-penetrating radar to investigate karst landscape in north-central Florida. Geoderma 61:1–15

Collins, M.E., and J.A. Doolittle. 1987. Using ground-penetrating radar to study soil microvariability. Soil Sci. Soc. Am. J. 51:491–493.

Collins, M.E., J.A. Doolittle, and R.V. Rourke. 1989. Mapping depth to bedrock on a glaciated landscape with ground-penetrating radar. Soil Sci. Soc. Am. J. 53:1806–1812.

Collins, M.E., G.W. Schellentrager, J.A. Doolittle, and S.F. Shih. 1986. Using ground-penetrating radar to study changes in soil map unit composition in selected histosols. Soil Sci. Soc. Am. J. 50:408–412.

Comas, X., L. Slater, and A. Reeve. 2004. Geophysical evidence for peat basin morphology and stratigraphic controls on vegetation observed in a northern peatland. J. Hydrol. 295:173–184.

Comas, X., L. Slater, and A. Reeve. 2005. Stratigraphic controls on pool formation in a doomed bog inferred from ground penetrating radar (GPR). J. Hydrol. 315:40–51.

Cook, P.G., and G.R. Walker. 1992. Depth profiles of electrical conductivity from linear combinations of electromagnetic induction measurements. Soil Sci. Soc. Am. J. 56:1015–1022.

Corwin, D.L. 2008. Past, present, and future trends in soil electrical conductivity measurements using geophysical methods, p. 17–44. *In* B.J. Allred, J.J. Daniels, and M.R. Ehsani (eds.) Handbook of agricultural geophysics. CRC Press, Boca Raton, FL.

Corwin, D.L., S.R. Kaffka, J.W. Hopmans et al. 2003. Assessment and field-scale mapping of soil quality properties of a saline sodic soil. Geoderma 114:231–259.

Corwin, D.L., and S.M. Lesch. 2003. Application of soil electrical conductivity to precision agriculture: Theory, principles, and guidelines. 2003. Agron. J. 95:455–471.

Corwin, D.L., and S.M. Lesch. 2005. Apparent electrical conductivity measurements in agriculture. Computers and Electronics in Agriculture 46:11–43.

Corwin, D.L., S.M. Lesch, J.D. Oster, and S.R. Kaffka. 2006. Monitoring management-induced spatio-temporal changes in soil quality through soil sampling directed by apparent electrical conductivity. Geoderma 131:369–387.

Corwin, D.L., S.M. Lesch, P.J. Shouse, R. Soppe, and J.E. Ayers. 2008. Delineating site-specific management units using geospatial EC$_a$ measurements, p. 247–254. *In* B.J. Allred, J.J. Daniels, and M.R. Ehsani (eds.) Handbook of agricultural geophysics. CRC Press, Boca Raton, FL.

Corwin, D.L., and J.D. Rhoades. 1982. An improved technique for determining soil electrical conductivity—Depth relations from above ground electromagnetic induction measurements. Soil Sci. Soc. Am. J. 46:517–520.

Corwin, D.L., and J.D. Rhoades. 1984. Measurements of inverted electrical conductivity profiles using electromagnetic induction. Soil Sci. Soc. Am. J. 48:288–291.

Corwin, D.L., and J.D. Rhoades. 1990. Establishing soil electrical conductivity—Depth relations from electromagnetic induction measurements. Commun. Soil Sci. Plant Anal. 21:861–901.

Curtis, J.O. 2001. Moisture effects on the dielectric properties of soils. Trans. Geosci. Remote Sens. 39:125–128.

Daniels, D.J. 2004. Ground-penetrating radar. 2nd Ed. The Institute of Electrical Engineers, London, U.K.

Daniels, J.J., B. Allred, M. Collins, and J. Doolittle. 2003. Geophysics in soil science, p. 1–5. *In* R. Lal (ed.) Encyclopedia of soil science. Marcel Dekker, New York.

Daniels, J.J., M. Reza Ehsani, and B.J. Allred. 2008. Ground-penetrating radar methods, p. 129–145. *In* B.J. Allred, J.J. Daniels, and M.R. Ehsani (eds.) Handbook of agricultural geophysics. CRC Press, Boca Raton, FL.

Davis, J.L., and A.P. Annan. 1989. Ground-penetrating radar for high resolution mapping of soil and rock stratigraphy. Geophys. Prospect. 37:531–551.

Davis, J.L., R.W.D. Killey, A.P. Annan, and C. Vaughan. 1984. Surface and borehole ground-penetrating radar surveys for mapping geologic structures, p. 681–712. *In* Proc. National Water Well Association/Environmental Protection Agency Conference on Surface and Borehole Geophysical Methods in Ground Water Investigations. 7–9 February 1984, San Antonio, TX.

de Jong, E., A.K. Ballantyne, D.R. Cameron, and D.L. Read. 1979. Measurement of apparent electrical conductivity of soils by an electromagnetic induction probe to aid salinity surveys. Soil Sci. Soc. Am. J. 43:810–812.

Diaz, L., and J. Herrero. 1992. Salinity estimates in irrigated soils using electromagnetic induction. Soil Sci. 154:151–157.

Doolittle, J.A. 1987. Using ground-penetrating radar to increase the quality and efficiency of soil surveys, p. 11–32. *In* Soil survey techniques. Special Publication No. 20. Soil Science Society of America, Madison, WI.

Doolittle, J.A., and L.E. Asmussen. 1992. Ten years of applications of ground-penetrating radar by the United States Department of Agriculture, p. 139–147. *In* P. Hanninen and

S. Autio (eds.) Proc. Fourth Int. Conf. Ground-Penetrating Radar. Rovaniemi, Finland, 8–13 June 1992. Geological Survey of Finland, Special Paper 16.

Doolittle, J.A., and M.E. Collins. 1995. Use of soil information to determine application of ground-penetrating radar. J. Appl. Geophys. 33:101–108.

Doolittle, J.A., and M.E. Collins. 1998. A comparison of EM induction and GPR methods in areas of karst. Geoderma 85:83–102.

Doolittle, J.A., M.E. Collins, and H.R. Mount. 1998. Assessing the appropriateness of GPR with a soil geographic database, p. 393–400. *In* R.G. Plumb (ed.) Proc. Seventh Int. Conf. Ground-Penetrating Radar, 27–30 May 1998, Radar Systems and Remote Sensing Laboratory, University of Kansas, Lawrence, KS.

Doolittle, J., J. Daigle, J. Kelly, and W. Tuttle. 2005. Using GPR to characterize plinthite and ironstone layers in ultisols. Soil Survey Horiz. 46:179–184.

Doolittle, J., P. Fletcher, and J. Turenne. 1990a. Estimating the thickness of organic materials in cranberry bogs. Soil Survey Horiz. 31:73–78.

Doolittle, J.A., M.A. Hardisky, and M.F. Gross. 1990b. A ground-penetrating radar study of active layer thicknesses in areas of moist sedge and wet sedge tundra near Bethel, Alaska, U.S.A. Arct. Alp. Res. 22:175–182.

Doolittle, J., G. Hoffmann, P. McDaniel, N. Peterson, B. Gardner, and E. Rowan. 2000. Ground-penetrating radar interpretations of a fragipan in Northern Idaho. Soil Survey Horiz. 41:73–82.

Doolittle, J.A., S.J. Indorante, D.K. Potter, S.G. Hefner, and W.M. McCauley. 2002a. Comparing three geophysical tools for locating sand blows in alluvial soils of southeast Missouri. J. Soil Water Conserv. 57:175–182.

Doolittle, J.A., B. Jenkinson, D. Hopkins, M. Ulmer, and W. Tuttle. 2006a. Hydropedological investigations with ground-penetrating radar (GPR): Estimating water-table depths and local ground water flow pattern in areas of coarse-textured soils. Geoderma 131:317–329.

Doolittle, J.A., F.E. Minzenmayer, S.W. Waltman, and E.C. Benham. 2002b. Ground-penetrating radar soil suitability map of the conterminous United States, p. 7–12. *In* Proc. Ninth Int. Conf. Ground-Penetrating Radar. 30 April–2 May 2002. Santa Barbara, CA, Proceedings of SPIE Volume 4158.

Doolittle, J.A., F.E. Minzenmayer, S.W. Waltman, and E.C. Benham. 2003. Ground-penetrating radar soil suitability maps. Environ. Eng. Geophys. J. 8:49–56.

Doolittle, J.A., F.E. Minzenmayer, S.W. Waltman, E.C. Benham, J.W. Tuttle, and S. Peaslee. 2006b. State ground-penetrating radar soil suitability maps, p. 1–8, Paper HYD; 13_mdx. *In* Proc. Eleventh Int. Conf. Ground-Penetrating Radar, 19–22 June 2006. Columbus, OH.

Doolittle, J.A., F.E. Minzenmayer, S.W. Waltman, E.C. Benham, J.W. Tuttle, and S. Peaslee. 2007. Ground-penetrating radar soil suitability map of the conterminous United States. Geoderma 141:416–421.

Doolittle, J., R. Murphy, G. Parks, and J. Warner. 1996. Electromagnetic induction investigations of a soil delineation in Reno County, Kansas. Soil Survey Horiz. 37:11–20.

Doolittle, J.A., and F.E. Nelson. 2009. Characterizing relict cryogenic macrostructures in mid-latitude areas of the USA with three-dimensional ground-penetrating radar. Permafrost Periglac. Process. 20:257–268.

Doolittle, J.A., M. Peterson, and T. Wheeler. 2001. Comparison of two electromagnetic induction tools in salinity appraisals. J. Soil Water Conserv. 56:257–262.

Doolittle, J.A., R.A. Rebertus, G.B. Jordan, E.I. Swenson, and W.H. Taylor. 1988. Improving soil-landscape models by systematic sampling with ground-penetrating radar. Soil Survey Horiz. 29:46–54.

Doolittle, J.A., K.A. Sudduth, N.R. Kitchen, and S.J. Indorante. 1994. Estimating depth to claypans using electromagnetic inductive methods. J. Soil Water Conserv. 49:552–555.

Doolittle, J.A., R.D. Windhorn, D.L. Withers, S.E. Zwicker, F.E. Heisner, and B.L. McLeese. 2008. Soil scientists revisit a high-intensity soil survey in northwest Illinois with electromagnetic induction and tradition methods. Soil Survey Horiz. 49:102–108.

Eigenberg, R.A., J.W. Doran, J.A. Nienaber, R.B. Ferguson, and B.L. Woodbury. 2002. Electrical conductivity monitoring of soil condition and available N with animal manure and a cover crop. Agric. Ecosyst. Environ. 88:183–193.

Eigenberg, R.A., S.M. Lesch, B. Woodbury, and J.A. Nienaber. 2008. Geospatial methods for monitoring a vegetative treatment area receiving beef-feedlot runoff. J. Environ. Qual. 35:S68–S77.

Farahani, H.J., G.W. Buchleiter, and M.K. Brodahl. 2005. Characterization of soil electrical conductivity variability in irrigated sandy and non-saline fields in Colorado. Trans. Am. Soc. Agric. Eng. 48:155–168.

Farahani, H.J., and R.L. Flynn. 2007. Map quality and zone delineation as affected by width of parallel swaths with mobile agricultural sensors. Biosyst. Eng. 96:151–159.

Farrish, K.W., J.A. Doolittle, and E.E. Gamble. 1990. Loamy substrata and forest productivity of sandy glacial drift soils in Michigan. Can. J. Soil Sci. 70:181–187.

Fenton, T.E., and M.A. Lauterbach. 1999. Soil map unit composition and scale of mapping related to interpretations for precision soil and crop management in Iowa, p. 239–251. *In* Proc. 4th Int. Conf. Precision Agriculture. Am. Soc. Agron, Madison, WI.

Fine, H. 1954. An effective ground conductivity map for continental United States. Proc. Inst. Radio Eng. 42:1405–1408.

Freeland, R.S., R.E. Yoder, J.T. Ammons, and L.L. Leonard. 2002. Mobilized surveying of soil conductivity using electromagnetic induction. Appl. Eng. Agric. 18:121–126.

Frogbrook, Z.L., and M.A. Oliver. 2007. Identifying management zones in agricultural fields using spatially constrained classification of soil and ancillary data. Soil Use Manage. 23:40–51.

Galagedara, L.W., G.W. Parkin, and J.D. Redman. 2003. An analysis of the ground-penetrating radar direct ground wave method for soil water content measurements. Hydrol. Process. 17:3615–3628.

Gerber, R., P. Felix-Henningsen, T. Behrens, and T. Scholten. 2010. Application of ground-penetrating radar as a tool for nondestructive soil-depth mapping of Pleistocene periglacial slope deposits. J. Plant Nutr. Soil Sci. 173:173–184.

Gish, T.J., W.P. Dulaney, K.-J.S. Kung, C.S. Daughtry, J.A. Doolittle, and P.T. Miller. 2002. Evaluating use of ground-penetrating radar for identifying subsurface flow paths. Soil Sci. Soc. Am. J. 66:1620–1629.

Godwin, R.J., and P.C.H. Miller. 2003. A review of the technologies for mapping within-field variability. Biosyst. Eng. 84:393–407.

Grant, J.A., and P.H. Schultz. 1994. Erosion of ejecta at Meteor Crater: Constraints from ground-penetrating radar, p. 789–803. *In* Proc. Fifth Int. Conf. Ground-Penetrating Radar. 12–14 June 1994, Kitchner, Ontario, Canada, Waterloo Centre for Groundwater Research and the Canadian Geotechnical Society.

Grasmueck, M. 1996. 3-D ground-penetrating radar applied to fracture imaging in gneiss. Geophysics 61:1050–1064.

Grasmueck, M., and A.G. Green. 1996. 3-D georadar mapping: Looking into the subsurface. Environ. Eng. Geosci. 2:195–220.

Grasmueck, M., R. Weger, and H. Horstmeyer. 2004. Three-dimensional ground-penetrating radar imaging of sedimentary structures, fractures, and archaeological features at submeter resolution. Geology 32:933–936.

Greenhouse, J.P., and D.D. Slaine. 1983. The use of reconnaissance electromagnetic methods to map contaminant migration. Ground Water Monitor. Rev. 3:47–59.

Greve, M.H., and M.B. Greve. 2004. Determining and representing width of soil boundaries using electrical conductivity and multigrids. Comput. Geosci. 30:569–578.

Grote, K., S. Hubbard, and Y. Rubin. 2003. Field-scale estimation of volumetric water content using ground-penetrating radar ground wave techniques. Water Resour. Res. 39:1321: SBH 5-1 to 5-13.

GSSI. 2008. New mapping module for RADAN 6. Subsurface solutions. Official newsletter of Geophysical Survey Systems, Inc., May 2008:4.

Gustafsson, J. 2007. Widening the scope of GPR surveys. GeoDrilling Int. 136:26–27.

Halverson, A.D., J.D. Rhoades, and C.A. Reule. 1977. Soil salinity four-electrode conductivity relationship for soils of the northern Great Plains. Soil Sci. Soc. Am. J. 41:966–971.

Hanninen, P. 1992. Application of ground-penetrating radar techniques to peatland investigations, p. 217–221. *In* P. Hanninen and S. Autio (eds.) Proc. Fourth Int. Conf. Ground-Penetrating Radar. Rovaniemi, Finland, 8–13 June 1992, Geological Survey of Finland, Special Paper 16.

Harari, Z. 1996. Ground-penetrating radar (GPR) for imaging stratigraphic features and groundwater in sand dunes. J. Appl. Geophys. 36:43–52.

Harvey, O.R., and C.L.S. Morgan. 2009. Predicting regional-scale soil variability using single calibrated apparent soil electrical conductivity model. Soil Sci. Soc. Am. J. 73:164–169.

Hedley, C.B., I.J. Yule, C.R. Eastwood, T.G. Sheperd, and G. Arnold. 2004. Rapid identification of soil textural and management zones using electromagnetic induction sensing in soils. Aust. J. Soil Res. 42:389–400.

Hendrickx, J.M., H.B. Baerends, Z.I. Raza, M. Sadig, and M.A. Chaudhry. 1992. Soil salinity assessments by electromagnetic induction of irrigated lands. Soil Sci. Soc. Am. J. 56:1933–1941.

Hezarjaribi, A., and H. Sourell. 2007. Feasibility study of monitoring the total available water content using non-invasive electromagnetic induction-based and electrode-based soil electrical conductivity measurements. Irrig. Drain. 56:53–65.

Hinkel, K.M., J.A. Doolittle, J.G. Bockheim et al. 2001. Detection of subsurface permafrost features with ground-penetrating radar, Barrow, Alaska. Permafrost Periglac. Process. 12:179–190.

Holden, J., T.P. Burt, and M. Vilas. 2002. Application of ground-penetrating radar to the identification of subsurface piping in blanket peat. Earth Surf. Process. Landf. 27:235–249.

Holloway, A.L., and J.C. Mugford. 1990. Fracture characterization in granite using ground probing radar. CIM Bull. 83:61–70.

Hubbard, S.S., K.S. Grote, and Y. Rubin. 2002. Mapping the volumetric soil water content of a California vineyard using high-frequency GPR ground wave data. Lead. Edge Explor. 21:552–559.

Huisman, J.A., S.S. Hubbard, J.D. Redman, and A.P. Annan. 2003. Measuring soil water content with ground penetrating radar. A review. Vadose Zone J. 2:476–491.

Huisman, J.A., J.J.C. Snepvangers, W. Bouten, and G.B.M. Heuvelink. 2002. Mapping spatial variation in surface water content: Comparison of ground-penetrating radar and time domain reflectometry. J. Hydrol. 269:194–207.

Huisman, J.A., C. Sperl, W. Bouten, and J.M. Verstraten. 2001. Soil water content measurements at different scales: Accuracy of time domain reflectometry and ground-penetrating radar. J. Hydrol. 245:48–58.

Huth, N.I., and P.L. Poulton. 2007. An electromagnetic induction method for monitoring variations in soil moisture in agroforestry systems. Aust. J. Soil Res. 45:6372.

Iivari, T.A., and J.A. Doolittle. 1994. Computer simulations of depths to water table using ground-penetrating radar in topographically diverse terrains, p. 11–20. *In* K. Kovar and J. Soveri (eds.) Ground water quality management. Proc. GQM 93, International Association of Hydrological Sciences. September 1993, Tallinn, Estonia.

James, I.T., T.W. Waine, R.I. Bradley, J.C. Taylor, and R.J. Godwin. 2003. Determination of soil type boundaries using electromagnetic induction scanning techniques. Biosyst. Eng. 86:421–430.

Jaynes, D.B. 1995. Electromagnetic induction as a mapping aid for precision farming, p. 153–156. *In* Clean water, clean environment, 21st century: Team agriculture working to protect water resources. 5–8 March 1995, Kansas City, MO.

Jaynes, D.B. 1996. Mapping the areal distribution of soil parameters with geophysical techniques, p. 205–216. *In* Application of GIS to the modeling of non-point sources of pollution in the vadose zone. Vol. 48. SSSA Special Publication.

Jaynes, D.B., T.S. Colvin, and J. Ambuel. 1995. Yield mapping by electromagnetic induction, p. 383–394. *In* P.C. Robert, R.H. Rust, and W.E. Larson (eds.) Proc. Second Int. Conf. Precision Management for Agricultural Systems. 27–30 March 1994, Minneapolis, MN. American Society of Agronomy, Madison, WI.

Jaynes, D.B., J.M. Novak, T.B. Moorman, and C.A. Cambardella. 1994. Estimating herbicide partition coefficients from electromagnetic induction measurements. J. Environ. Qual. 24:36–41.

Johnson, D.G. 1987. Use of ground-penetrating radar for determining depth to the water table on Cape Cod, Massachusetts, p. 541–554. *In* Proc. First National Outdoor Action Conf. Aquifer Restoration, Ground Water Monitoring Geophys. Methods. 18–21 May 1987. Las Vegas, NV. National Water Well Association, Dublin, OH.

Johnson, C.K., J.W. Doran, H.R. Duke, B.J. Wienhold, K.M. Eskridge, and J.F. Shanahan. 2001. Field-scale conductivity mapping for delineating soil condition. Soil Sci. Soc. Am. J. 65:1829–1837.

Johnson, R.W., R. Glaccum, and R. Wojtasinski. 1979. Application of ground penetrating radar to soil survey. Soil Crop Sci. Soc. Fla. Proc. 39:68–72.

Johnston, M.A., M.J. Savage, J.H. Moolman, and H.M. du Plessis. 1997. Evaluation of calibration methods for interpreting soil salinity from electromagnetic induction measurements. Soil Sci. Soc. Am. J. 61:1627–1633.

Junck, M.B., and H.M. Jol. 2000. Three-dimensional investigation of geomorphic environments using ground penetrating radar, p. 314–318. *In* D. Noon (ed.) Proc. Eighth Int. Conf. Ground-Penetrating Radar. 23–26 May 2000. Gold Coast, Queensland, Australia. The University of Queensland, Queensland, Australia.

Jung, W.K., N.R. Kitchen, K.A. Sudduth, R.J. Kremer, and P.P. Motavalli. 2005. Relationship of apparent soil electrical conductivity to claypan soil properties. Soil Sci. Soc. Am. J. 69:883–892.

Kachanoski, R.G., E. de Jong, and I.J. van Wesenbeeck. 1990. Field scale patterns of soil water storage from non-contacting measurements of bulk electrical conductivity. Can. J. Soil Sci. 70:537–541.

Kachanoski, R.G., E.G. Gregorich, and I.J. van Wesenbeeck. 1988. Estimating spatial variations of soil water content using noncontacting electromagnetic inductive methods. Can. J. Soil Sci. 68:715–722.

Kettridge, N., X. Comas, A. Baird, L. Slater, M. Strack, D. Thompson, H. Jol, and A. Binley. 2008. Ecohydrologically important subsurface structures in peatlands revealed by ground-penetrating radar and complex conductivity surveys. J. Geophys. Res. G: Biogeosci. 113:G04030.

Khakural, B.R., P.C. Robert, and D.R. Hugins. 1998. Use of non-contacting electromagnetic inductive method for estimating soil moisture across a landscape. Commun. Soil Sci. Plant Anal. 29:2055–2065.

King, J.A., P.M.R. Dampney, R.M. Lark, T.R. Mayr, and R.I. Bradley. 2001. Sensing soil spatial variability by electromagnetic induction (EMI): It's potential in precision farming, p. 419–424. *In* G. Grenier and S. Blackmore (eds.) Proc. Third European Conf. Precision Agriculture. Agro Montpellier, Ecole Nationale Supérieure Agronomique, France.

King, J.A., P.M.R. Dampney, R.M. Lark, H.C. Wheeler, R.I. Bradley, and T.R. Mayr. 2005. Mapping potential crop management zones within fields: Use of yield-map series and patterns of soil physical properties identified by electromagnetic induction sensing. Precis. Agric. 6:167–181.

Kitchen, N.R., T.S. Drummond, E.D. Lund, K.A. Sudduth, and G.W. Buchleiter. 2003. Soil electrical conductivity and topography related yield for three contrasting soil-crop systems. Agron. J. 95:483–495.

Kitchen, N.R., K.A. Sudduth, D.B. Meyers, S.T. Drummond, and S.Y. Hong. 2005. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. Comput. Electron. Agric. 46:285–308.

Klein, K.A., and J.C. Santamarina. 2003. Electrical conductivity of soils: Underlying phenomena. J. Environ. Eng. Geophys. 8:263–273.

Kravchenko, A.N. 2008. Mapping soil drainage classes using topographic and soil electrical conductivity, p. 255–261. *In* B.J. Allred, J.J. Daniels, and M.R. Ehsani (eds.) Handbook of agricultural geophysics. CRC Press, Boca Raton, FL.

Kravchenko, A.N., G.A. Bollero, R.A. Omonode, and D.G. Bullock. 2002. Quantitative mapping of soil drainage classes using topographical data and soil electrical conductivity. Soil Sci. Soc. Am. J. 66:235–243.

Kung, K.-J.S., and S.V. Donohue. 1991. Improved solute-sampling in a sandy vadose zone using ground-penetrating radar. Soil Sci. Soc. Am. J. 55:1543–1545.

Kung, K.-J.S., and Z.-B. Lu. 1993. Using ground-penetrating radar to detect layers of discontinuous dielectric constant. Soil Sci. Soc. Am. J. 57:335–340.

Lane, J.W., M.L. Buursink, F.P. Haeni, and R.J. Versteeg. 2000. Evaluation of ground-penetrating radar to detect free-phase hydrocarbons in fractured rocks—Results of numerical modeling and physical experiments. Ground Water 38:929–938.

Lapen, D.R., B.J. Moorman, and J.S. Price. 1996. Using ground penetrating radar to delineate subsurface features along a wetland catena. Soil Sci. Soc. Am. J. 60:923–931.

Lebron, I., D.A. Robinson, S. Goldberg, and S.M. Lesch. 2004. The dielectric permittivity of calcite and arid zone soils with carbonate minerals. Soil Sci. Soc. Am. J. 68:1549–1559.

Lehmann, F., and A.G. Green. 1999. Semiautomatic georadar data acquisition in three dimensions. Geophysics 64:719–731.

Lesch, S.M. 2005. Sensor-directed response surface sampling designs for characterizing spatial variation in soil properties. Comput. Electron. Agric. 46:153–180.

Lesch, S.M., J. Herrero, and J.D. Rhoades. 1998. Monitoring for temporal changes in soil salinity using electromagnetic induction techniques. Soil Sci. Soc. Am. J. 62:232–242.

Lesch, S.M., J.D. Rhoades, and D.L. Corwin. 2000. ESAP-95 version 2.10R user manual and tutorial guide. Research Report 146. USDA-ARS George E. Brown, Jr. Salinity Laboratory, Riverside, CA.

Lesch, S.M., J.D. Rhoades, L.J. Lund, and D.L. Corwin. 1992. Mapping soil salinity using calibrated electromagnetic measurements. Soil Sci. Soc. Am. J. 56:540–548.

Lesch, S.M., D.J. Strauss, and J.D. Rhoades. 1995a. Spatial prediction of soil salinity using electromagnetic induction techniques. 1. Statistical prediction models: A comparison of multiple linear regression and cokriging. Water Resour. Res. 31:373–386.

Lesch, S.M., D.J. Strauss, and J.D. Rhoades. 1995b. Spatial prediction of soil salinity using electromagnetic induction techniques. 2. An efficient spatial sampling algorithm suitable for multiple linear regression model identification and estimation. Water Resour. Res. 31:387–398.

Lesmes, D.P., R. Herbstzuber, and D. Wertz. 1999. Terrain permittivity mapping: GPR measurements of near-surface soil moisture, p. 575–582. *In* Proc. SAGEEP'99. Oakland, CA. Environmental and Engineering Geophysical Society, Denver, CO.

Lowe, D.J. 1985. Application of impulse radar to continuous profiling of tephra-bearing lake sediments and peats: An initial evaluation. N. Z. J. Geol. Geophys. 28:667–674.

Lowry, C.S., D. Fratta, and M.P. Anderson. 2009. Ground penetrating radar and spring formation in a groundwater dominated peat wetland. J. Hydrol. 373:68–79.

Lück, E., R. Gebbers, J. Ruehlmann, and U. Sprangenberg. 2009. Electrical conductivity mapping for precision farming. Near Surf. Geophys. 7:15–25.

Lund, E.D., C.D. Christy, and P.E. Drummond. 2000. Using yield and soil electrical conductivity (EC$_a$) maps to derive crop production performance information. *In* P.C. Roberts, R.H. Rust, and W.E Larson (eds.) Proc. 5th Int. Conf. Precision Agric. (CD-ROM). Minneapolis, MN. 16–19 July 2000. American Society of Agronomy, Madison, WI.

Lyons, J.C., C.A. Mitchell, and T.M. Zobeck. 1988. Impulse radar for identification of features in soils. J. Aerosp. Eng. 1:18–27.

McBride, R.A., A.M. Gordon, and S.C. Shrive. 1990. Estimating forest soil quality from terrain measurements of apparent electrical conductivity. Soil Sci. Soc. Am. J. 54:290–293.

McKenzie, R.C., W. Chomistek, and N.F. Clark. 1989. Conversion of electromagnetic inductance readings to saturated paste extract values in soils for different temperature, texture, and moisture conditions. Can. J. Soil Sci. 69:25–32.

McNeill, J.D. 1980. Electrical conductivity of soils and rock. Technical Note TN-5. Geonics Limited, Mississauga, ON, Canada.

McNeill, J.D. 1985. Rapid, accurate mapping of soil salinity by electromagnetic ground conductivity meters. Soil Sci. Soc. Am. Spec. Publ. 30:209–229.

McNeill, J.D. 1996. Why doesn't Geonics Limited build multi-frequency EM31 or EM38 meter? Technical Note TN-30. Geonics Ltd., Mississauga, ON, Canada.

Mellett, J.S. 1995. Profiling of ponds and bogs using ground-penetrating radar. J. Paleolimnol. 14:233–240.

Mokma, D.L., R.J. Schaetzl, J.A. Doolittle, and E.P. Johnson. 1990a. Ground-penetrating radar study of ortstein continuity in some Michigan Haplaquods. Soil Sci. Soc. Am. J. 54:936–938.

Mokma, D.L., R.J. Schaetzl, E.P. Johnson, and J.A. Doolittle. 1990b. Assessing Bt horizon character in sandy soils using ground-penetrating radar: Implications for soil surveys. Soil Survey Horiz. 30:1–8.

Moorman, B.J., and F.A. Michel. 1997. Bathymetric mapping and sub-bottom profiling through lake ice with ground-penetrating radar. J. Paleolimnol. 18:61–73.

Moorman, B.J., S.D. Robinson, and M.M. Burgess. 2003. Imaging periglacial conditions with ground-penetrating radar. Permafrost Periglac. Process. 14:319–329.

Mueller, T.G., N.J. Hartsock, T.S. Stombaugh, S.A. Shearer, P.L. Cornelius, and R.I. Barnhisel. 2003. Soil electrical conductivity map variability in limestone soils overlain by loess. Agron. J. 95:496–507.

Mukherjee, D., E. Heggy, and S.D. Khan. 2010. Geoelectric constraints on radar probing of shallow water-saturated zones within karstified carbonates in semi-arid environments. J. Appl. Geophys. 70:181–191.

Mulla, D.J. 1993. Mapping and managing spatial patterns in soil fertility and crop yields, p. 15–26. *In* P.C. Robert, R.H. Rust, and W.E. Larson (eds.) Proc. Second Int. Conf. Precision Management Agricultural Systems. 27–30 March 1994. Minneapolis, MN. American Society of Agronomy, Madison, WI.

Munroe, J.S., J.A. Doolittle, K.M. Hinkel et al. 2007. Application of ground-penetrating radar imagery for three-dimensional visualization of near-surface structures in ice-rich permafrost, Barrow, Alaska. Permafrost Periglac. Process. 18:309–321.

Nadler, A. 1981. Field application of the four-electrode technique for determining soil solution conductivity. Soil Sci. Soc. Am. J. 45:30–34.

Nascimento da Silva, C.C., W.E. de Medeiros, E.F. Jarmin de Sa, and P.X. Neto. 2004. Resistivity and ground-penetrating radar images of fractures in a crystalline aquifer: A case study in Caicara farm—NE Brazil. J. Appl. Geophys. 56:295–307.

Neal, A. 2004. Ground-penetrating radar and its use in sedimentology: Principles, problems and progress. Earth Sci. Rev. 66:261–330.

Nobes, D.C., and B.G. Warner. 1992. Application of ground penetrating radar to a study of peat stratigraphy: Preliminary results. Geol. Surv. Can. Pap. 90–4:133–138.

Olhoeft, G.R. 1986. Electrical properties from 10−3 to 10+9 Hz—Physics and chemistry, p. 281–298. *In* Physics and chemistry of porous media II, American Institute of Physics Conference Proceedings, Ridgefield, CT.

Olhoeft, G.R. 1998. Electrical, magnetic, and geometric properties that determine ground-penetrating radar performance, p. 177–182. *In* Proc. Seventh Int. Conf. Ground-Penetrating Radar. 27–30 May 1998, University of Kansas, Lawrence, KS.

Olson, C.G., and J.A. Doolittle. 1985. Geophysical techniques for reconnaissance investigations of soils and surficial deposits in mountainous terrain. Soil Sci. Soc. Am. J. 49:1490–1498.

Palacky, G.J., and L.E. Stephens. 1990. Mapping of quaternary sediments in northeastern Ontario using ground electromagnetic methods. Geophysics 55:1596–1604.

Pelletier, R.E., J.L. Davis, and J.R. Rossiter. 1991. Peat analysis in the Hudson Bay Lowlands using ground-penetrating radar, p. 2141–2144. *In* Proc. Int. Geosci. Remote Sensing Sympos. June 1991, Helsinki, Finland.

Pipan, M., L. Baradello, E. Forte, and A. Prizzon. 2000. GPR study of bedding planes, fractures and cavities in limestone, p. 682–687. *In* Proc. Eighth Int. Conf. Ground-Penetrating Radar. 23–26 May 2000, Gold Coast, Queensland, Australia: Proceedings of SPIE—The International Society of Optical Engineering, Bellingham, WA.

Porsani, J.L., V.R. Elis, and F.Y. Hiodo. 2005. Geophysical investigations for the characterization of fractured rock aquifer in Itu, SE Brazil. J. Appl. Geophys. 57:119–128.

Puckett, W.E., M.E. Collins, and G.W. Schellentrager. 1990. Design of soil map units on a karst area in West Central Florida. Soil Sci. Soc. Am. J. 54:1068–1073.

Raper, R.L., L.E. Asmussen, and J.B. Powell. 1990. Sensing hard pan depth with ground-penetrating radar. Trans. Am. Soc. Agric. Eng. 33:41–46.

Rebertus, R.A., J.A. Doolittle, and R.L. Hall. 1989. Landform and stratigraphic influences on variability of loess thickness in northern Delaware. Soil Sci. Soc. Am. J. 53:843–847.

Rhoades, J.D., F. Chanduvi, and S. Lesch. 1999. Soil salinity assessment: Methods and interpretation of electrical conductivity measurements. FAO Irrigation and Drainage Paper #57. Food and Agriculture Organization of the United Nations, Rome, Italy.

Rhoades, J.D., and D.L. Corwin. 1981. Determining soil electrical conductivity-depth relations using an inductive electromagnetic soil conductivity meter. Soil Sci. Soc. Am. J. 45:255–260.

Rhoades, J.D., and R.D. Ingvalson. 1971. Determining salinity in field soils with soil resistance measurements. Soil Sci. Soc. Am. Proc. 35:54–60.

Rhoades, J.D., S.M. Lesch, P.J. Shouse, and W.J. Alves. 1989a. New calibrations for determining soil electrical conductivity depth relations from electromagnetic measurements. Soil Sci. Soc. Am. J. 53:74–79.

Rhoades, J.D., N.A. Manteghi, P.J. Shouse, and W.J. Alves. 1989b. Soil electrical conductivity and soil salinity: New formulation and calibrations. Soil Sci. Soc. Am. J. 53:433–439.

Rhoades, J.D., P.A. Raats, and R.J. Prather. 1976. Effects of liquid-phase electrical conductivity, water content, and surface conductivity on bulk soil electrical conductivity. Soil Sci. Soc. Am. J. 40:651–655.

Rial, F.I., M. Pereira, H. Lorenzo, and P. Arias. 2005. Acquisition and synchronism of GPR and GPS data. Application on road evaluation, p. 20–22. *In* L. Bruzzone (ed.) Image and signal processing for remote sensing XI. Article no. 598219, September 2006, Bruges, Belgium. Proceedings of SPIE.Vol. 5982. The International Society of Optical Engineering, Bellingham, WA.

Robinson, D.A., I. Lebron, S.M. Lesch, and P. Shouse. 2004. Minimizing drift in electrical conductivity measurements in high temperature environments using the EM-38. Soil Sci. Soc. Am. J. 68:339–345.

Rosa, E., M. Larocque, S. Pellerin, S. Gagné, and B. Fournier. 2008. Determining the number of manual measurements required to improve peat thickness estimations by ground penetrating radar. Earth Surf. Proc. Landf. 34:377–383.

Roy, A., and A. Apparao. 1971. Depth of investigation in direct current methods. Geophysics 36:943–959.

Saarenketo, T. 1998. Electrical properties of water in clay and silty soils. J. Appl. Geophys. 40:73–98.

Saey, T., D. Simpson, H. Vermeersch, L. Cockx, and M. Van Meirvenne. 2008. Comparing the EM38DD and DUALEM-21S sensors for depth-to-clay mapping. Soil Sci. Soc. Am. J. 73:7–12.

Sass, O., A. Friedman, G. Haselwanter, and K.F. Wetzel. 2010. Investigating thickness and internal structure of alpine mires using conventional and geophysical techniques. Catena 80:195–203.

Schellentrager, G.W., and J.A. Doolittle. 1991. Systematic sampling using ground-penetrating radar to study regional variation of a soil map unit. Soil Sci. Soc. Am. Spec. Publ. 28:199–214.

Schellentrager, G.W., J.A. Doolittle, T.E. Calhoun, and C.A. Wettstein. 1988. Using ground-penetrating radar to update soil survey information. Soil Sci. Soc. Am. J. 52:746–752.

Schenk, C.J., D.L. Gautier, G.R. Olhoeft, and J.E. Lucius. 1993. Internal structure of an eolian dune using ground-penetrating radar. Spec. Publ. Int. Assoc. Sediment. 16:61–69.

Sellmann, P.V., S.A. Arcone, and A.J. Delaney. 1983. Radar profiling of buried reflectors and the groundwater table. USA Cold Region Research and Engineering Laboratory Report 83-11, Hanover, NH.

Sheets, K.R., and J.M.H. Hendrickx. 1995. Noninvasive soil water content measurements using electromagnetic induction. Water Resour. Res. 31:2401–2409.

Shih, S.F., and J.A. Doolittle. 1984. Using radar to investigate organic soil thickness in the Florida Everglades. Soil Sci. Soc. Am. J. 48:651–656.

Shih, S.F., J.A. Doolittle, D.L. Myhre, and G.W. Schellentrager. 1986. Using radar for ground water investigation. J. Irrig. Drain. Eng. 112:110–118.

Simeoni, M.A., P.D. Galloway, A.J. O'Neil, and R.J. Gilkes. 2009. A procedure for mapping the depth to the texture contrast horizon of duplex soils in south-western Australia using ground penetrating radar, GPS and kriging. Aust. J. Soil Res. 47:613–621.

Slavich, P.G. 1990. Determining EC_a-depth profiles from electromagnetic induction measurements. Aust. J. Soil Res. 28:443–452.

Slavich, P.G., and J. Yang. 1990. Estimation of field scale leaching rates from chloride mass balance and electromagnetic induction measurements. Irrig. Sci. 11:7–14.

Smith, D.G., and H.M. Jol. 1995. Ground-penetrating radar: Antenna frequencies and maximum probable depths of penetration in Quaternary sediments. J. Appl. Geophys. 33:93–100.

Smith, M.C., G. Vellidis, D.L. Thomas, and M.A. Breve. 1992. Measurement of water table fluctuations in a sandy soil using ground penetrating radar. Trans. Am. Soc. Agric. Eng. 35:1161–1166.

Soil Survey Division Staff. 1993. Soil survey manual. USDA Agric. Handbook No. 18. U.S. Government Printing Office, Washington, DC.

Soil Survey Division Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd Ed. Handbook 436, USDA–NRCS, Washington, DC.

Soil Survey Staff. 2009. Web soil survey. USDA–NRCS, Lincoln, NE. Available online: http://websoilsurvey.nrcs.usda.gov/. Accessed 10 February 2009.

Sommer, M., M. Wehrhan, M. Zipprich et al. 2003. Hierarchical data fusion for mapping soil units at field scale. Geoderma 112:179–196.

Stafford, J.V. 2000. Implementing precision agriculture in the 21st century. J. Agric. Eng. Res. 76:267–275.

Steelman, C.M., and A.L. Endres. 2009. Evolution of high-frequency ground-penetrating radar direct ground wave propagation during frozen soil layer development. Cold Region Sci. Technol. 57:116–122.

Steenhuis, T.S., K.-J.S. Kung, and L.M. Cathles. 1990. Finding layers in the soil. Ground-penetrating radar as a tool in studies of ground water contamination Engineering. Cornell Q. (autumn):15–19.

Steenhuis, T., K. Vandenheuvel, K.W. Weiler et al. 1998. Mapping and interpreting soil textural layers to assess agri-chemical movement at several scales along the eastern seaboard (USA). Nutr. Cycl. Agroecosyst. 50:91–97.

Stevens, K.M., G.S. Lodha, A.L. Holloway, and N.M. Soonawala. 1995. The application of ground penetrating radar for mapping fractures in plutonic rocks within the Whiteshell Res. Area, Pinawa, Manitoba. Can. J. Appl. Geophys. 33:125–141.

Stogryn, A. 1971. Equations for calculating the dielectric constant of saline water. IEEE Trans. Microwave Theory Technol. MIT 19:733–736.

Stroh, J., S.R. Archer, L.P. Wilding, and J. Doolittle. 1993. Assessing the influence of subsoil heterogeneity on vegetation patterns in the Rio Grande Plains of south Texas using electromagnetic induction and geographical information system. College Station, TX. The Station:39–42.

Sudduth, K.A., S.T. Drummond, and N.R. Kitchen. 2001. Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. Comput. Electron. Agric. 31:239–263.

Sudduth, K.A., and N.R. Kitchen. 1993. Electromagnetic induction sensing of claypan depth. *In* Winter Meetings of the American Society of Agricultural Engineers, Paper No. 93-1550. Am. Soc. Agric. Eng., St. Joseph, MI.

Sudduth, K.A., N.R. Kitchen, G.A. Bollero, D.G. Bullock, and W.J. Wiebold. 2003. Comparison of electromagnetic induction and direct sensing of soil electrical conductivity. Agron. J. 95:472–482.

Sudduth, K., N.R. Kitchen, and S. Drummond. 1999. Soil conductivity sensing on claypan soils: Comparison of electromagnetic induction and direct methods, p. 979–990. *In* Proc. 4th Int. Conf. Precision Agric. ASA, CSSA and SSSA, Madison, WI.

Sudduth, K.A., N.R. Kitchen, D.H. Hughes, and S.T. Drummond. 1995. Electromagnetic induction sensing as an indicator of productivity on claypan soils, p. 671–681. *In* P.C. Robert, R.H. Rust, and W.E. Larson (eds.) Proc. Second Int. Conf. Precision Management Agricultural Systems. 27–30 March 1994, Minneapolis, MN. Am. Soc. Agron., Madison, WI.

Taylor, K.R., and M.E. Baker. 1988. Use of ground-penetrating radar in defining glacial outwash aquifers, p. 70–98. *In* Proc. FOCUS Conf. Eastern Regional Ground Water Issues. 27–29 September 1988, Stamford, CT. National Water Well Association, Dublin, OH,

Theimer, B.D., D.C. Nobes, and B.G. Warner. 1994. A study of the geoelectric properties of peatlands and their influence on ground-penetrating radar surveying. Geophys. Prospect. 42:179–209.

Tolonen, K., A. Rummukainen, M. Toikka, and I. Marttilla. 1984. Comparison between conventional and peat geological and improved electronic methods in examining economically important peatland properties, p. 1–10. *In* Proc. 7th Int. Peat Congress. 18–23 June 1984. Dublin, Ireland.

Tolonen, K., M. Tiuri, M. Toikka, and M. Saarilahti. 1982. Radiowave probe in assessing the yield of peat and energy in peat deposits in Finland. Suo 4–5:105–112.

Tomer, M.D., J. Boll, K.-J.S. Kung, T. Steenhuis, and J.L. Anderson. 1996. Detecting illuvial lamellae in fine sand using ground-penetrating radar. Soil Sci. 161:121–129.

Toshioka, T., T. Tsuchida, and K. Sasahara. 1995. Application of GPR to detecting and mapping cracks in rock slopes. J. Appl. Geophys. 33:119–124.

Triantafilis, J., S.M. Lesch, K. La Lau, and S.M. Buchanan. 2009. Field level digital mapping of cation exchange capacity using electromagnetic induction and a hierarchical spatial regression model. Aust. J. Soil Res. 47:651–663.

Tromp-van Meerveld, H.J., and J.J. McDonnell. 2009. Assessment of multi-frequency electromagnetic induction for determining soil moisture patterns at the hillslope scale. J. Hydrol. 368:56–67.

Truman, C.C., H.F. Perkins, L.E. Asmussen, and H.D. Allison. 1988. Some applications of ground-penetrating radar in the southern coast plain region of Georgia. GA Agric. Exp. Stn. Res. Bull. 362.

Turenne, J.D., J.A. Doolittle, and R. Tunstead. 2006. Ground-penetrating radar and computer graphic techniques are used to map and inventory histosols in southeastern Massachusetts. Soil Survey Horiz. 47:13–17.

Ulriksen, P. 1980. Investigation of peat thickness with radar, p. 126–129. *In* Proc. 6th Int. Peat Congress. Duluth, MN.

Ulriksen, C.P.F. 1982. Application of impulse radar to civil engineering. Ph.D. Thesis. Department of Engineering Geo., Lund University of Technology. Lund, Sweden.

United States Salinity Laboratory Staff. 1954. Diagnosis and improvement of saline and alkali soils. USDA Agric. Handbook No. 60. U.S. Government Printing Office, Washington, DC.

van der Lelij, A. 1983. Use of electromagnetic induction instrument (Type EM-38) for mapping soil salinity. Water Resources Commission, Murrumbidgee Division, New South Wales, Australia.

van Overmeeren, R.A. 1998. GPR and wetlands of the Netherlands, p. 251–268. *In* R.G. Plumb (ed.) Proc. Seventh International Conf. Ground-Penetrating Radar. 27–30 May 1998. Radar Systems and Remote Sensing Laboratory, University of Kansas, Lawrence, KS.

van Overmeeren, R.A. 2004. Georadar for hydrogeology. First Break 12:401–408.

Vellidis, G., M.C. Smith, D.L. Thomas, and L.E. Asmussen. 1990. Detecting wetting front movement in a sandy soil with ground-penetrating radar. Trans. Am. Soc. Agric. Engin. 33:1867–1874.

Violette, P. 1987. Surface geophysical techniques for aquifers and wellhead protection area delineation. Report No. EPA/440/12–87/016. Environmental Protection Agency, Office of Ground Water Protection, Washington, DC.

Waine, T.W., B.S. Blackmore, and R.J. Godwin. 2000. Mapping available water content and estimating soil textural class using electromagnetic induction. Eur. Agric. Eng. Paper No. 00-SW-044, Warwick, U.K.

Walther, E.G., A.M. Pitchford, and G.R. Olhoeft. 1986. A strategy for detecting subsurface organic contaminants, p. 357–381. *In* Natl. Water Well Association/API Conf. Petrol. Hydrocarb. Org. Chem. Ground Water Prevention, Detection, Restoration. Houston 12–14 November 1986. National Water Well Association, Dublin, OH.

Warner, B.G., D.C. Nobes, and B.D. Theimer. 1990. An application of ground-penetrating radar to peat stratigraphy of Ellice Swamp, southwestern Ontario. Can. J. Soil Sci. 27:932–938.

Wastiaux, C., L. Halleux, R. Schumacker, M. Streel, and J.M. Jacqmotte. 2000. Development of the Hautes-Fagnes peat bogs (Belgium): New perspectives using ground-penetrating radar. Suo 51:115–120.

Weller, U., M. Zipprich, M. Sommer, W. Zu Castell, and M. Wehrhan. 2007. Mapping clay content across boundaries at the landscape scale with electromagnetic induction. Soil Sci. Soc. Am. J. 71:1740–1747.

Welsby, J. 1988. The utilization of geo-radar in monitoring cutover peatlands. *In* Proc., 8th Int. Peat Congress, 99–107. Leningrad, USSR, International Peat Society, Jyska, Finland.

Williams, B.G., and S. Arunin. 1990. Inferring recharge/discharge areas from multifrequency electromagnetic induction measurements. CSIRO Division of Water Resources. Technical Memorandum 90/11 November.

Williams, B.G., and G.C. Baker. 1982. An electromagnetic induction technique for reconnaissance surveys of soil salinity hazards. Aust. J. Soil Res. 20:107–118.

Williams, B.G., and D. Hoey. 1987. The use of electromagnetic induction to detect the spatial variability of the salt and clay contents of soils. Aust. J. Soil Res. 25:21–27.

Wollenhaupt, N.C., J.L. Richardson, J.E. Foss, and E.C. Doll. 1986. A rapid method for estimating weighted soil salinity from apparent soil electrical conductivity measured with an aboveground electromagnetic induction meter. Can. J. Soil Sci. 66:315–321.

Won, I.J. 1980. A wideband electromagnetic exploration method—Some theoretical and experimental results. Geophysics 45:928–940.

Won, I.J. 1983. A sweep-frequency electromagnetic exploration method, p. 39–64. *In* A.A. Fitch (ed.) Development of geophysical exploration methods. Elsevier Appl. Sci. Publishers, Ltd., London, U.K.

Won, I.J., A. Dean, G. Keiswetter, R.A. Fields, and L.C. Sutton. 1996. GEM-2: A new multifrequency electromagnetic sensor. J. Environ. Eng. Geophys. 1:129–137.

Worsfold, R.D., S.K. Parashar, and T. Perrott. 1986. Depth profiling of peat deposits with impulse radar. Can. Geotech. J. 23:142–145.

Wraith, J.M., and D. Or. 1999. Temperature effects on soil bulk dielectric permittivity measured by time domain reflectometry: Experimental evidence and hypothesis development. Water Resour. Res. 35:361–369.

Wright, D.L., G.R. Olhoeft, and R.D. Watt. 1984. Ground-penetrating radar studies on Cape Cod, p. 666–680. *In* D.M. Nielsen (ed.) Surface and borehole geophysical methods in ground water investigations. National Water Well Association, Worthington, OH.

Zalasiewicz, J.A., S.J. Mathers, and J.D. Cornwell. 1985. The application of ground conductivity measurements to geological mapping. Quart. J. Eng. Geol. Hydrogeol. 18:139–148.

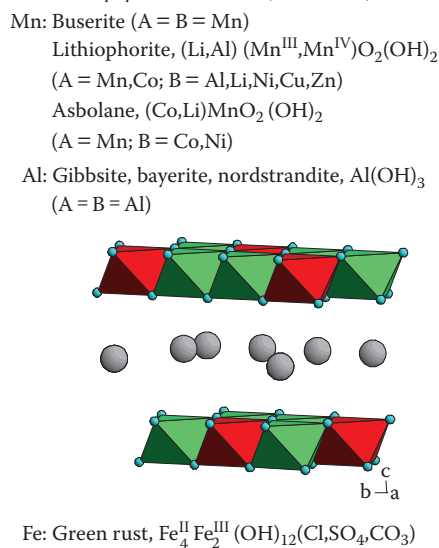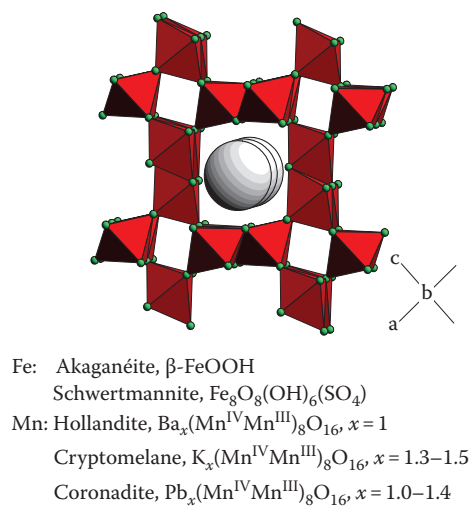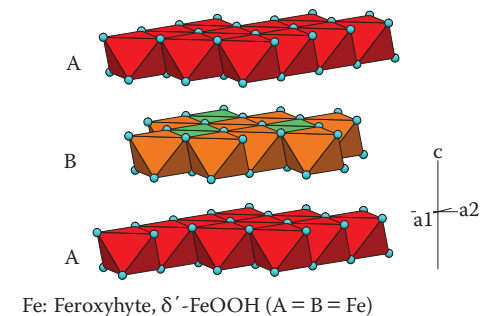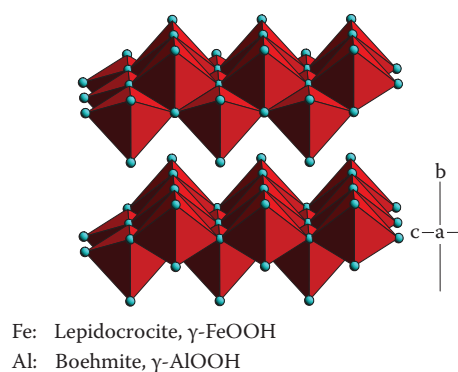Mn: Pyrolusite, MnO_2
 Manganite, $\gamma\text{-MnOOH}$
Ti: Rutile, TiO_2

Mn: Romanechite, $Ba_{0.66}(Mn^{IV}Mn^{III})_5O_{10} \cdot 1.3H_2O$

Fe: Goethite, $\alpha\text{-FeOOH}$
Mn: Ramsdellite, MnO_2
 Groutite, $MnOOH$
Al: Diaspore, $\alpha\text{-AlOOH}$

Mn: Todorokite,
 $(Na,Ca,K)_{0.3-0.5}(Mn^{IV}Mn^{III})_6O_{12} \cdot 3.5H_2O$

Fe: Lepidocrocite, $\gamma\text{-FeOOH}$
Al: Boehmite, $\gamma\text{-AlOOH}$

Fe: Feroxyhyte, $\delta'\text{-FeOOH}$ (A = B = Fe)
Mn: Buserite (A = B = Mn)
 Lithiophorite, $(Li,Al)(Mn^{III},Mn^{IV})O_2(OH)_2$
 (A = Mn,Co; B = Al,Li,Ni,Cu,Zn)
 Asbolane, $(Co,Li)MnO_2(OH)_2$
 (A = Mn; B = Co,Ni)
Al: Gibbsite, bayerite, nordstrandite, $Al(OH)_3$
 (A = B = Al)

Fe: Akaganéite, $\beta\text{-FeOOH}$
 Schwertmannite, $Fe_8O_8(OH)_6(SO_4)$
Mn: Hollandite, $Ba_x(Mn^{IV}Mn^{III})_8O_{16}$, $x = 1$
 Cryptomelane, $K_x(Mn^{IV}Mn^{III})_8O_{16}$, $x = 1.3-1.5$
 Coronadite, $Pb_x(Mn^{IV}Mn^{III})_8O_{16}$, $x = 1.0-1.4$

Fe: Green rust, $Fe_4^{II}Fe_2^{III}(OH)_{12}(Cl,SO_4,CO_3)$

FIGURE 22.1 Structural schemes for oxide minerals in soils.

Mn: Chalcophanite, $ZnMn_3O_7 \cdot 3H_2O$
          Birnessite, (Na,Ca,Mn^{II}) $Mn_7O_4 \cdot 2.8H_2O$

Si: α-Cristobalite, SiO_2

Fe: Hematite, α-Fe_2O_3
Al: Corundum, α-Al_2O_3
Ti,Fe: Ilmenite, $FeTiO_3$

Ti: Anatase, TiO_2

Ti: Brookite, TiO_2

Fe: Magnetite, Fe_3O_4
          Maghemite, γ-Fe_2O_3
Mn: Hausmannite, Mn_3O_4
Ti,Fe: Titanomagnetite, $Fe_{3-x}Ti_xO_4$
          Titanomaghemite, $Fe_{2-x}Ti_xO_3$

Ti,Fe: Pseudobrookite, $Fe_2^{III}TiO_5$

Ti,Fe: Pseudorutile, $Fe_2^{III}Ti_3O_8(OH)_2$

Si: α-Quartz, SiO_2
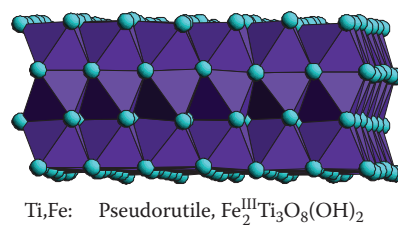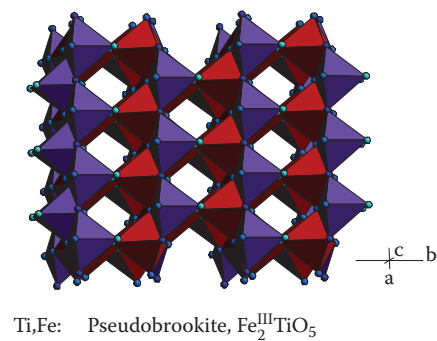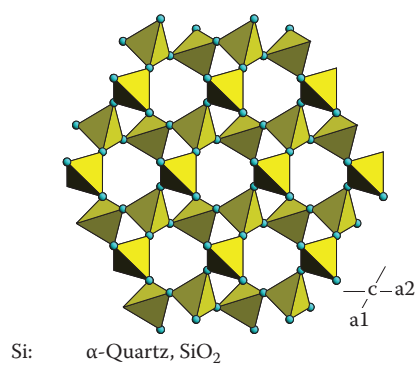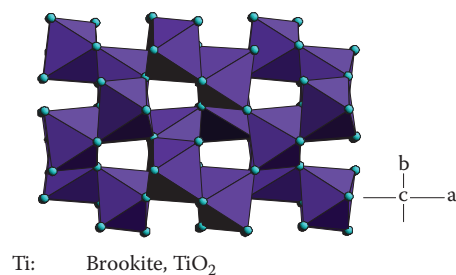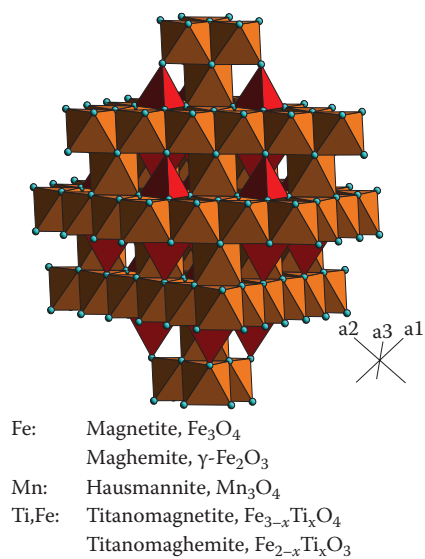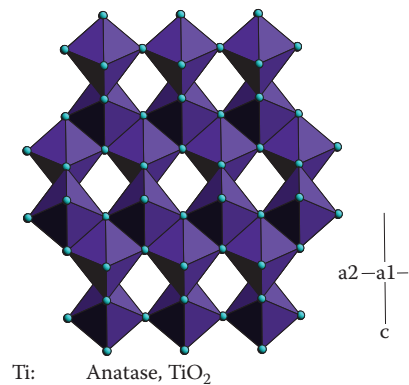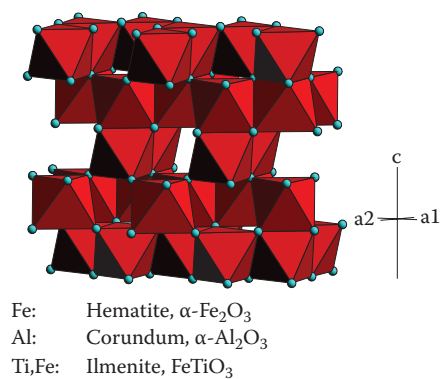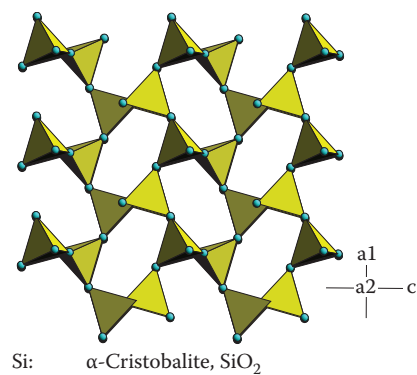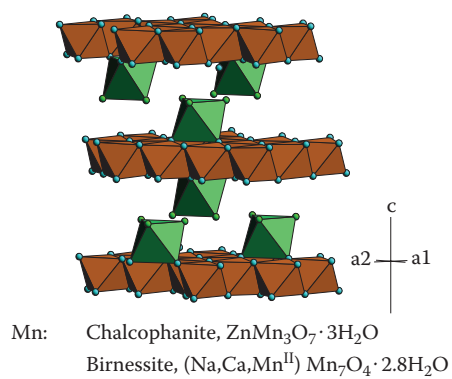
FIGURE 22.1 (continued)
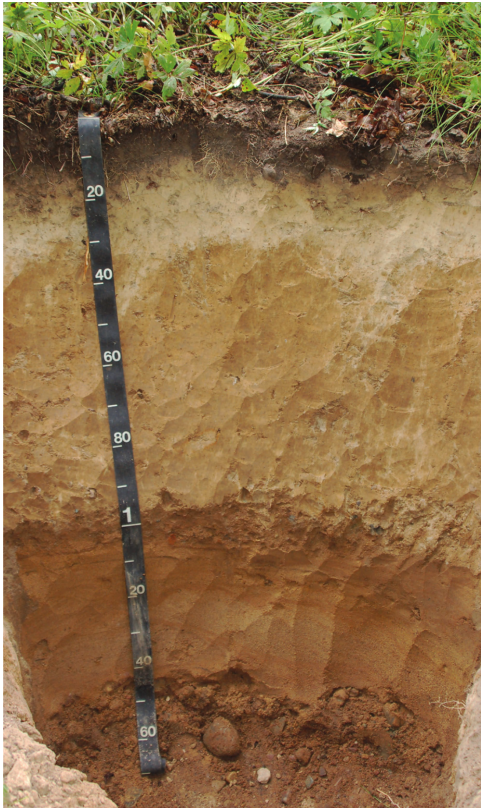
Alfisol

Andisol

Andisol (2)
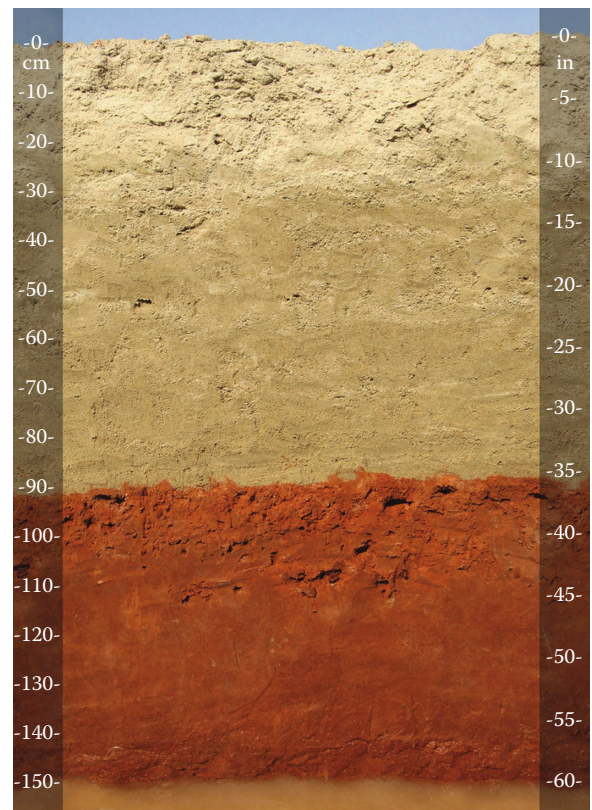
Aridisol
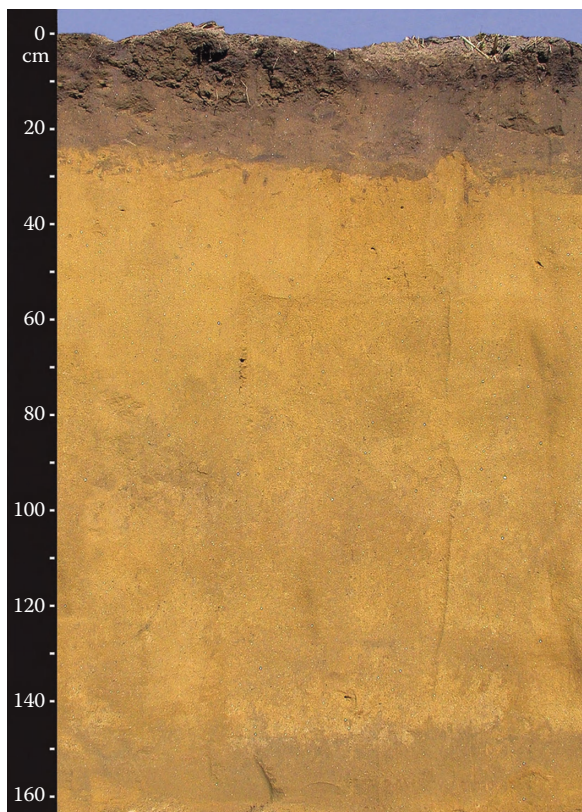
Entisol

Gelisol

Histosol
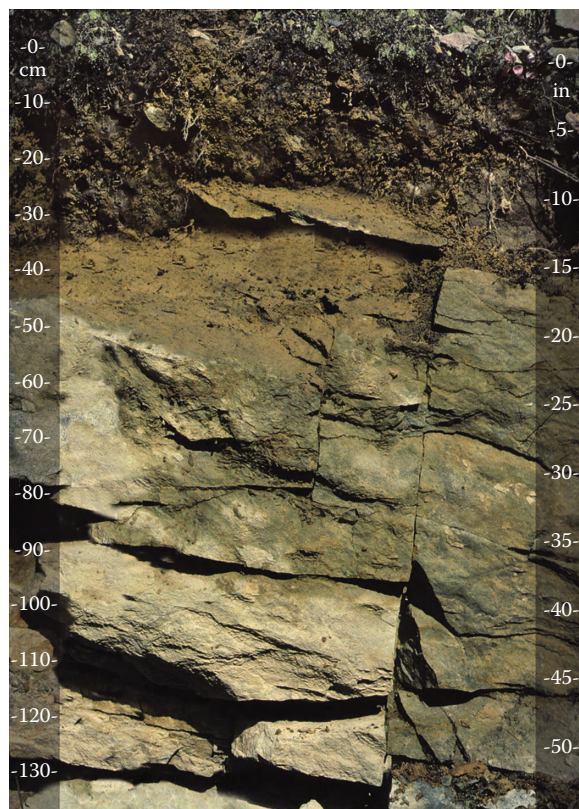
Inceptisol

Mollisol

Mollisol (2)
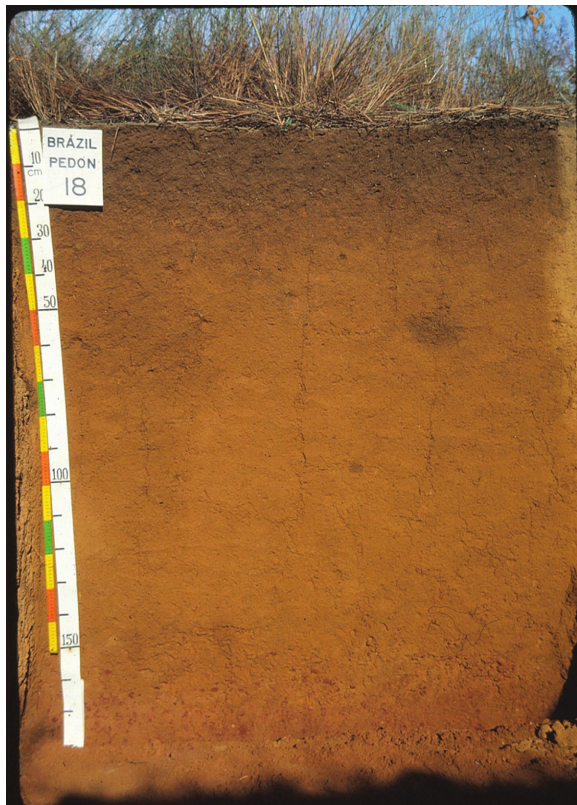
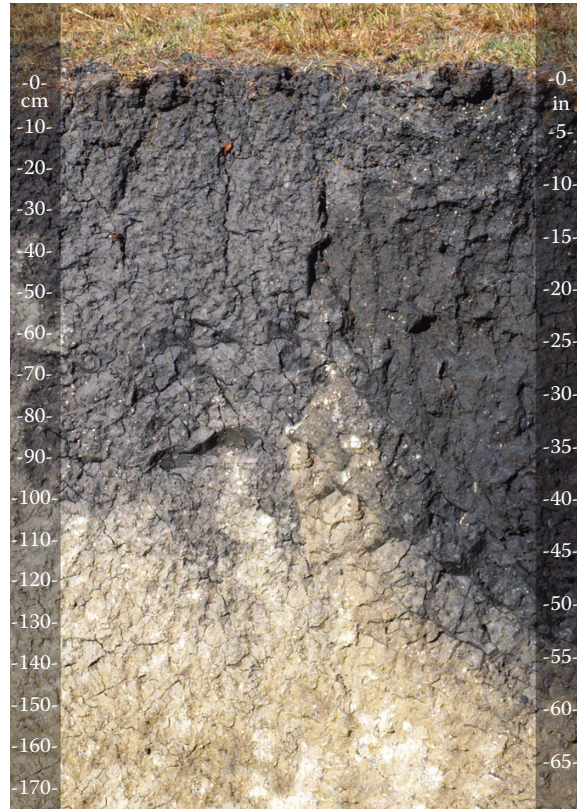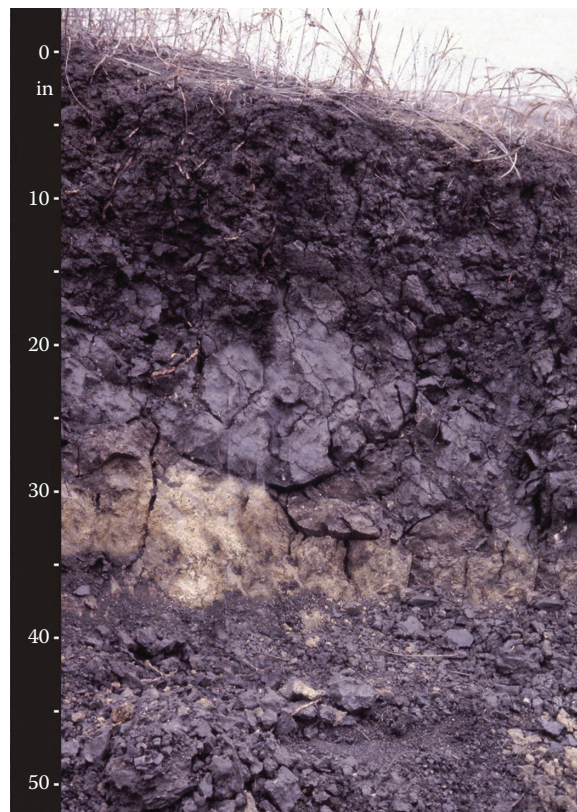
Oxisol
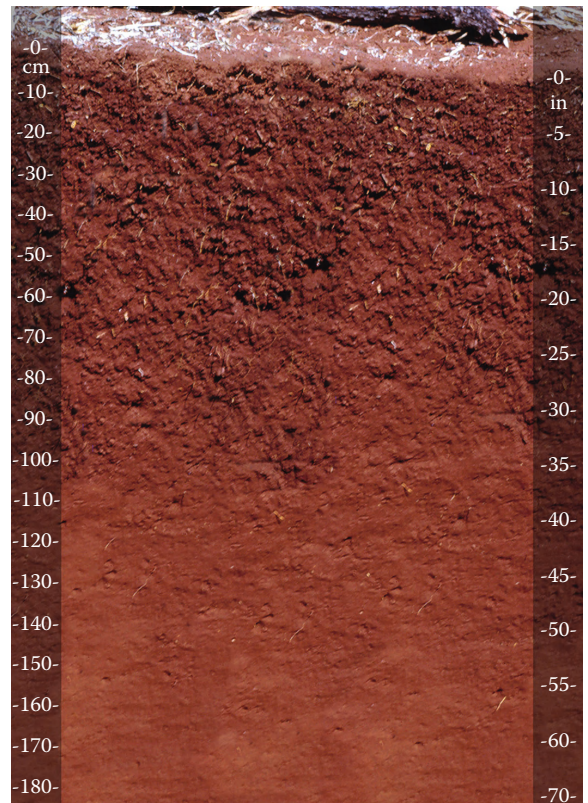
Oxisol (2)
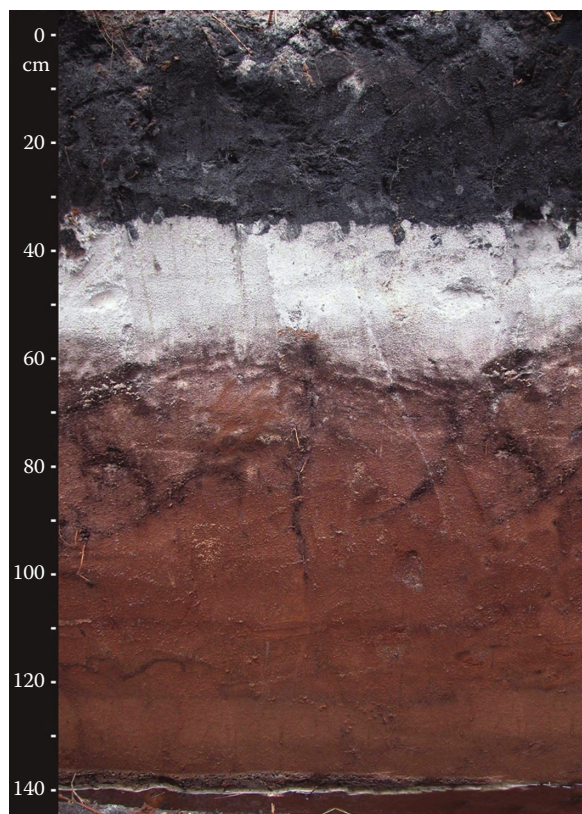
Spodosol
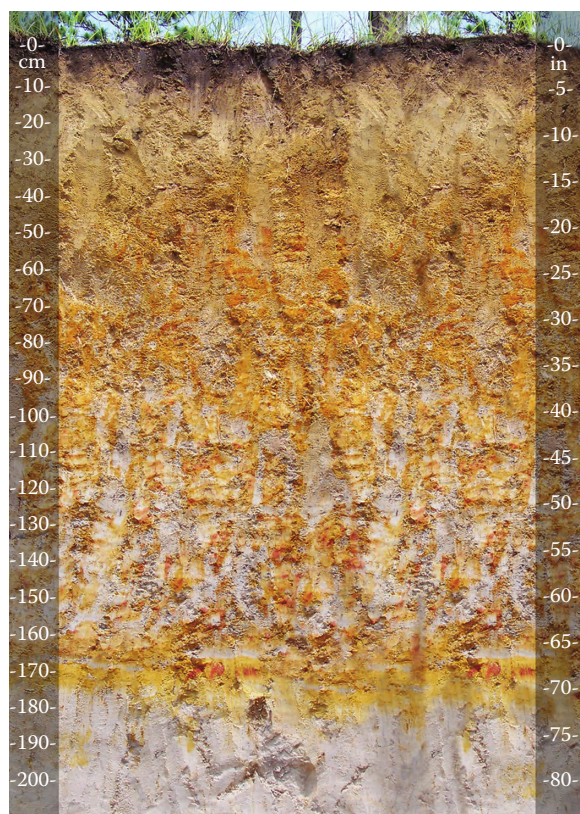
Ultisol
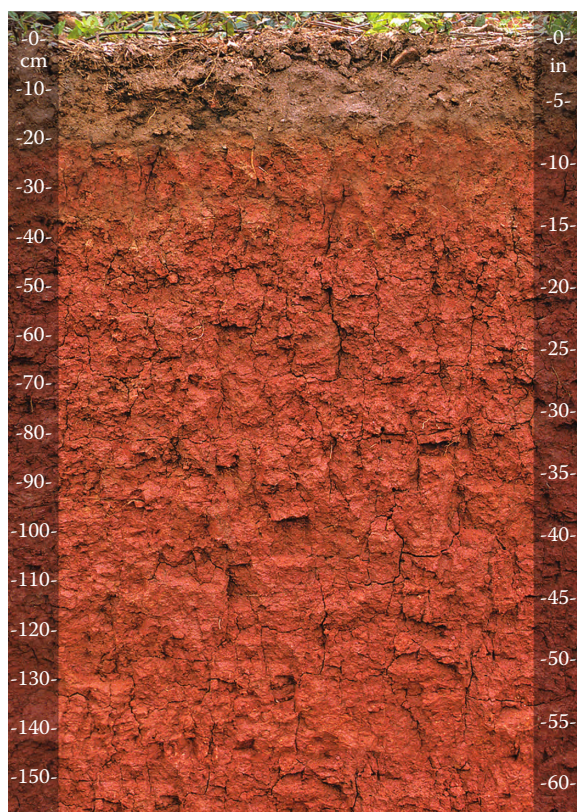
Ultisol (2)

Vertisol

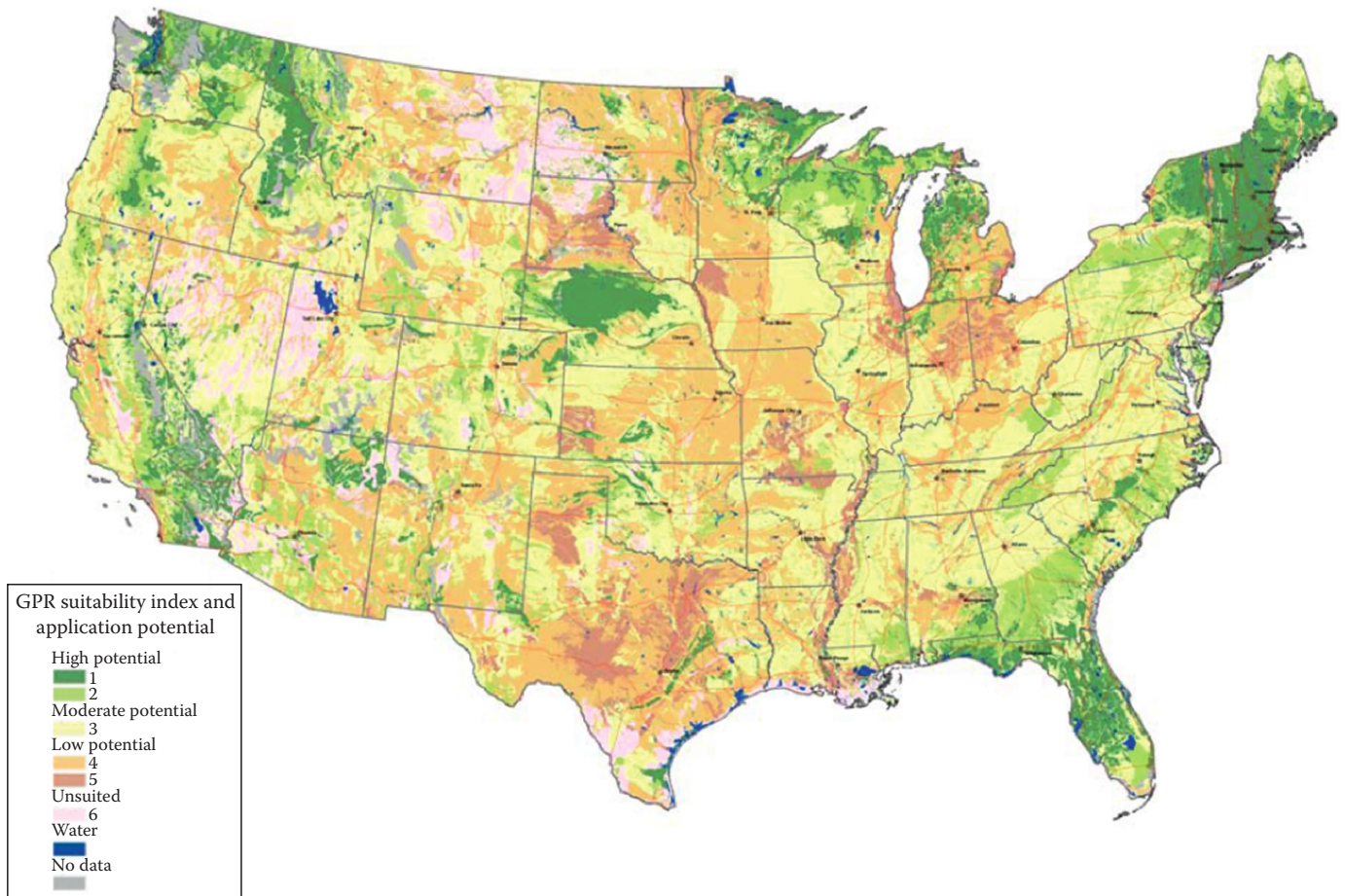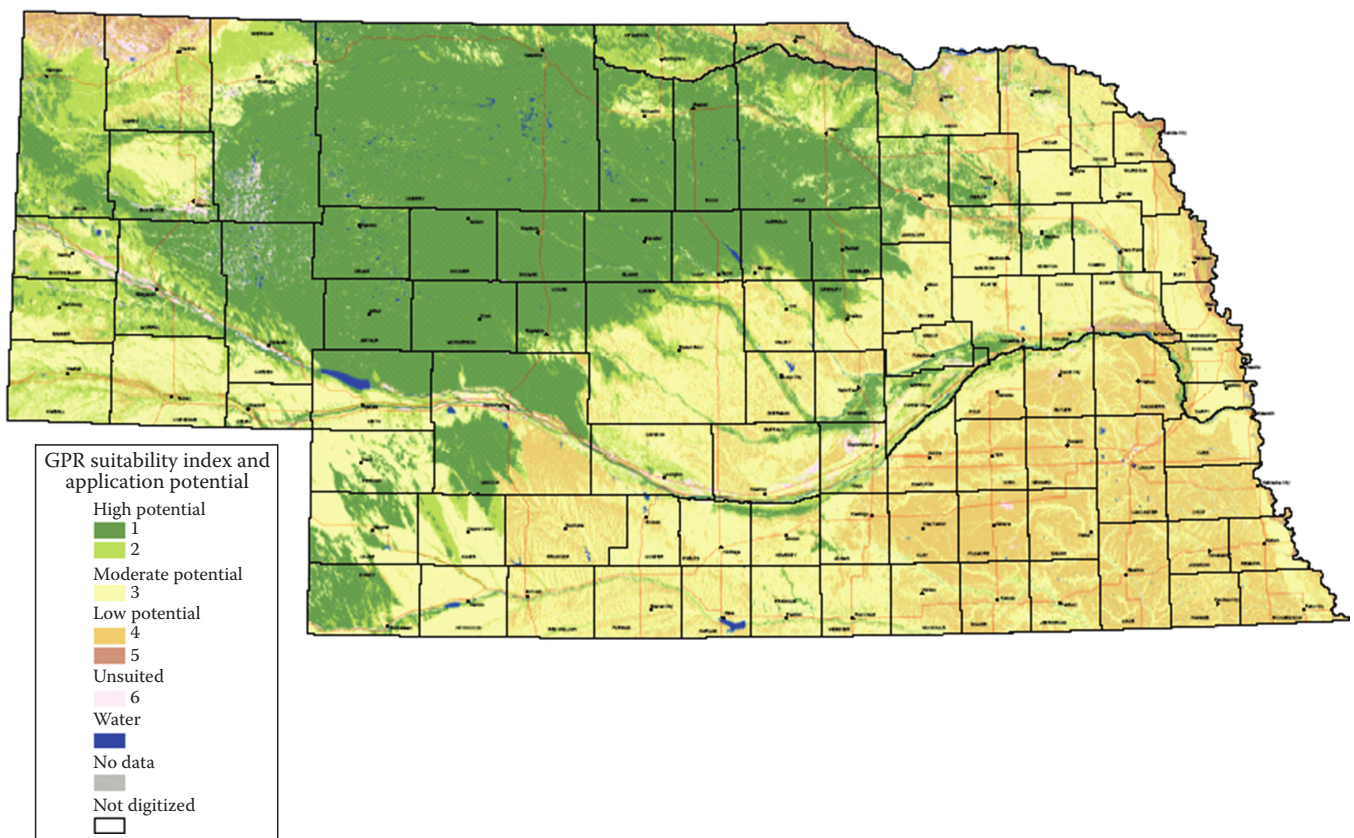Vertisol (2)

FIGURE 39.2 The GSSM-USA.

FIGURE 39.3 The State GPRSSM-NE.